5-2001

Mathematics Achievement in Tennessee Schools in the Context of Opportunity to Learn.

Richard Lee Kitzmiller
East Tennessee State University

Mathematics Achievement in Tennessee Schools In The Context of Opportunity to Learn

---

A dissertation

presented to

the faculty of the Department of Educational Leadership and Policy Analysis

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Doctor of Education

---

by

Richard Kitzmiller

May 2001

Louise MacKay, Chair
Gunapala Edirisooriya
Russell Mays
Jack Rhoton

Keywords: Opportunity to Learn, Student Achievement, Mathematics, Accountability, Value-added

ABSTRACT

Mathematics Achievement in Tennessee Schools
In The Context of Opportunity to Learn

by

Richard Kitzmiller

This study examines the relationship of student achievement in mathematics and factors purported to influence such achievement. The factors used in the study were selected from those contained in accountability reports issued by the state of Tennessee in 1995-96. The student achievement measures were based on four high school end-of-course mathematics tests. Student performance on these tests is examined both in terms of *actual* scores and *value-added* measures. This study organizes the available accountability measures in an Opportunity-to-Learn (OTL) framework for the purpose of determining relationships that can inform practice and give policy guidance.

The study examined 65 (of 139) school systems in Tennessee in a research design of correlation and multiple regression analyses. Twenty-three independent variables were organized into three OTL categories (fiscal, educational process, and teacher quality) and an "external factors" category. Eight dependent variables represented *actual* and *value-added* results on the four tests.

Results revealed a number of significant relationships; there were relatively fewer and weaker relationships involving *value-added* measures than *actual* measures of achievement. Conclusions emphasize the need for continued refinement of the accountability and research goals for the state. Specific recommendations are that the number of variables measured be increased and that the focus and specificity of the variables be increased.

CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

There is mounting federal and state commitment to using educational achievement tests as a way to monitor and promote the success of schools in the United States. In addition there are increasing attempts to estimate the effects of a wide range of societal and school variables on educational outcomes for the purpose of informing educational policy. As policymakers focus on different aspects of the education problem, evaluations will need to address such topics as fiscal equity, opportunity to learn, differences based on ethnicity and socioeconomic status, and a host of other issues. Often the accountability efforts, the research efforts, and the focus on social issues are occurring simultaneously with school finance reform efforts.

There are two perspectives that have dominated recent discussions among advocates of school finance reform. School finance reform traditionally has meant the reduction in aggregate differences in spending between school districts, especially when the differences are based on economic status of the students and/or communities (Arnold, 1998). A more recent perspective places an emphasis on productivity and calls for efficiency, with spending only on proven strategies for improvement. Wenglinsky (1997) has given an interesting, generalized statement of the two perspectives; he referred to the two as “pure” types of characterizations, with many researchers and policymakers borrowing from both perspectives. He said, “The traditional view holds that nearly all forms of education spending have an effect on student performance; therefore, additional funds for all these forms will raise student achievement” (p. 2). “The productivity view, on the other hand, holds that few (if any) of the conventional forms of spending, when increased, result in achievement gains” (p. 3).

Key Concepts and Terms

There are several terms and concepts that are critical to an understanding of this study. One of these is the *Tennessee Comprehensive Assessment Program*, or *TCAP*. TCAP is the entire package of state-mandated assessments of student performance. The TCAP program was initiated in 1990 by the Tennessee Department of Education and has continued with some modifications since that date. From this package of assessments, two are used in this study.

The two assessments included in this study are the TCAP Achievement test and the High School Subject Matter Tests. The *TCAP Achievement Test* is given in elementary and middle schools in Tennessee and is very typical of achievement tests given across the country. *The High School Subject Matter Tests,* also called *end-of-course tests*, are custom-generated to correlate to the Tennessee curriculum but are similar to other states' criterion-referenced mathematics end-of-course tests or those published by textbook companies.

Value-added methodology, in general, measures the growth in individual student performance from one point in time to another. The emphasis is on *gains* rather than absolute levels of student achievement. One of the most sophisticated value-added accountability systems is in place in Tennessee.

The *Tennessee Value-Added Assessment System (TVAAS)* is a statistical process specifically mandated by the 1992 Education Improvement Act. Initially, the TVAAS processing and reporting (including the analyses included in this study) was conducted under a contract with the Value-Added Research and Assessment Center of the University of Tennessee-Knoxville. The TVAAS methodology was championed by and is administered by Dr. William Sanders.

The TVAAS statistical model calculates system, school, and teacher effects from the aggregate of individual students' performance on mandated tests. Bock and Wolfe (1996) said of TVAAS, "The system and school models are typical two-stage hierarchical models for educational survey data.... The teacher model is three-stage hierarchical with teachers the third stage" (pp. 39-40). Fisher (1996) said the TVAAS model is based on the "Henderson mixed-model methodology." Of the model he said:

> This is a very complex statistical methodology that fundamentally is a repeated measure, regression system that makes use of all available information, not being negatively affected by missing values in the data set. The label "mixed model" is derived from the fact that a score is a sum of both fixed and random effects.... In the Tennessee situation, teacher effects are considered to be random while school, program, etc. are considered to be fixed. The model is able to estimate the effects of a teacher, school, or district without actually having the "gain score" as one of the data entries. (p. 35)

While discussing the development of the underlying statistical theory over more than 25 years, Bock and Wolfe said:

> Multilevel models in education are now well known from the work of Aitken & Longford (1986), Bryk & Raudenbush (1992), Goldstein (1987), and others.... They are now the accepted standard for analyzing growth and other effects in social and educational survey data, including large-scale assessment. (p. 40)

Although there is an apparent sound statistical basis for the TVAAS analyses, there are unique features that are still questioned. The biggest controversial aspects, however, deal with the conclusions drawn from the results, the applications of the findings, and the emphasis placed on the single value-added measure.

Another key concept that is central to this study deals with students' opportunity to learn. Traditionally, discussion in this area has been general and often has centered around the availability of a program of studies. With the emphasis placed on standards-based reform, which gained momentum in the early 1990s, the concept of "opportunity to learn" was more specifically defined and gained increased attention.

Standards-based reform has included content standards and performance standards that describe what students should know and how well they should know it. In a discussion of the three types of educational standards distinguished in the Clinton administration's Goals 2000 Act, Linn (1994) has included Opportunity-to-Learn standards with the other two categories of standards. *Opportunity-to-Learn (OTL)* standards address the problem of equity in education, that is, ensuring fairness and equal possibility of success for all students. Porter (1995b) viewed OTL standards as the "criteria for assessing the efficiency or quality of the resources, practices, and conditions necessary at each level of the educational system" to provide students an opportunity to learn the material (p. 23). This concept extends the notion of curriculum standards by emphasizing the role of the school and units of government in being accountable to students for proving adequate resources and conditions for learning.

In Tennessee, the *Career Ladder* program is an incentive program in which high-status professional educators are recognized and compensated with extra pay. To participate in the voluntary Career Ladder program's upper levels (II and III), eligible teachers undergo an extensive process of self-report, observations, and structured interviews to formulate evidence of professionalism and productivity.

Statement of the Problem

Most of the attention on value-added measures of accountability in Tennessee has been centered on the calculations for elementary and middle schools based on results of the TCAP achievement tests. The methodology used to determine value-added scores for the high-school end-of-course tests has not received the same external scrutiny as the methodology used for elementary and middle school programs. Neither has there been the depth of discussion within the state nor a comparable degree of published research. Policy discussion has been limited, as has media attention concerning end-of-course mathematics test results. This study attempts to fill this void and examine factors influencing these results.

Background of the Problem

The state of Tennessee has been considering educational funding and accountability issues over the past few years. In 1988, the Tennessee Small School Systems, representing 77 of the 139 districts in Tennessee, filed suit in chancery court against the state, claiming that, because of disparities in funding, the state school finance system violated the education clause and the equal protection requirements of the state constitution. The suit passed through the various levels of the court system, ending up in the state supreme court. In 1993, the state supreme court ruled that Tennessee's school finance system was unconstitutional. The disparity in funding that was evidenced in the case led the court to conclude that students in the plaintiff schools did not have equal access to adequate educational opportunities, such as laboratory facilities; computers; current textbooks; buildings; and music, art, and foreign language courses, some of which were required by the state. Plaintiff schools had difficulty retaining teachers, funding needed administrators, and providing sufficient physical education and other programs (U.S. General Accounting Office [GAO], 1995a, pp. 39-40).

The court also linked inadequate funding of the plaintiff schools to their educational outcomes. This ruling was unique among school finance court decisions; Tennessee was one of only three states where a court ruled on student achievement (U.S. GAO, 1995a, p. 7). The court noted that, in the 10 wealthiest districts for the 1988-89 school year, 66% of the elementary schools and 77% of the secondary schools were accredited by the Southern Association of Colleges and Schools compared with 7% and 40%, respectively, in the 10 poorest districts. The court observed that graduates from accredited high schools have better success in securing college acceptance. Students in the plaintiff schools had poor standardized test results and more need for remedial courses in college (pp. 39-40).

In the decade of the ‘90s, the state dramatically increased funding and addressed court mandates to make educational funding more equitable. Several reports showed the need for Tennessee to increase funding. The U.S. General Accounting Office (1995b) reported that in

1992 Tennessee ranked very low among the states in its willingness to raise revenue for education. The comparisons and analyses in the report used measures of the *actual* state revenue relative to the state's *ability* to raise revenue. The report cited Alabama, Delaware, Hawaii, Nevada, and Tennessee as "states with very low willingness to raise revenue" (pp. 11-12). The index for Tennessee's willingness to raise educational revenue was 0.68 (with 1.00 as the national average); the second lowest index was Delaware's 0.79 (pp. 52-53).

In the 10-year period preceding this analysis, 1982 - 1992, Tennessee's willingness to raise revenue *overall* had roughly matched the national pattern. However, the percent change in Tennessee's willingness to raise revenue *for education* had dropped 13.4%, while nationally there was an 8.4% increase in this measure over the same period (U.S. GAO, 1995b, p. 54).

In 1995-96, Tennessee spent an average of $4,172 on each of its 893,770 students; this compares to the national average of $5,689 expended per pupil (National Center for Education Statistics, 1998a, p. 8). The $5,689 per pupil expenditure represents a 0.6% increase from 5 years earlier, after adjusting for inflation (p. 1).

The Education Improvement Act of 1992 was Tennessee's major policy initiative to bring about the desired reforms. As a companion to the increased funding, the legislation also included policy efforts to increase expectations for students, staff, schools, and districts and to increase accountability. The accountability has in large part been focused on results from the Tennessee Comprehensive Assessment Program (TCAP). A new component in Tennessee's system of accountability has been a value-added component, which for the most part, is unique to the state of Tennessee.

The Tennessee Value-Added Assessment System (TVAAS) makes use of a statistical mixed-model methodology. Dr. William L. Sanders and Dr. Robert A. McLean of the University of Tennessee proposed this statistical approach to eliminate what had previously been cited as impediments in educational outcome-based assessment systems. The TVAAS, according to its proponents, will "eliminate, or at least trivialize," the following problems: missing student records, various modes of teaching (self-contained classroom versus departmentalized instruction versus team teaching), teachers changing assignments over the years, transient students, regression to the mean, different variance-covariance structures across school systems, and the need to include concomitant covariables (Sanders & Horn, 1994, p. 299).

The Tennessee Comprehensive Assessment Program (TCAP) achievement test program was first administered in the 1989-90 school year. Administered annually in the spring, the TCAP achievement tests include norm-referenced subtests that are not included in TVAAS analyses (such as listening and study skills) and other norm-referenced subtests on which TVAAS does focus. For the norm-referenced portions of the TCAP, the state contracted with the CTB/McGraw-Hill Company and administered their product, CTBS/4, through the 1996-97 school year. This test was nationally normed in 1989 and makes no particular reference to the Tennessee state curriculum (no more than to other states'). Also included in the achievement test battery, to relate directly to the Tennessee curriculum, were two criterion-referenced subtests, in language and mathematics. Beginning with the 1997-98 school year, the CTBS/4 was replaced with the newer CTB/McGraw-Hill product, the TerraNova test.

Although not included in value-added analyses, it could be noted that the comprehensive nature of the TCAP program includes a writing assessment (Grades 4, 7, and 11) and a mandated proficiency test that is required for graduation. Additionally, all students are required to take an exit examination; this can be either the College Board’s SAT, the ACT, or the Work Keys test.

TVAAS is a statistical process that provides measures of the influence that school systems, schools, and teachers have on indicators of student learning (Sanders & Horn, 1994, p. 301). The initial focus of TVAAS was on grades 3 through 8, analyzing the scale scores students made on certain norm-referenced sections of the TCAP achievement test. The initial TVAAS analyses and reports focused on the areas of reading, language, mathematics, science, and social studies, with an emphasis on “gain” scores.

The Education Improvement Act of 1992 also required value-added assessment for high school students. The addition of value-added measures at the high-school level widened the scope. The first high school tests administered for that purpose were in the area of mathematics. At the end of the 1995-96 school year, five mathematics courses were tested in the initial administration. The courses and the number of students who took each test are as follows: Pre-Algebra - 20,040; Algebra 1 - 41,248; Algebra 2 - 27,339; Geometry - 30,599; Math for Technology - 7,621 (Tennessee Department of Education, 1996, p. 79). Other subjects are to be added as tests are developed. For some reason, the number of test takers is considerably smaller than the number of students enrolled in these courses; the Tennessee State Board of Education reported 48,849 students enrolled in Algebra 1 in 1995-96, with 33,118 in Algebra 2, and 36,314 in Geometry (Tennessee State Board of Education & Tennessee Higher Education Commission, 1997, p. 7).

The questions in the High School Subject Matter Tests were developed collaboratively by Tennessee teachers and the CTB/McGraw-Hill company. As contrasted with the nationally-normed TCAP achievement tests used through the eighth grade, performance on these tests cannot be compared to other states or the nation as a whole. The tests were customized for Tennessee and were not nationally normed.

With the initial administration of the High School Subject Matter Tests, also called end-of-course tests, Tennessee entered a new and unique phase of accountability. This aspect of Tennessee’s accountability model is unique for several reasons. One feature is that the achievement scores on the subject tests are compared to previous measures of achievement, primarily scores on previous mathematics achievement tests. This comparison is made in spite of differences in the types of tests. The TCAP achievement tests are reflective of a national curriculum and are primarily norm-referenced. The TCAP end-of-course tests are criterion-referenced and reflect the Tennessee curriculum. The end-of-course tests are not on the same scale as the TCAP achievement tests, although the Pre-Algebra; Algebra 1, and Algebra 2 tests themselves share a common scale.

The methodology used to determine value-added scores for the end-of-course tests has not received the same external scrutiny as the methodology used previously for value-added measures for elementary and middle school programs. The accountability methodology places a fairly strong focus on teacher-, school-, and district-level evaluation, using an instrument that is "no-stakes" for the student.

Because of the uniqueness of the situation, there is a corresponding unique opportunity to examine the relationship of student achievement to a variety of factors related to the teaching-learning process.

Purpose of the Study

This study attempts to contribute to the debate of using achievement measures to monitor and influence policy intentions. In their discussion of assessing Opportunity to Learn (OTL), Herman and Klein (1997) cited "at least three points where OTL data can serve the critical functions of verifying the policy assumptions and contributing to system decision making" (p. 4).

Specifically, this study examines the High School Subject Matter Tests in Tennessee with the goal of comparing policymakers' *desired* results to *actual* results. Herman and Klein's first point was that the central theme of the policymakers' reform logic "is the notion that school curriculum, teaching, and instruction will change to provide students with appropriate opportunities to learn and to achieve the high standards that are held for them" (1997, p. 4). They stated that the policy logic is that, without such changes, "improvement in student performance is unlikely and the policy tool ineffective in achieving its ends. Yet, absent data on OTL, how will policymakers know?" (p. 4).

In the spirit of providing such data, this study will examine fiscal factors to see if the move toward financial equity is resulting in student performance equity. There will be an examination of factors relating to educational practices. For instance, the use of a block schedule has been effected by some schools as a response to the increased graduation requirements (and for other qualitative reasons). The impact on achievement of block scheduling, course-taking patterns, attendance and dropout rates, and other factors dealing with the educational process is examined. Other Opportunity-to-Learn (OTL) factors examined include measures relating to the school district's teachers, including Career Ladder status, educational background, and the average teacher's salary. The influence of factors generally considered to be beyond the control of schools is also studied.

Dorn (1998) has said, "Accountability should connect student performance with classroom practice. Statistical accountability, with the abstraction of student performance into numbers without context, removes classroom practices from the discussion of educational reform" (paragraph 50). This study attempts to provide context to the abstraction of the accountability measures. This context reflects both classroom practice and school- and district-level characteristics.

Research Questions

The value-added measure called *school effect*, calculated for the end-of-course mathematics tests, will serve as one of the measures of student achievement. The other measure, the *actual* measure of student achievement, will be the mean scale score for each subject tested.

Although five subjects were tested in 1995-96, only four were included here (Pre-algebra, Algebra 1, Algebra 2, and Geometry). Math for Technology results are not included because of the sporadic implementation of the course across the state. This study will address the following questions:

1) Is there a relationship between mathematics achievement and student population characteristics and/or community characteristics?

2) Is there a relationship between mathematics achievement and fiscal factors ?

3) Is there a relationship between mathematics achievement and characteristics of the educational process?

4) Is there a relationship between mathematics achievement and measures of teacher quality?

Significance of the Problem

Considering the enormity of the effort that has taken place in Tennessee in the recent past, it is time to examine the results. Looking at the accountability efforts, the efforts in equalizing resources, and the curricular and instructional changes in an integrated fashion seems to be quite appropriate and useful. It has been said that large-scale assessments used by the states have been analyzed statistically for technical appropriateness, but not for whether the assessments mean what users think they mean. Stake (1998) says, “Presumption that assessments indicate quality of teaching, appropriateness of curricula, and progress of the reform movement--commonplace presumptions in political and media dialogue--is unwarranted. Proper validation would tell us the strength or weakness of our conclusions about student accomplishment” (paragraph 10). He concludes that this (the most important) type of validation of statewide assessment programs has not taken place.

Studies such as this one hopefully will yield insight into what works and what doesn’t. The relative importance of myriad simultaneous factors may be realized and perhaps more will be learned about the nature of accountability measures--the instruments and methods as well as the interpretation and uses of the data. Herman and Klein (1997) have advocated that OTL data can provide policy makers with early feedback on system progress. They stated, “We would argue that data on students’ opportunity to learn can provide an interim measure of system progress as well as important data to inform mid-course corrections” (p. 5).

Part of the urgency in examining policy implementation is because the very nature of policy logic has changed. The current focus on accountability and the emphasis on results is a departure from the thinking that has prevailed. The current policy transition to results is somewhat unsettling in that it follows nearly a century of focus on inputs. During the time of focus on inputs, states implemented policies to insure quality and equity by standardization and common regulations. The quest for equal education opportunity has focused on school finance equalization. Over time, small disparate school districts have been consolidated into a smaller number of larger districts, to become more efficient and to provide better learning opportunities. Federal emphases have focused on additional resources for at-risk students and students with special needs. The focus to a "results" orientation, however, has changed student achievement from an implicit to an explicit policy goal.

Another reason to examine the policy implementation in Tennessee is because of expressed concern from neutral, outside observers. In the GAO case study of Tennessee's efforts to equalize funding and educational opportunities, the authors reported that the officials they interviewed expressed concern, among other things, about the students' **opportunity to learn** [emphasis added], based on the implementation of performance goals prior to the full funding of the Basic Education Program (U.S. General Accounting Office, 1995b, p. 47). Many have cautioned that meaningful change in educational practices takes time (Fullan, 1991; Hall & Hord, 1987; Huberman & Miles, 1984), perhaps even time after full funding.

Adding to the significance of this study is the fact that the state of Tennessee has entered into a new era of accountability. In 1998, the Tennessee State Board of Education approved the Algebra 1 end-of-course test to be one of three high-stakes tests, called Gateway tests, with a passing grade being a graduation requirement for students. This added level of accountability (for the student) sets in place a higher level of technical expectations for the test(s) and a burden of proof that OTL issues have been addressed by the school systems of the state. Minimum-competency tests have been expected, by judicial rulings, to have curriculum or instructional validity. As Linn (1994) related:

> The match between what was tested and what students were taught was one of the key issues in determining whether students had been provided with a fair opportunity to learn the knowledge and skills required by the minimum-competency test in the *Debra P. vs Turlington* case (474, F Supp. 244, M. D. Fla., 1979). (pp. 8-9)

Linn concluded, “If OTL standards are required in the future either as the result of legislation or judicial decisions, they will pose daunting measurement challenges” (p. 17). Herman and Klein said, “Absent data on OTL, how can policy makers and the public verify that all students have a chance to succeed, and that all subgroups of the population have the opportunity to engage in the kinds of curriculum and instruction that would prepare them to achieve expected standards?” (1997, p. 4).

A U.S. Department of Education report on education research and reform supports Linn’s contention that OTL standards will be of critical importance. The editor of the report (Tomlinson, 1994) stated:

> As the states hold their students to uniform and higher standards of achievement, they must also assure that all students have a fair and equitable opportunity to meet the new expectations. Accordingly, states would be expected to develop and meet opportunity-to-learn standards in the same manner as they develop and meet the curriculum and assessment standards.... This step is also necessary before academic failure may be fairly ascribed to reasons within the students’ control--for example, lack of effort--rather than those outside of it, such as incompetent teaching. (p. 178)

Willms and Raudenbush (1989) have said, “Without an understanding of how much variation between schools in their performance that can be attributed to school practice, to wider social factors, or simply to random fluctuation, there is potential for the misuse of data on schooling indicators" (pp. 209-210). Tomlinson has called for a redirection and resuscitation of the federal role in education research, development, and dissemination to add to this understanding. He said, “When the states may be required to spend great sums of money to meet the opportunity-to-learn standards, it is not unreasonable to establish a valid and causal relationship between school resources and academic achievement before the reforms are implemented" (1994, p. 179).

Limitations

One limitation of the study is that some data (financial, primarily) is available only at the county level, with the desired unit of analysis being the school or district. In these instances, potential data for this study were not appropriate and not included in the analyses.

Another issue that is ignored in this study, as in almost all fiscal studies, is that of outside funding sources. Information on revenue received from foundations or grant sources is seldom included in such analyses. Similarly, schools and school districts vary in their willingness and ability to generate additional discretionary funds through fund-raising activities (Arnold, 1998, p. 36).

Another limitation is that the plan of equalizing funding is not yet complete. "Full funding" for the plan was achieved in the 1997-98 school year, but a decision on equalization of teacher salaries is currently pending. In short, the equalization process is evolving and there has been insufficient elapsed time for the effects of the equalization to be fully actualized. Similarly, the effects of the implementation of block scheduling may be influenced by its relative "newness."

Overview of the Study

This study was organized and is presented in five chapters. Chapter 1, Introduction, includes a statement of the problem, the purpose of the study, the significance of the problem, limitations of the study, and an overview of the study. The major research questions are introduced in Chapter 1.

Chapter 2, Review of the Literature, presents a review of literature related to various aspects of accountability, including measuring effectiveness at various levels and accountability by student test scores. This chapter also presents a review of literature related to certain technical aspects of tests, scales, and scores, including scale characteristics, scale linkage, and gain scores. External influences that influence measures of student achievement are examined. Chapter 2 also presents a review of literature related to opportunity-to-learn factors. Fiscal factors are one category of opportunity-to-learn factors examined. Other OTL factors examined pertain to the educational process, including the course-taking pattern, implementation of the block schedule, and time on task. Another OTL factor examined related to teacher quality.

Chapter 3, Research Design, describes the methods and procedures used to obtain and compile research data in the study. Hypotheses are presented that were examined to test the major research questions. The chapter also describes the research design and procedures used to analyze the data.

Chapter 4, Presentation and Analysis of Data, provides a description and interpretation of the results of the statistical analyses.

Chapter 5, Discussion and Conclusions, summarizes the research findings, discusses conclusions and implications, and presents recommendations.

Willms and Raudenbush (1989) said, "Although the literature provides some guidance on what variables are relevant, we feel that the processes are complex and interactive, and that theories about school effects are in their early stages of development" (p. 228). This study addresses many of these relevant variables in a framework that addresses issues in Tennessee. Willms and Raudenbush continued, "As successive waves of data on schooling indicators become available, researchers will be able to conduct more powerful analyses that address questions concerning the effects of school organization and practice on changes in school performance (p. 228). It is hoped that this study will be at least a small part of one of "the successive waves of data."

CHAPTER 2

LITERATURE REVIEW

Because the primary focus of the TVAAS is to provide accountability information on teachers, schools, and districts, an examination of research in these areas is a major emphasis in this chapter. The technical aspects of tests, scales, scores, and statistical analyses are also examined in this chapter. External influences that confound the issue of accountability are also explored. Finally, opportunity-to-learn concepts and issues are examined to provide a framework for this study. Monk and Rice (1999) have said:

> The advent of high performance standards has renewed efforts to understand and enhance the productivity of educational systems. Analysts are struggling to grasp the resource implications of these standards in the face of inadequate conceptualizations of educational productivity, imperfect data, and inadequately developed statistical tools and research methods. (p. 115)

This literature review addresses each of the areas described by Monk and Rice to establish a background for this study.

Accountability

Accountability has multiple meanings, although the term is often used imprecisely as if there were a single, consensus meaning. Multiple meanings exist both in a general sense and in the sense of specific statistical systems designed to measure accountability. Dorn has said that the apparent consensus of meaning hides the differences and conflicts among the various meanings of "accountability" (1998, paragraph 15).

Dorn (1998) described six different concepts or purposes of accountability and illustrated the types of statistical analyses accompanying each. One concept is judging public schools (as a collective set of institutions); the annual release of average SAT scores is one example of statistical measures related to this concept. Another is judging teachers and other educators, in a comparative sense, using such statistical reports as the U.S. Department of Education's "Wall Charts". The Tennessee Value-Added Assessment System was specifically listed as an example of a statistical methodology with this intent. Judging students is another form of accountability. The use of minimum competency tests as a requirement for graduation is one example of this high-stakes variety of accountability.

One might also use standardized test scores to judge public policy. In 1969, the National Assessment of Educational Progress (NAEP) tests were initiated as a low-stakes means to evaluate public school policy with objective data. Recently, there has been growing interest in using NAEP data to make judgements in a more high-stakes manner. These attempts have been at the center of debate over the dual intents (Jones, 1996). Other manifestations of accountability include the building of organizations (public and private organizations producing educational statistics) and marketing (the use of student statistics to bolster public support and/or to attract students).

Dorn (1998) contended that these varied purposes are not necessarily congruent. Indeed, the purposes may be at odds with each other. Additionally, the aims of accountability do not easily accommodate other relevant educational issues, including fostering equity, giving a sense of direction to curriculum development, and educating for a changing world (Darling-Hammond, 1992).

Accountability systems are built on the assumption that there is a consensus on educational goals and that a single system can be developed that measures whether or not these goals are met. The creation and establishment of a statistical accountability system may freeze the given foundational assumption as a political and functional framework for years to come.

There are, however, other viewpoints as to how the foundational assumptions of accountability systems should be established. Fetler (1994) said that most accountability systems currently in place have a rational perspective with a strong emphasis on structural characteristics. In addition to the structural theory of organization, there are also human resource, cultural, and political theories--each with a different focus. Fetler added that reports on student achievement are helpful--especially if they can be causally linked to structural factors that can be changed (e.g., staffing, budgeting, planning, facilities, curriculum, instructional techniques, technology, etc.). He said that an understanding of how performance reports affect organizations, however, should include consideration of non-structural dimensions. The rational approach pays less attention to the way people interact to do work, cultural values and norms, and the political process of dividing up resources.

Dorn (1998) said there are alternatives to the statistical accountability systems that currently dominate. He said the alternatives would be both open and political. Accountability, in this alternative scenario, connects schools "in a meaningful and explicitly political way with broader communities.... No statistics can substitute for the type of immediate contact such external evaluation provides" (paragraph 48). Dorn cited Sizer (1992) as a supportive example, saying that Sizer advocates opening up schools to external evaluation for pedagogical reasons, to keep teachers in touch with reasonable expectations of what students should do.

Dorn (1998) also said, "Statistical accountability systems intensify educational triage, encouraging schools to isolate and devote fewer resources to students whom schools judge as difficult to teach" (paragraph 50). Numerous reports show instances in which students are selectively excluded for participating in assessments to insure higher reported average scores. By law, TVAAS excludes the scores of students receiving special education services from the calculation of value-added scores. Such situations raise the question of whether the education of *all* students is undervalued as a goal, when the accountability model rewards a selective focus toward *some* students.

Linn (2000) says that the most recent wave of reform continues to emphasize accountability, as have previous reform efforts; however, there are some significant new features. Among these are the inclusion or emphasis on performance-based approaches to assessment, the concept of tests worth teaching to, and "the politically controversial and technically challenging" issue of opportunity to learn (p. 8). He also discussed the new features of (a) the emphasis on the development and use of ambitious content standards as the basis of assessment and accountability, (b) the dual emphasis on setting demanding performance standards and on the inclusion of all students, and (c) the attachment of high-stakes accountability mechanisms for schools, teachers, and, sometimes, students. This last point leads to a discussion of accountability at various levels.

Measuring Effectiveness at Various Levels

Many current accountability efforts include several levels of emphasis. After analyzing decades of previous research relative to relationships between school inputs and student outcomes, Willms and Raudenbush (1989) concluded that at least four principles are important to those collecting data on indicators of school performance. Among those principles are that:

> Relationships between variables at one level influence relationships at other levels. An indicators program aimed at determining school quality must include individual-level measures of students' outcomes and background characteristics, school-level variables describing school policy and practice, and community-level data describing factors such as local employment rates, opportunities for further education, and community resources. (p. 214)

These authors based their conclusions on applications of hierarchical linear models to cross-sectional data.

In a meta-review and synthesis of research (Wang, Haertel, & Walberg, 1993), the state- and district-level policy influences are reported to be relatively weak. The authors said their research shows "distal policy variables are less important to schooling outcomes than quantity or quality of instruction, home environment, or student characteristics" (p. 37). Variables included in the study's category labeled *State and District Variables* included state curriculum and textbook policies, testing and graduation requirements, and teacher licensure, as well as specific provisions in teacher contracts, and some district-level administrative and fiscal variables.

In his 1990 book, *Educational Finance: An Economic Approach*, David Monk of Syracuse University, discussed productivity functions in education. He noted that at various times researchers have been interested in outcomes of individual students, classes of students, schools, school districts, states, nations, ethnic groups, age groups, gender groups, and other population subsets. He concluded there is no one best approach, as he stated:

> It is not always the case that microlevel data are better than macrolevel data. The proper level of analysis depends largely on the nature of the phenomenon being studied. Some phenomenon are district rather than school or classroom phenomenon and have effects that are felt throughout entire school districts. (p. 327)

Centra and Potter (1980) presented a structural model depicting the relationships among many of the variables that contribute to variance in student learning. Their model is illustrative of many models showing a wide range of factors affecting student achievement. They said "parents, peers, teachers, schools, and, most of all, students themselves" are among the many factors affecting student learning (p. 273). They further illustrated the complexity of the issue by saying:

> Although the point is less frequently made, it is equally true that the effects of any one of the variables (or classes of variables)... on student achievement are likely to be small when compared with the combined interactive effect of all the other variables. (p. 287)

Barr and Dreeben (1977) stated:

> Such gains may be more or less likely to occur with a given form of instruction, depending upon school and school system conditions, community conditions, short- and long-term allocative decisions by

administrators, the state of the job market, the cost of going to college, the nature of administrative supervision, and the like.... A complete formulation of school effects must treat the full range of organizational levels and their interconnections. (pp. 101-102)

Another principle described by Willms and Raudenbush (1989) is that part of a school's effect is associated with the overall social class and ability composition of the school (p. 214). Researchers have struggled to determine how to properly separate the relative influences of the students' social class from school effects in studies. Bowles and Levin (1968) and McPartland, Epstein, Karweit, and Slavin (1976) argued that school effects often have been seriously underestimated because student background variables (i.e., social class) have been confounded with school differences. By controlling for social class, many of the studies have also controlled for school differences.

Researchers in Oregon (Schalock, Schalock, Cowart, & Myton, 1993) have reported on research supporting the relationship of *context* to teaching effectiveness. They reported that, in order to be effective facilitators of learning, teachers must consider and make decisions based on "not only differences in students, but also differences in subject matter, various learning goals within a subject area, available instructional resources, and available time for teaching" (p. 110). Lara and Medley (1987) found that teaching behaviors that fostered learning in students not only varied by level and subject area but also by high- and low- ability students within grades and subjects.

Schalock et al. (1993) argued that the overall context and expectation for learning is provided by the school and not by the individual teacher. An individual teacher is more likely to be "socialized" by the school than to influence the prevailing school culture. They argue that good teachers are a *necessary*, but not *sufficient*, condition for a good school. The authors cited other resources (besides teachers) that impact student learning within a school; these include time for learning, clarity of goals for learning, resources teachers have available to draw upon to aid learning, resources teachers have available to draw upon to assess student progress in learning, and the time a teacher has to reflect upon information about student learning and plan supplementary or corrective instruction if needed. The student learning demonstrated in a situation with few or none of these other resources may reflect poorly on an otherwise effective teacher (p.125). Resources are important opportunity-to-learn factors. Teachers are generally powerless to control the complete range of available resources; they are, however, impacted by the decisions that affect their working environment.

Willms and Raudenbush (1989) said that difficulties in measuring effectiveness exist not only at the teacher level, but also at the school level. "Schools can have varying effects on different types of students. Some schools may be particularly effective for high-ability students, but not for low-ability students, or vice-versa" (p. 214). Similarly, Mandeville and Anderson (1987) have reported findings on the lack of stability in school effectiveness indices across grade levels and subject areas; schools could be more effective in certain grades/courses and less effective in others.

Accountability Based on Student Test Data

Observers note a recent trend in the increased use of tests as educational policy tools. Stecher and Barron (1999) stated that the last two decades have seen a rapid increase in the numbers of state tests and that testing practices have changed. These changes include the introduction of new types of assessments, an increase in the stakes attached to the scores, and the incorporation of tests into formal accountability systems. Stecher and Barron claimed that these rapid increases have produced unresolved questions about the validity of scores produced by high stakes tests and the impact of these tests on classroom practices (p. 1).

Dorn (1998) also stated that the public judging of schools by test scores is relatively new in the United States. He said, “School statistics have existed since the late 19th century, and claims to objective measurement of student achievement from the turn of the 20th, but achievement scores have typically been only for internal consumption *within* school bureaucracies until recently” (paragraph 10).

Cronbach (1963) discussed the early history of standardized testing and classified the use of the evaluation results into three categories: course improvement, decisions about individuals, and administrative regulation. He defined administrative regulation as “judging how good the school system is, how good individual teachers are, etc.” (p. 673); today this would likely be called accountability.

In 1963, Cronbach was writing during a period of interest in course improvement. The National Science Foundation, in particular, was involved in efforts to improve mathematics and science curriculum and instruction. Cronbach reminded readers that course improvement was the goal of the first use of standardized testing. In 1897, Joseph Rice had administered a standardized spelling test in a number of American schools to collect evidence to provoke curriculum revision. Cronbach said that, in the 1920s, the emphasis shifted to an assessment of efficiency of the teacher or school. This “administrative testing” was often used “injudiciously and heavyhandedly” and was replaced in the 1930s by a use of tests almost exclusively for judgements about individuals (p. 673).

Despite reservations over the misuse or overuse of student testing as an instrument of accountability policy, the practice is still widespread. Linn (2000) stated, “There are several reasons for the great appeal of assessment to policymakers as an agent of reform” (p. 4). The reasons given by Linn included: 1) Tests and assessments are relatively inexpensive; 2) Testing changes can be implemented relatively quickly; 3) Results are visible and draw media attention; and 4) Testing can create externally mandated changes that would be difficult to legislate.

Cronbach (1963) said that the three different purposes of standardized testing “call for measurement procedures having somewhat different qualities” (p. 677). The discussion of the purposes and techniques of standardized testing continues as a current topic. The next section deals with characteristics of current standardized tests, labeled by Cronbach as focusing on the individual student. During the discussion, we will be reminded of Cronbach's caution that different purposes for tests call for different types of qualities.

Characteristics of Standardized Tests

There has been a history of criticism of standardized tests. Some say external testing minimizes local control or reduces teacher autonomy. Other criticisms have included accusing the evaluators/researchers of being biased or having ulterior motives, etc. Some have pointed out methodological faults with the research.

Airasian and Madaus (1983) raised the issue of using a test with one intended use for another purpose. They discussed the increased use of achievement test results in the policy sphere, with a much less emphasized use of test results by classroom teachers, the original intent of the tests of this era. They concluded that analyses in this area become increasingly pertinent as this shift continues. They said, "When achievement test results are intended to inform extra-classroom policy decisions where there is the potential for serious harm either to individuals or programs, then the problems of linking testing and instruction become acute" (p.103).

Linn (2000) said that the increased popularity of high-stakes accountability systems placed increased technical burdens on the tests used. One example is the choice of constructs; he said that content areas assessed will receive emphasis, while those that are left out will languish. Even within a content area, the emphasis given to subareas matters. Another technical demand is for each successive administration of the test to be with a new, well-equated form of the test. This will counter the complaints about test coaching or narrow test preparation. These technical demands will be demanding and more expensive. Linn said, "We should not expect, however, inexpensive tests designed for other low-stakes purposes to withstand the pressures now being placed on them by high-stakes accountability systems" (p. 12).

Test Design and Use

Typical standardized achievement tests differ in many ways from the teacher-made classroom tests. Although standardized tests consist of items that are judged to reflect important aspects of widely used curriculum materials, they may or may not match the instruction in a particular classroom during a specific period of time. Topics on the test may not have been covered yet or may have received little emphasis relative to other topics.

Airasian and Madaus (1983) have claimed “Standardized achievement tests are designed and constructed to provide assessments of individual pupils. Test items are selected to maximize differences among individual pupils, and a pupil’s performance is interpreted by comparing it to the performance of other pupils” (p. 104). However, Linn (1983) said that standardized tests have seldom been relied upon to provide specific feedback or to flag critically important concepts or skills for students, despite their design intent. They have also not been used by teachers to assign grades. In short, the links between these tests and classroom instruction have often been relatively weak and indirect (p. 180).

By and large, the typical standardized achievement tests have served different functions that are more general and usually a bit vague. For the school or district as a whole, the traditional standardized test provides a general summary of the achievement of its students. By themselves, the scores do not reveal anything about the causes of performance (Linn, 1983, p. 180), but they do provide a general reading of the current status in comparison to the national norm group. Schools or school districts can get an indication of improvement or decline in performance by comparing scores from one year to another.

When the focus of test results shift from the individual student to a group, school, or program level, a different set of issues arises. One basic concern that has been raised about using standardized norm-referenced tests to measure differential school or program effectiveness is whether such tests are sensitive enough to detect school or program related differences when such differences exist? (Airasian & Madaus, 1983, p. 104).

This concern is based on features associated with norm-referenced, standardized achievement tests. One feature centers around individual versus group differences. Norm-referenced, standardized achievement tests are designed and constructed to provide assessments of individual pupils. Test items are selected to maximize differences among individual pupils, and a pupil's performance is interpreted by comparing it to the performance of other pupils. Many of the newer uses of tests, however, require group (not individual) comparison and differentiation. Because educational effects may occur and differ according to the level examined (i.e. individual, class, school), the sensitivity of using a type of test intended to assess performance of one level (individual) for assessing performance of a different level (school, program) has been questioned.

Another key concern centers around content selection. Tests assess "lowest common denominator" to be "fair" across all schools and pupils; content idiosyncrasies are eliminated as much as possible. If a test does not adequately reflect specific content and objectives, it misses the very idiosyncrasies designed to make a school or program unique, and hopefully better.

Other concerns center around the nature of the test items. The form of measurement embodied in tests is one example. In many cases, the way things are tested is not congruent with how they are taught. For example, teachers seldom ask students to select from possible answers, as on a multiple-choice test.

Airasian and Madaus (1983) also expressed a concern about the construct validity of tests. Because of the concerns cited above, they questioned the assumption that standardized tests measuring specific achievement can (and do) separate out non-school or non-program factors (e.g., general pupil ability, social class). Evidence suggests they measure a general factor that is not much different from the general factor measured by many ability tests. Supporting this finding is a technical report detailing analyses of mathematics data from the National Assessment of Educational Progress (NAEP) and from the National Education Longitudinal Study (Muthén et al., 1995). The authors found that the 1992 NAEP Grade 8 algebra items appear not to be sensitive to the completion of an algebra course. They found that, for a given achievement level, algebra students did not have a significantly higher probability than nonalgebra students of giving a correct answer to any of the 28 algebra items (p. 40).

Another concern is the use of summary performance scores rather than more specific information. Airasian and Madaus (1983) have said that using a total score on an achievement test as the dependent variable in studies of school or program effectiveness “hides more than it reveals” about differences in achievement between schools or programs (p. 105).

As we have seen, some of the criticisms of using test performance data center around technical attributes of the tests used to generate the data. Other concerns arise when scores from tests given at *different occasions* are used to measure growth (gains) or when results from *different* tests are compared for accountability purposes.

The National Education Longitudinal Study of 1988 (NELS:88) is a Federal study designed *specifically* to measure progress (growth). By contrast, the TVAAS is a statistical treatment of the results from legislatively imposed tests that were *not* specifically designed to measure growth. A look at the test specifications of the NELS:88 will provide a base for comparison to the TCAP testing and the value-added interpretation of the longitudinal data.

According to the authors of the psychometric report for the NELS:88 (Rock, Pollack, & Quinn, 1995), the test specifications of the longitudinal test battery were dictated by its primary purpose: accurate measurement of the status of **individuals** at a given point in time, as well as their growth over time. The NELS:88 was designed to measure growth between the eighth, tenth, and twelfth grades. Several of the stated objectives for the NELS:88 are pertinent to this discussion. Two of the objectives were that there should be little evidence of floor or ceiling effects and that the reliabilities of the component tests should be psychometrically acceptable for the purpose of measuring individual status as well as growth (p. 3).

In order to achieve the desired reliabilities and to minimize floor or ceiling effects, the authors said that some sort of adaptive testing was necessary. The adaptive procedures for the mathematics and reading tests were necessitated because of the potentially large growth trajectories over a four-year period. One possible solution, which was not used because of testing time constraints, would have been to use lengthy tests at each level, with many very difficult as well as many very easy items.

The adaptive strategy that was selected was a multilevel strategy, in which the eighth grade test served as a pretest for the tenth grade test. Students received a tenth grade test that was either easy, average, or difficult based on their eighth grade score. A similar strategy was followed between the tenth and twelfth grade tests. The multilevel strategy was selected because the same information could be obtained within the time constraints of the testing situation and with a greater degree of accuracy for the high- and low-scoring students.

In contrast, in the TVAAS, there is a less-defined process of sequencing the tests used to measure academic growth (or gains). Although all eighth-grade students take the same eighth-grade TCAP achievement test, the next test in a particular student’s sequence depends on the student’s course-taking sequence. The course-taking sequence is based, presumably, on the student’s abilities and readiness; however, there is a great deal of discretion that enters into the decision. This discretion involves some combination of student/parent choice, recommendations and guidance from school personnel, and course offerings and policies of the school and district.

Another key objective was that the NELS:88 test battery should share sufficient common items both across and within grade level forms, and with the High School & Beyond test battery (a related assessment), to provide articulation of scores for vertical equating and cross-sectional equating (Rock et al., p. 4). There is no evidence of common items between the TCAP achievement tests and the High School Subject Matter tests, which form the basis of the TVAAS analysis.

Another key specification is that the NELS:88 tests should be sufficiently reliable to support change measurement and be characterized by a sufficiently dominant underlying factor to support the Item Response Theory (IRT) model. The fit to the IRT model was necessary to support the vertical equating between retestings as well as linking to such tests as the National Assessment of Educational Progress (NAEP). The IRT vertical equating puts the scores within a given content area on the same scale regardless of the grade in which the score was obtained. (Rock et al., p. 4). There is no evidence of a fit to the IRT model in the high school subject matter tests, used in the TVAAS analysis.

Scale Characteristics

Many types of scales have been devised to measure educational achievement (Yen, 1986). Differences among these different scales and underlying methods have caused educators and researchers to make choices when attempting to measure and report academic growth. Two of the most prominent procedures used to develop standard score scales are the Thurstone method and the (more recent) item response theory (IRT) methods.

The results of the empirical studies (including analyses using the CTB/McGraw-Hill Comprehensive Tests of Basic Skills used in Tennessee) associated with this debate have shown that the nature of the growth models implied by Thurstone and IRT-based scales differs dramatically, especially at the 90th & 10th percentiles (Becker & Forsyth, 1992, pp. 341-2). Becker and Forsyth, who applied Thurstone's absolute scaling method and IRT scalings to the same set of achievement test data, concluded that "the particular scaling method chosen to scale an achievement test will have an impact on interpretations of and expectations for achievement growth" (p. 352).

Linn (1990) focused on the question of the contribution of IRT to the validity of interpretations of achievement test results in the context of the construction of scales for achievement tests. He concluded that, although it has sometimes been claimed that equal interval scales are produced by both Thurstone scaling and IRT scaling, "It is obvious that they cannot both be equal interval when one scale suggests that above-average students tend to develop at a faster rate than below-average students, whereas the other scale suggests just the opposite" (p. 119).

He concluded that the validity of widely accepted inferences concerning growth in student achievement that are dependent on the particular properties of scales has been called into question by the debate about scale properties. He noted, "Increased awareness of the scale dependent nature of certain interpretations is itself a contribution that IRT has made to validity of inferences about achievement that are made from test scores regardless of whether one uses scores based on Thurstone scaling or IRT scaling" (p. 125).

A paper by Baglin (1986) relates an analysis of results on the Metropolitan Achievement Tests for the Rochester City Schools. Results showed "different methods of calculating group scores often produce different results" (p. 57). By "different methods", the author is referring both to *type of score* (raw, percentile, normal curve equivalent (NCE), grade equivalent (GE), and to *measure of central tendency* (mean, median). He concluded:

> The attempt to accomplish both these ends (group scores and continuity) through the same mechanism (the scaled score system) has been responsible for many of the sorts of problems identified in this article. Statistical procedures should not be forced to attempt to accomplish more than that of which they are capable. Psychometrically aware test users are unlikely to regret the elimination of the current brand of scaled score, which allegedly serves both ends but in reality is questionable for each. (p. 67)

Since the era of Baglin's study, much effort has gone into the development of scales--scales with properties that can deliver appropriate cross-sectional data and longitudinal data.

As we have seen, there are studies at the national level that are designed to give a longitudinal (as opposed to cross-sectional) look at student performance. Newer state-level accountability models call for longitudinal data. Willms and Raudenbush (1989) have concluded that a longitudinal design has two chief advantages over cross-sectional designs. One advantage is that it measures the effects of variables describing school policy and practice because it assesses whether changes in school policy affect changes in school performance. Second, there is the advantage that the stable effects of schools can be disentangled from the changing component of their effects (p. 225).

These longitudinal studies call for linking scales, either equated scales or scales that have some degree of comparability. In the next section we will examine the linking of these scales.

Scale Linkage

Equipercentile and linear methods have been used traditionally to equate tests, but item response (latent trait) theory methods have recently been advocated as a potential improvement over traditional methods (Kolen, 1981, p.1). Lord (1977) argued from theoretical considerations that traditional equating methods are not appropriate for equating tests of differing difficulty, whereas item response theory methods have the capacity to provide an appropraite equating in this case.

Mislevy (1994) described the development of "a separate stream of test theory research" concerned with "the analysis of relationships among scores from different tests." This stream of research, which he says is in addition to Classical Test Theory (CTT) and Item Response Theory (IRT), addresses “patterns in correlations among scores of several tests.” He cites factor analysis, structural equations modeling, and multitrait-multimethod analysis as examples of this type of inquiry (p. 13).

As described by Mislevy (1992) and Linn (1993), the central problems of “linking assessments” are determining the relationships between the evidence that two measures give about the performance of interest and interpreting such evidence correctly. Mislevy and Linn defined four types of linking: equating, calibration, projection, and moderation. These approaches are described below, in an ordering from strongest to weakest, in terms of the strength of the link produced.

Equating is the strongest link; it is possible when the two assessments are built to the same specifications. The tests must share the same underlying conception of achievement and matches in content coverage, difficulty, mode of administration, test length, and measurement accuracy at each score point. When tests have been equated, they can be used interchangeably.

Calibration is a somewhat weaker link than equating. Calibration is a process of linking tests that measure the same dimension of achievement but differ in reliability. In this case, the results can be adjusted so that the same score can be expected on both assessments for the same level of achievement. Even equal scores from the two tests, however, cannot be used interchangeably for all purposes.

Projection involves linking scores on tests that measure different dimensions of achievement. Since projection does not meet the more strict assumptions underlying equating and calibration, it is considered a weaker form of linking. Projection uses statistical methodology, often regression, to derive predictions from one set of data about characteristics of another data set distribution. The adequacy of this approach to linking tests depends on the strength of the correlation between the tests involved and on the similarity of the individuals in the sample group (used to estimate the prediction equation) and the individuals in the other group.

Moderation is a process in which scores are aligned from two or more tests that are not assumed to be measuring the same construct. Moderation makes it possible to compare scores on two tests by placing the scores on a common metric; the tests and scores, however, may not be used interchangeably. Moderation is the weakest of the linking techniques.

Those attempting to design longitudinal studies often must make difficult choices to relate tests given at various points in time. Sometimes the tests compared have a common scale. Other studies, such as the TVAAS study of performance on the end-of-course mathematics tests that is a focus in this study, rely on regression analysis to predict performance on later tests (projection). Many other options exist. The choice of model used to examine waves of student data is critical; different models may lead to different conclusions.

Gain Scores

In describing the TVAAS program, Greg Camilli of Rutgers said, “Specifically, it is a ‘gain-oriented’ statistical tool for collecting and analyzing student achievement test score data; that is, gains are the focus rather than absolute levels of achievement” (Camilli, 1996, paragraph 1). Centra and Potter (1980) said, “Gains in student learning are most evident at the point that teachers and students interact, although it should not be concluded that other factors are unimportant" (p. 276). The concept of varying levels of gain scores has been incorporated into Centra and Potter’s model and many others, including the TVAAS model. Although the achievement measure is sometimes referred to as an “effect” (e.g., teacher effect, school effect, and district effect), both formally and informally the “effect” is sometimes called a “gain” score. Even the “value-added” adjective in the name (Tennessee Value-Added Assessment System), shows the focus on improvement (as opposed to absolute levels of achievement).

Thinking concerning the advisability of using gain scores has evolved over time. As a review of literature will show, there was an early reluctance on the part of statisticians and researchers to endorse what seems such an intuitively enlightening concept.

Cronbach and Furby (1970) have listed four purposes for estimating gains or differences: (a) to provide a dependent variable in an experiment, (b) to provide a measure of growth rate or learning rate, (c) to provide an indicator of deviant development, as a basis for identifying individuals to be given a special treatment or to be studied clinically, and (d) to provide an indicator of a construct that is thought to have significance in a certain theoretical network. The indicator may be used as an independent variable, covariate, dependent variable, etc. They concluded that "much of the confusion in the literature arises from a failure to distinguish these purposes and to match distinct methodological recommendations to them" (pp. 77-78).

While they argued that certain statistical methods can minimize statistical weaknesses in approaches of calculating gain scores, Cronbach and Furby (1970) remained unconvinced of the advisability of calculating gain scores to measure improvement. Their concluding statement in a paper that typifies early literature on change or gain scores said, "It appears that investigators who ask questions regarding gain scores would ordinarily be better advised to frame their questions in other ways" (p. 80). Lord has said, “Differences between scores tend to be much more unreliable than the scores themselves”(1956, p. 429).

Titles such as "Gain Scores in Research Can be Highly Reliable" (Zimmerman & Williams, 1982) illustrate a more recent direction in the literature on change or gain scores. In a discussion of the article by Zimmerman and Williams, Rogosa and Willett (1983) expressed their general agreement but added the constraint that the key to understanding the reliability of the difference score is the *amount* of individual differences in true change (pp. 339-340). They concluded:

> Certainly, we do not claim that the difference score has high reliability in all, or even many, practical applications. Instead, we formalize and exposit what common sense dictates: The difference score cannot detect individual differences in change that do not exist, but it will show good reliability when individual differences in true change are appreciable. Although large individual differences in true change may be outside the experience of many researchers, some longitudinal studies do show the requisite increase in variability of test scores over time. (p. 341)

Williamson, Appelbaum, and Epanchin (1991) explained that there has been a focus on the use of growth curves for modeling individual growth in recent literature about the measurement of change. They said this approach has numerous advantages over previous models. “Foremost, it focuses attention an the use of multiwave data for determining characteristics of growth among individuals. The statistical rewards for using multiple occasions of measurement include improved estimation, increased precision, and better reliability” (p. 73).

Schalock et al. (1993) also call for multiple occasions of measurement; they call for “gain score data ... from at least three separate learning (work) samples.” In addition, they call for the gain score data to be “analyzed separately for various groups of learners, for example, for students who enter a course or began a unit of instruction in the upper or lower quartile of goal-related knowledge or skill, or for students enrolled in an entitlement program or who have English as a second language” (p. 124).

Value-added calculations make use of multiple occasions of measurement. Hanushek (1998a) said, “One type of statistical investigation--those employing a value-added specification--is generally regarded as being conceptually superior and likely to provide the most reliable estimates of education production functions” (p. 23). In an earlier work, Hanushek (1979) explained that the advantage of this approach comes from the use of prior achievement to ameliorate any problems arising from missing data about past school and family factors and from differences in innate abilities of students. Proponents of the TVAAS methodology agree on this point; Sanders and Horn (1994) described this concept as follows:

> By focusing on measures of academic gain, each student serves as his or her own ‘control’ - or, in other words, each child can be thought of as a ‘blocking factor’ that enables the estimation of school system, school, and teacher effects on the academic gain with the need for few, if any, of the exogenous variables. (p. 305)

A report from the National Education Longitudinal Study of 1988 (Scott, Rock, Pollack, & Ingels, 1995) calls for a look at another dimension of gain scores. The authors said, "The reporting of the second part of the cognitive growth story, *qualitative* information about the gain scores, is necessary if proper interpretations are to be made when comparing subpopulations" (p. 65). To illustrate they said:

> Comparing groups on raw gains alone can lead to serious misinterpretation in that two or more subpopulations may have the same raw score gains, but the gains may be occurring at quite different points along the scale. For example, one group could be gaining five points at the upper end of the test score scale while another group may be showing the same amount of gain (in terms of raw score points), but at the lower end of the scale. While both groups appear to be growing at the same rate given equivalent raw score gains, they are likely to be exhibiting skill gains in qualitatively different areas. (p. 65)

It is quite possible that this changing of attitudes toward the use of change or gain scores parallels the trend discussed by Mislevy (1994) in his *Test Theory Reconceived*. Mislevy discusses the history of classical test theory (CTT), which he traces to the work of Edgeworth and Spearman around the turn of the 20th century, and the more recent Item Response Theory (IRT). As Mislevy suggested, it is possible that statistical methods have progressed to the point that the conflicts between the CTT and IRT paradigms have been ameliorated. Similarly, it is possible that statistical methods have progressed to the point that earlier reluctance toward gain scores is overcome.

However, there are other issues concerning IRT that serve to lead into a discussion of opportunity-to-learn factors. An observation by Mislevy (1994) illustrates what many consider a weakness of IRT:

> From a cognitive perspective, what makes a task difficult for a particular individual is the match-up between her knowledge structure and the demands of the task.... These match-ups can vary substantially from one person to another for any given task. An IRT item difficulty parameter captures only a population-level characteristic, the relative ordering of items *on the average*. The summaries of the difficulties of items and the proficiencies of persons that the IRT parameters embody miss information to the extent that items are hard for some people and easy for others. (p. 30)

Whether or not it is a weakness, it certainly points out that test scores (a compilation of item responses) can be influenced by factors other than an individual's inherent ability and the difficulty of the test items. If indeed, as mentioned above, an individual's knowledge structure influences the difficulty of test items, then questions relative to the formation of that knowledge structure arise.

The formation of an individual’s knowledge structure is influenced by factors both within and without the formalized, organized system of schooling. Many of the questions relative to the formation of that knowledge structure are grouped into what many call opportunity-to-learn (OTL) issues. Before looking at opportunity-to-learn factors, a look at another category of influences would be appropriate. There are factors that exist outside the school program that have an influence on measures of academic learning.

External Influences

Analysts tend to agree that learning is influenced significantly by factors outside the school. Monk (1990) indicated that a vast array of home and background variables have been used at various times as part of the specification of the inputs of schooling; the use of the variables has not always been accompanied by a strong theoretical rationale for their importance (p. 324). Even when identified, these input variables are difficult to measure.

Poor performance in mathematics and other content areas has been attributed to “a myriad of scapegoats ranging from biological and family influences to unresponsive teaching methods and lack of motivation ” (Signer, Beasley & Bauer, 1997, p. 378). These researchers expressed their belief that “only a few of the factors that potentially influence student achievement are within the control of teachers and administrators” (p. 378).

Haertel, Walberg, and Weinstein (1983) identified nine theoretical constructs that exhibit consistent causal influences on academic learning; they showed that previous models of school learning neglected extramural and social-psychological influences. More recent research has included such variables. In a meta-review and synthesis of research (Wang, Haertel, & Walberg, 1993), the category of variables called *Out-of-School Contextual* variables influenced learning to nearly the same degree as student aptitude and classroom instruction and climate. Of the six broad categories of variables, *Out-of-School Contextual* variables trailed only *Program Design* variables in ratings of importance (p. 35).

The broad category, *Out-of-School Contextual* variables, consisted of four scales: Home Environment/Parental Support, Peer Group, Community Influences, and Out-of-Class Time. The Peer Group scale ranked sixth highest (out of 28 scales); this ranking was caused primarily by the emphasis placed on peers' educational and occupational aspirations (p. 36). Not surprisingly, the scale, Home Environment/Parental Support, was also determined to be relatively influential. Parenting influences have been shown to vary by ethnic group (Okagaki & Frensch, 1998). The benefits of family involvement have been well documented, including the effects of family involvement on improving school attendance and on reducing delinquency, pregnancies, and rate of dropping out (Epstein, 1984; Garnier, Stein, & Jacobs, 1997; Moles, 1982; Peterson, 1989; Walberg, 1984). Unfortunately, there is less reported parent involvement (between parent and school) and less effort (by the school) to reach out to parents in high poverty schools than in low poverty schools (Young & Smith, 1997, p.19).

Willms and Raudenbush (1989) also discussed the ways in which school performance is influenced by “wider social, economic, and political factors.” They cited research supporting the influences of such factors as students’ opportunities for local employment, community support for public education, secular changes in student values, and local deprivation (p. 210).

A RAND Institute study (Grissmer, Kirby, Berends, & Williamson, 1994) found that parents’ education was the most important family characteristic influencing student performance. Young and Smith (1997) pointed out the difficulty in disentangling the separate effects of parents’ education and family income on children’s education. They said, “Independent of income, however, parents' level of education may influence the value that parents place on education, which in turn can influence their children's educational attainment" (p. 10).

Lamdin (1996) said that most student performance studies using the production function do not consider student attendance as an independent variable. He reported that data from Baltimore City public elementary schools indicate that student attendance positively and consistently correlates to student achievement test performance. Willms and Raudenbush state that traditional research shows that academic achievement and levels of truancy have small correlations--less than .3--when the measures are background-adjusted (1989, p. 210). Teachers in high poverty public secondary schools are more likely to report that student absenteeism and tardiness are serious problems in their schools than teachers in low poverty schools (Young & Smith, 1997, p.17).

Researchers have studied the effects of motivation on student achievement. One study (O'Neil, Sugrue, Abedi, Baker, & Golan,1997) summarizes studies on this relationship and says "the observed correlation between motivation and achievement ranges from .12 to about .33 ..., with a maximum of approximately 10% of variance in achievement being explained by motivational factors" (p. 5). This study, analyzing the effect of student motivation on NAEP test performance, reported that the findings indicate "we may be underestimating the achievement of students when we use scores on 'low-stakes' tests as the indicators of performance" (p. xii).

Levin (1993) and others have suggested that most production function studies have seen schooling as something that is done to students, rather than thinking about education as something that students essentially do for themselves. "In the final analysis, learning is something children do (sometimes in schools), not something which schools or teachers do to them.... The variable closest in the causal chain to student achievement is student behavior" (Centra & Potter, 1980, p. 287).

Another complicating factor is that there is often a blurring of what constitutes an external factor and what is under the influence of the school program. Willms (1992) pointed out that a variable like “parental press for academic achievement” is problematic. While many researchers would classify this as an external factor (as defined here), others would recognize that this variable might be affected greatly by a school staff through its homework policy, incentive structure, and pattern of involvement with parents.

Opportunity To Learn (OTL)

In discussing the concept of Opportunity to Learn, Schmidt (1983) shared three situations involving students performance and the match between a test and the instruction delivered/received. In one scenario, students perform well on the test and they did not receive instruction in the area. In this case, either (a) the students had knowledge prior to coming to that class, (b) they had learned it outside school, or (c) learning transfer took place which enabled the students to use knowledge from other content areas. In another scenario, students do poorly and there is a good match between the test and the instruction domain for the students. In this case, the results imply either (a) deficiencies in the instruction received, (b) deficiencies in the students' prerequisite knowledge, (c) lack of ability, (d) lack of motivation, or (e) some combination of these factors. He concluded, “But if the students did poorly and had not been taught the content, the most likely cause of poor performance is lack of opportunities to be taught” (p. 170).

The definition of Opportunity to Learn has evolved over time. For instance, Carroll defined the phrase in 1963 as “the time allowed for learning.” Years later, the National Council on Education Standards and Testing, in its 1992 report *Raising Standards for American Education*, called for the development of “school delivery standards” to help assess a school’s capacity and performance. During the debate around the development of the Clinton administration’s Goals 2000 Act and similar policy discussion by the National Governors’ Association, the terminology shifted from school delivery standards to “opportunity-to-learn” standards (OTL). Porter said “opportunity to learn” describes the “enacted curriculum as experienced by the student" (1995a, p.57). Yoon and Resnick (1998) said that opportunity to learn (OTL) "has come to refer not only to the overlap between what has been taught and what is tested, but to a more proactive concern with providing appropriate learning opportunities for all groups of students" (pp. 2-3).

In discussing the development of the OTL concept, Porter (1995a) said that the greatest emphasis has been placed on the content of instruction--the particular concepts, skills, and applications that are to be taught. But the concept of OTL has also been expanded to include the pedagogical quality of instruction and the resources available to the students and teachers.

In the NAEP report mentioned earlier (Muthén et al., 1995), the authors also discussed ways in which OTL might vary. First, although the OTL effects are the same across subgroups, the starting points might be different. For instance, the criteria or mechanisms for selection into 8th-grade algebra classes might be different; the students may not have the same 7th-grade math ability across subgroups, or may otherwise differ in their preparation prior to their 8th-grade algebra studies. Second, the quality of OTL may differ across subgroups. For example, the starting points may be the same, but OTL factors such as curriculum or quality of instruction may differ. Third, the OTL reporting may differ across subgroups in terms of reliability and validity (p. 2).

Herman and Klein (1997) said, "Although almost everyone seems to agree that OTL indicators should describe the resources, school conditions, curriculum, and instruction to which students have access..., there clearly are many ways to structure these measures" (p. 5). They reported that OTL researchers typically have distinguished three overlapping categories of concern: curriculum content, instructional strategies, and instructional resources (p. 2). The three categories are described in detail with illustrative indicators, with instructional resources including the broad categories of teacher preparation and material resources. Teacher preparation issues in the Herman and Klein case study included education (degree and major/minor), amount and type of teaching experience, participation in relevant inservice activities, and attitudes toward mathematics instruction. Material resources included appropriate instructional materials and suitable tools and supplemental materials.

The framework used in this study was proposed by Allan Odden, of the University of Wisconsin-Madison. Odden is Co-Director of the Finance Center of the Consortium for Policy Research in Education (CPRE). He has expressed his belief that decisions about a set of opportunity-to-learn variables should take a broad rather than a narrow perspective. “The notion is to be as parsimonious as possible in deciding what variables to collect but not to limit the scope of variables as to prematurely eliminate important factors that might be strongly related to student learning” (Odden, 1996, p. 122). He has suggested collecting variables in three specific categories: fiscal, educational process, and teacher quality.

Although others (such as Herman and Klein, above) have proposed alternatives to Odden’s framework, it seems to be an acceptable basis for this study. Porter, in another framework option, said, “School funding is sufficiently important and complicated that it deserves consideration in its own right” (1995a, p. 57). Porter considered adequate funding to be a *prerequisite* to meeting school delivery and opportunity-to-learn standards. Notwithstanding the availability of alternate frameworks, Odden’s proposed framework was chosen for this study. The following discussion of fiscal factors serves as a starting point for a discussion organized by the three categories in Odden’s framework.

Fiscal Factors

Since the 1966 publication of the report entitled *Equality of Educational Opportunity*, commonly known as the Coleman Report (Coleman et al., 1966), policymakers (and educational researchers) have been divided as to the most appropriate courses of action. Those who advocate for the traditional approach to school finance reform call for equity in funding. They aim to reduce disparities in resources available between school districts and, generally, to raise aggregate spending in all school districts in the state. Some even suggest that poor school districts should have greater resources provided than affluent districts, to compensate for other inequalities experienced by students in poor districts.

A study of 1989-90 educational spending (Parrish, Matsumoto, & Fowler, 1995) has given a picture of the state of financial equity in place when Tennessee (and other states) were making policy adjustments. This multivariate study, which made use of adjustments for weighted student needs and “buying power” costs, yielded the following findings (among others):

1. Greater education expenditures per student are associated with higher community socioeconomic status as measured by the value of owner-occupied housing... or by educational attainment. However, this relationship is less pronounced when socioeconomic status is defined in terms of median household income. When the relationship between this variable and education expenditures is considered in isolation, only the wealthiest group is considerably different from the other groups.
2. More money is spent in districts with the highest percentages of minority students compared to districts with the lowest percentages of minority students.... These findings suggest that while inequalities may remain for students in poverty, they do not appear to be driven by minority status.
3. Public education expenditures per student are highest in low poverty districts.... However, this relationship is not linear and affects only the 11% of students in the wealthiest districts. Among the other 89% of students, the variation is only $8 per student.
4. Districts with the highest percentages of students in special education show higher overall actual expenditures than do districts with the lowest percentages of special education students. However, when differences in cost of living and the added cost of serving students with supplemental needs are included, an opposite expenditure pattern is observed. (pp. xvii-xviii)

The study also concludes that the amount of local support for public education rises with the wealth and socioeconomic condition of the community and that state funds are the primary equalizing force in resource allocations (p. 86). Another finding is that lower-wealth districts appear to be investing a larger percentage of their spending on core instruction, rather than on other areas. “Although such findings mitigate total expenditure differences by socioeconomic status, they may also indicate that poor districts are deferring needed school construction, renovation, and the purchase of instructional equipment” (p. xxii).

Other policymakers argue for a “productivity” approach to school finance reform. They are skeptical that significant increases in spending have, or will, result in corresponding gains in student achievement. They advocate targeting increases in funding to techniques and areas of the budget most likely to be productive. Conventional spending patterns and carte blanche increases in revenue are not favored by advocates of this approach. The “productivity” approach demands that new and innovative approaches for linking dollars to achievement must be developed and tested.

Since the publication of the Coleman Report, educational researchers have undertaken many studies to measure the impact of economic inputs on academic achievement. Using a methodology generally known as “production function” research, most of these studies have been designed to measure the amount of each school resource that will maximize student achievement levels and other educational outcomes. The production function approach relies heavily on multiple regression (or other correlation-based) analysis to relate a series of inputs (such as cost factors) to an output (such as student achievement). Levin (1993) defined a production function as “a mathematical expression of the relationship between inputs and outputs in education” (paragraph 3).

Two major studies have tended to dominate recent discussion on the production function studies. Individual studies have incorporated inputs ranging from pure spending measures (such as per pupil expenditures) to the types of services these expenditures buy (such as teacher salaries and teacher-student ratios). An influential review by Hanushek (1989) and a meta-analyses by Hedges and his colleagues (Hedges, Laine, & Greenwald, 1994) identified and considered 38 such studies, conducted between 1967 and 1987, that examine the relationship between economic resources and student achievement. These 38 studies contained a total of 187 estimates of relationships between resources and student achievement, generally expressed in regression equations. The relationships were categorized by Hanushek into seven inputs: (1) teacher/pupil ratio; (2) teacher education; (3) teacher experience; (4) teacher salary; (5) per pupil expenditure; (6) administrative inputs; and (7) facilities.

Hanushek noted that, for each input, from 7% to 29% of the relationships to educational outcomes were positive and statistically significant. For the expenditure measure (per pupil expenditure), 29% of the relationships showed a positive significant relationship. Hanushek's overall conclusion was that "there is no strong or systematic relationship between school expenditures and student performance" (1989, p. 47).

Hedges, Laine, and Greenwald (1994) reviewed the same studies as Hanushek and reached considerably different conclusions. Their analysis suggested that money may in fact be more important in determining how well students are likely to achieve. Their meta-analysis eliminated those studies that had insignificant results and the sign on the coefficient that could not be determined. They then analyzed the remaining studies using statistical techniques other than the "vote-counting" procedure used by Hanushek. They argued that the vote-counting procedure has limited power in finding significant effects and argued that earlier work by Hedges and Olkin (1980) shows that as the number of studies reviewed increases, the probability that a vote count will correctly detect an effect decreases. They concluded that:

> These analyses are persuasive in showing that, with the possible exception of facilities, there is evidence of statistically reliable relations between educational resource inputs and school outcomes, and that there is much more evidence of positive relations than of negative relations between resource inputs and outcomes. (p.11)

Hanushek (1994) then responded to the arguments of Hedges et al., noting that while their meta-analysis provided evidence that relationships between spending and achievement sometimes exist, it still did not constitute evidence of a "strong" or "systematic" relationship, because so many of the studies evinced either no relationship or a negative relationship between the two.

Other work conducted with alternate methodologies like Hierarchical Linear Modeling rather than the "production function" framework used in the economic community often showed positive effects of resources (Wenglinsky, 1998).

Interpretations of such production function research vary and are often used to reinforce already established beliefs. Much of well-publicized findings cited in the policy arena is based on gross oversimplifications. Interest groups on both sides of the issue use the data of Hanushek and Hedges as proof that spending does or does not have a significant impact on school outcomes. "When you compare expenditures and student performance, you don't find a lot of direct effect because of all the complicating factors", says William Cooley of the University of Pittsburg. "It appears to be more a failure of educational research rather than a failure of education" (quoted in Sadowski, 1995, p. 2).

Although most who disagree with Hanushek think his proposals are somewhat harsh and unfeeling and are based too much on the theoretical (as opposed to practical) side of economics, there are critics on the other side as well. Herbert Gintis, Professor of Economics at the University of Massachusetts at Amherst, has called Hanushek's proposals "tentative to the point of impotence in the face of the entrenched interests that benefit the maintenance of the *status quo*" (Gintis, 1995, paragraph 3).

From among the efforts to study production function research, there has been an identification of problems associated with many of these studies. One such identified phenomenon has been labeled a *threshold effect*. This effect is illustrated in situations in which the correlation analysis in production function research studies will not show the effects of small differences in funding. Fortune (1993) said that some larger, aggregate differences in funding, perhaps $600 or $700 per student, are needed in order to demonstrate observable differences in student achievement (paragraph 9).

Another problem associated with the use of the simple, linear correlation method is the absence of information on cost disparities that are essential to demonstrate differences in educational purchasing power. A ranking of districts by amount of instructional expenditures does not necessarily rank the same districts by their educational purchasing power. Arnold, of the Mid-continent Regional Education Laboratory (McREL) said, “For states, the key problem in incorporating a cost-of-living adjustment in the funding formula lies in deciding what factors to adjust for and how much to adjust” (1998, p. 36).

There are those who challenge the correlation-based analyses used in production function studies. O’Neil (1994), in a paper presented to the Mid-South Educational Research Association, said that correlation-based analyses are not accurate reflections of the relationship between educational expenditures and student achievement; the authors offered an alternative methodology based on t-tests and other tests of means differences. The report, based on data from six states involved in equity lawsuits, described results of both production-function analysis and t-test-based methods. The data produced by tests of means difference indicated a significant relationship between instructional expenditures and student achievement, a relationship that production function analysis failed to show. Additionally, it is quite possible that potential non-linear relationships between the input and output variables will not be revealed by the simple, linear correlation method.

Some critics point out problems that can arise in handling shared variance or commonality of explanation in the multiple regression approach. If correlated independent variables are entered in one order, one explanation of credit results; if the variables are entered in a different order, a markedly different explanation would result. Critics of Coleman have showed that his order of effects do not hold up across applications of different regression models (Pedhazur, 1982).

Another design problem is the inclusion of confounded data elements. Much research ignores the fact that many of the expenditures are earmarked. Federal dollars, for instance, are often targeted to programs not related to the specified measure of student achievement (LoVette, 1995; Rothstein & Miles, 1996). Mandated expenditures, say for health or safety concerns, may not impact student achievement in any measurable way. Similarly, overhead expenditures, such as utilities expenses, vary little from district to district. By not focusing on discretionary expenditures, some research underestimates the impact of additional fiscal resources.

The determination and specification of input variables often leads to problems. Problems occur when input measures are chosen that are not related to instruction. A frequent example is the use of teacher salary as an input variable. Teacher salary is based partly on seniority, partly on the training (degree level) of the teacher, and on the specifics of the applicable salary scale. It is quite possible that one district's average teacher salary may be higher than another's, based purely on a factor such as seniority, not on the quality of the teaching force. In addition, the effect of teacher salaries on fiscal needs varies widely among districts. Ferguson, of Harvard's Kennedy School of Government, based on studies of more than 1,000 school districts in Texas and Alabama, said "The 'Does money matter?' question has a lot to do with the ability to attract and retain good teachers" (quoted in Sadowski, 1995, p. 2). He concluded, "Teacher salaries may not matter much in the absolute sense, but they matter more in places where teachers have a lot of competing alternatives" (p. 2).

Another problem that can occur is in the selection of input variables that cannot be measured across all districts to be studied. Very large and very small districts may affect the accuracy of some measures, based on the economies of scale. Funding methods and accounting procedures that vary from state to state compound standardization efforts for possible input measures.

Output variables can also be the source of specification problems. It may be that the output variable being considered for study may have received little emphasis at the schools in the study. It may also be true that the output measure does not apply to the whole student body, although the other measures do. Another problem occurs when the output variables have no relationship to the selected input variable or to school quality.

Another problem in some educational research (such as that by Hanushek) is that of the truncated variable, or attenuation. Using “percent passing” on a test as a measure of achievement limits the variation to a dichotomy rather than variation across the full set of test scores. Variable truncation also occurs when the tests have either floor or ceiling effects, when only one segment of the enrollment is used (such as at-risk students) or when data are not available for the entire sample being analyzed.

Many of the production function studies still being considered in policy deliberations came from a different economic era. Many of the studies reported by Hanushek (including the Coleman Report) range from the late 1950s to the early 1980s. There have been so many developments in the economic and educational policy arenas that many question the applicability of these studies.

Hirth and Mitchell (1995) described two other criticisms of production-function analysis. One problem they cited is that such studies assume that all schools pursue the same goals and that the goals are related to student achievement. Second, production-function analysis inadequately identifies inputs and ignores processes, they said. As an example, they said that curriculum and instruction are overlooked as inputs.

Recent thinking calls for a shift in school finance thinking from equity to adequacy (e.g., Clune, 1994a; Clune, 1994b; Odden, 1998). Defining and determining adequacy seems to call for a better understanding of what helps facilitate learning in students (including fiscal resources).

Fiscal resources alone do not influence learning; however, one way they are often deployed is in the purchase of supplies and equipment for the classroom. A National Science Foundation (NSF) report (Weiss, Lawrenz, & Queitzsch, 1996) stated, “Overall, studies show that science and mathematics classes do not receive adequate support for supplies and equipment" (p. 65). The report cited that, in 1993, about 36% of all science teachers and 27% of all mathematics teachers reported, in a national survey, that a lack of funding for equipment and supplies is one of the most serious problems or barriers they encounter. This problem has grown considerably since 1977, when about 27% of all science teachers and 14% of all mathematics teachers cited this problem as serious. (pp. 65-66)

Student achievement does appear to be related to availability of resources. Average scale scores on the 8^{th} grade National Assessment of Educational Progress (NAEP) 1996 Mathematics Assessment were compared based on teachers' reports on the availability of resources in their general mathematics, pre-algebra, and Algebra classes. In each of the three subjects, average scale scores were highest for classes whose teachers reported having "all the resources I need." In turn, classes with "most" outscored those with "some or none" of the needed resources (Hawkins, Stancavage, & Dossey, 1998, p.66).

The NSF report (Weiss et al., 1996) also noted that the resources that are available "do not appear to be distributed equally across classes." As an example, the authors reported that high school science classes consisting of more than 40 percent minority students are more likely than other science classes to need various types of equipment that are not available (p. 66).

Mosburg (1996), in a Northwest Regional Educational Laboratory report, has made four points which serve as a good summary to this matter. Her first point was, "Money is related to student learning outcomes, but there is no one-to-one correlation" (p. 3). Second was, "School districts spend their money in remarkably consistent proportions, regardless of their wealth. Current spending patterns are inequitable due to large differences in wealth and student need across schools and districts" (p. 13). A third point the author stated was, "A chief way money matters is by maximizing a wide range of opportunities to learn. It is vital to make strategic investments geared to improving students' learning interactions, as well as fund 'the basics,' if more children are to learn and perform at high levels" (p. 17). The final point was, "After three decades of modest equalization efforts by the states, reliance on local funding of schools is currently rising, just as income inequality among community is growing. Poor children are suffering the most from the resulting funding problems" (p. 21).

Levin (1998) stated, "There may be no agreement on the importance of additional funding in promoting achievement, but there is growing acceptance that gross disparities in funding across schools and districts are undesirable" (paragraph 21). He used this point to illustrate his thesis that the role of research and evidence may not reconcile all differences in opinion and value sets, but "we should all be pleased that evidence is seen to be important and at least in principle accepted as a basis for bridging differences" (paragraph 22).

Educational Process

There are many different factors that could be included in any consideration of the educational process. Almost anything that happens in school or is planned by the school could be examined. A few that are pertinent to this study are discussed here. These include students' course-taking patterns, the use of block scheduling, and the related role of time as a factor in learning.

Course-taking Pattern

Course-taking Pattern. One of the most prevalent criticisms of U.S. schools in recent years centers around the U.S. curriculum; specifically the mathematics courses (and to a lesser extent, the science courses) are described as lacking rigor. Alberts, the president of the National Academy of Sciences, said, "The fact is, if you go out and look at the curriculum we're teaching kids in middle school and high school, you'll find it wanting" (cited in Viadero, 1998). Our middle school students are generally still doing arithmetic, while students elsewhere are beginning algebra and geometry.

Analyses related to the National Education Longitudinal Study of 1988 (NELS:88) have added to appreciation of the importance of students' course-taking pattern. One report (Hoffer, Moore, Quinn, & Suter, 1996) stressed the impact of higher-level courses in this finding:

> The strongest effects that emerge from this analysis are those associated with the students' class composition, which we have referred to as their class achievement levels. Here we find that students in higher-level classes learn much more over the two-year period than otherwise comparable students in lower-level classes. The results are strongest in mathematics but are present in science as well. The effects of class achievement-level are only partially explained by the instructional variables measured here. (p. iv)

The authors stated that students in higher-level classes learned more, even when the analysts controlled for the student's initial level of achievement and social background. (p. 105)

Rock and Pollack (1995) reported that when student gains in tested mathematics achievement were cross-classified by grade in school and highest level of mathematics course taken, in their analysis of the National Education Longitudinal Study of 1988 (NELS:88) high school transcript study, the following findings resulted:

> 1. Growth in arithmetic, algebra, and geometry achievement appears to be greater in the first two years of high school than in the last two years for almost all course-taking categories;
> 2. Students who take the more advanced mathematics courses show greater gains, both between 8th and 10th grade, and between 10th and 12th grade;
> 3. Students who do not take advanced courses make greater gains on test items dealing with computational skills, while students in the advanced courses make larger gains on test items requiring conceptual understanding and problem-solving skills. In fact, for these students, significant growth does not occur until they move into the pre-calculus level of coursework. (p. 7)

An analysis of data from the 1990 National Assessment of Educational Progress (NAEP) showed that, on average, students completed three mathematics courses in high school (Davenport et al., 1998). The authors concluded that standards of expectations for students should define not only the number of courses, but the content of these courses as well. They related that higher levels of achievement have generally been associated with coursework at or above the level of Algebra 1. These researchers said that low-scoring minority students have too little exposure to the standard sequence of math courses. They attributed this lack of exposure, at least in part, to a lack of readiness for the content of these courses; they say the solution to this situation lies in a better preparation prior to high school.

As part of an explanation for the variations in students' course-taking patterns and participation in higher-level courses, another report (Hoffer, Rasinski, & Moore, 1995) also noted an association of ethnicity and socio-economic status with participation in higher-level courses. Hoffer et al. noted that, "Historically, some of the largest fault lines in student participation coincided with social background differences, and these often had little to do with ability or motivation to succeed" (p. 2). While recognizing the importance of individual motivation, ability, and parental guidance, these authors stated that some schools have sharply limited opportunities for advanced coursework. Other students may have an interest but be excluded because of school tracking patterns or inappropriate guidance.

It is not clear, however, that policy mandates requiring a greater number of mathematics credits or requiring the completion of higher levels of mathematics courses would increase overall student achievement measures. Most high school students, 64%, complete more mathematics courses than required (Chaney, Burgdorf, & Atash, 1997). Increasing graduation requirements would impact only a minority of students, under current course-taking patterns.

Scores on the NAEP assessment in states requiring 3 or 4 years of mathematics courses reveal a situation similar to the principle of "diminishing returns." Of students assessed in states requiring 3 or 4 years of mathematics, 86% had completed Geometry; their average scale score was 309. In states requiring 2 years or less of mathematics courses, a lower 76% had completed Geometry. These students scored higher, with an average scale score of 312 (Hawkins, Stancavage, & Dossey, 1998, p.50). Similar patterns (higher percentages of students with lower average NAEP scale scores) were reported for second-year Algebra and pre-calculus (p. 49).

Factors other than state requirements affect course-taking patterns. Discretionary decisions in the school programs and the overall climate within a school affect the availability and desirability of more advanced courses. Additionally, Marsh and Yeung (1997) have reported findings that specific components of a student's academic self-concept are stronger influencing factors on subsequent course selection than are previous grades.

Another discretionary decision in the school program centers around the type of schedule used. The time allowed for courses is generally not a student decision nor is it driven by identified student needs. A recent school-level option to meet state time requirements is the block schedule.

Block Schedule . Block scheduling in high schools has received a great deal of attention in recent years. Much of the attention to the concept centers around providing a longer, uninterrupted class period--a particular advantage to laboratory classes and other activities characterized by sustained student involvement. Often block scheduling is cited as a means to encourage changes in instructional practice, less lecturing and more hands-on experiences, for instance. There are administrative advantages centering around the more efficient uses of resources including human resources, textbooks, etc. (Carroll, 1994). Increasing graduation requirements have been instrumental in block scheduling's increased popularity (Rettig & Canady, 1996).

In block scheduling, classes meet for a longer-than-traditional class period but for fewer than the traditional number of days during the school year. The two primary types are "intensive" block scheduling and "alternating-day" block scheduling. The most prevalent form of intensive block scheduling is the "semestered" schedule--typically four courses meeting for 80 to 90 minutes daily for about 90 days. In the alternating-day approach, eight classes meet for the entire school year in a four-course, every-other-day pattern.

Much of the research reported on the topic of block scheduling is anecdotal and/or descriptive. Although there remain many concerns toward the move to block scheduling, many positive results have been reported. Kramer (1996) summarizes research that shows that, under block scheduling, student discipline improves as well as student attitudes toward school. He also reports that semestered and other intensive forms of block scheduling appear to lead to reduced dropout rates (p. 758). Even most critics do not dispute these general statements of positive findings. They do, however, ask for more convincing proof relating to student achievement. The critics acknowledge that student attitudes improve; many conclude students like it better because the course work is not as comprehensive and/or as rigorous. The dropout rate improvement is understandable in that the goal of completing a course (and thus progress toward graduation) is more short-term and opportunities to earn a fresh start come more often.

There is not much recent research that qualifies as conclusive concerning the impact of block scheduling on student achievement. Following is a summary of relevant research that relates to student achievement and type of schedule. It should be noted that student grades (as noted by critics) are problematic as a measure of student achievement, because they are so subjective and vary from school to school, year to year, or even class to class. Grade comparisons for a given student from year to year are comparing different classes (for example, Algebra 1 one year and Geometry the next); course differences may have as much impact as the type of schedule.

The most significant, large-scale research comparing standardized test results under different scheduling options has been conducted in Canada. A number of studies, primarily from the 1980s (prior to the popularity of block scheduling in the U.S.), have investigated the effects of semestered block scheduling in Ontario. Raphael, Wahlstrom, and McLean (1986) compared the performance on the Second International Mathematics Study (SIMS) of semestered and nonsemestered Ontario students who were in Grade 12 or who were mathematics specialists in the college-preparatory Grade 13. The semestered (block) students were outperformed significantly (9 of 11 subtests) by their peers on a traditional schedule.

Raphael and Wahlstrom (1986) report a similar large-scale comparison of achievement in biology, chemistry, and physics. Ontario also participated in the Second International Science Study (SISS); the results from tests of more than 3600 students from 75 schools were analyzed by type of class schedule. While there was no significant difference on scores in physics, students in traditional, year-long biology and chemistry classes outscored those in semestered classes.

Bateson (1990) studied the effects of full-credit semester and all-year timetables on science achievement of Grade-10 students in British Columbia. Students who studied science for the full year performed better on the multiple matrix assessment instruments than those in the semestered (block-scheduled) classes. A further finding is that students who took the full-credit course over the term of the second semester performed better than those taking the same course in the first semester (with testing done at end of second semester.) This lends credence to the criticism that block scheduling leaves gaps in instruction, as students "wait out" a block of time before taking the next course in a given sequence. Foreign language and mathematics courses are cited as examples of such sequential courses.

In another Canadian study (Reid, Hierck, & Veregin, 1994), results from province-wide standardized final exams were analyzed. In particular, one high school compared two years of test results; one year was prior to the implementation of the Horizontal Timetable and one year following the implementation. The Horizontal Timetable was even more intensive than the semestered block schedule. Under this intensive schedule, students would study only two subjects (in large blocks of time) for a ten-week period. Results reported showed a general decline in failure rates on the standardized exams. There was no mention of mean scores or other statistics, leaving the reader to wonder if the positive impacts were spread across the range of achievement levels, or seen only at the lower levels where passing the test was an issue.

The Canadian studies on semestered classes are, by far, the most defensible studies from a research standpoint. In the United States, the timing of block-schedule implementation is several years later than the Canadian initiative. Much of the "evidence" from the U.S. is descriptive, small-scale and suffering from the lack of a strong research design.

There has been debate over the generalizability of the Canadian studies. Cautions on the Canadian studies include questions over the student population sampled (Kramer, 1997), including the age of students and the self-selection of students and their courses. The possible presence of a volunteer effect was also mentioned; schools "elected which timetable to adopt, and it is possible that variables such as prior student performance could have caused them to make the change to a block schedule and in turn could account for the differences reported" (p. 30). Other concerns are that the Canadian efforts to alter the schedule did not include any of the theoretical underpinnings and/or staff development initiatives that are considered crucial in U.S. implementation efforts.

An additional general concern is that the timing of an assessment may be crucial and potentially reflect poorly on block scheduling. As an example, consider tests given ten days from the end of a school year; the ten remaining days would, by design, have more content planned in the 10 longer class periods. The traditionally scheduled classes would have covered a higher percentage of the course's curriculum.

In two Dothan, Alabama high schools, no significant difference was found on measures of mathematics achievement between two random samples of algebra and geometry students (Lockwood, 1995). The study used a standardized national test, carefully designed and controlled student assignment procedures, and statistical controls for other factors. Lockwood concluded that the semesterized block schedule is a "viable option" for consideration for high schools which can be implemented "with no overall decline in student achievement" (p. 108-109).

One recent study which was carefully designed and national in scope was done by the College Board (1998). The study used PSAT/NMSQT scores as a covariate to ensure that any differences in student performance were primarily related to the schedule and not to existing differences in the ability level of students in a specific schedule. The Board reported that approximately seven percent of the Advanced Placement (AP) high schools had adopted a semesterized block schedule in 1997 (p. 1). The four highest volume AP examinations in 1997 (U.S. History, English Literature, Biology, and Calculus AB) were studied.

The College Board (1998) study concluded:

> The instructional schedule does affect student grades on the four AP Examinations in this study after they have been adjusted for group differences in student ability. The results from this study generally suggest that students, on average, obtain higher AP grades when instruction is given over an entire year rather than in a semesterized block schedule format. These results are consistent across four AP Examinations and are found on 15 of 16 comparisons between year-long and semester block courses. (pp. 8-9)

The report also found some evidence that higher AP grades may be obtained when testing immediately follows instruction (Fall or Spring block). A finding that is pertinent to a discussion of time factors is that students tend to obtain higher AP grades when more time is devoted to instruction, in both year-long and semesterized block courses (p. 10).

Carroll (1994), a leading proponent of block scheduling, says the Copernican plan (his name for a version of block scheduling) uses a change in scheduling not as an *end*; "rather it is a means to several important ends. The most important of these are to improve vastly the relationships between teachers and students and to provide teachers and students with more manageable workloads" (p. 27). Block scheduling is intended to provide a more efficient and effective use of time.

Time Factors . Proponents of block scheduling say the longer class periods allow for a higher degree of student involvement, which, in turn, will positively influence student learning. Stohr-Hunt (1996) reports on an analysis of the National Educational Longitudinal Study: 1988 (NELS:88) that supports this argument. This study, analyzing the relation between “hands-on” science time and science achievement, showed there was a significant, positive main effect at the school level and at the student level.

A comprehensive “meta-review” and synthesis of research on variables related to learning (Wang, Haertel, & Walberg, 1993) supports the idea that what happens in the class is important, not merely the length of the class period. The authors of this report, when analyzing factors within the Social and the Classroom Climate categories, conclude, “Together, the highly rated items in these two scales characterize a classroom in which the teacher and students interact considerably and cooperatively, where students work with several classmates, share common interests and values, and pursue cooperative goals” (p. 36). Buckman, King, and Ryan (1995) explained that two Orlando high schools restructured after studying flexible schedules used across the country. They said:

> In these schools teachers have reported major differences in presentation and planning and a more active role of students in learning. The flexible schedule allowed teachers to use cooperative learning, integrated curriculum, and multiple-intelligence instruction, all innovations that support constructivist and brain-based learning theories. (p. 11)

The *amount* of time spent in instructional activities, as well as the *quality*, has been shown to be important in influencing learning. The Wang et al. report rated the scale Quantity of Instruction as third most influential of the 30 scales studied.

A student’s opportunity to learn is decreased when the amount of time spent in the classroom is lessened by the student’s absence from school, arriving late, or cutting class. Also, when students disrupt classes by being late or absent, they interfere with lessons in progress and with other students’ opportunities to learn. On a typical day in the 1990-91 school year, 10% of the students in high poverty secondary schools are reported absent; this compares to 7% in low poverty schools (Young & Smith, 1997, p.18).

Because time is obviously essential for learning, and because educators view time as something they can manipulate, much has been written about the subject. There are two related aspects of the subject–time allocation, as in block vs. yearlong scheduling, and time usage, as in time on task, academic learning time, engaged time, etc. Time allocation is relatively easy to measure; however, much needs to be studied about the valuable time that students spend interacting with the teacher or engaged in activities designed by the teacher. The role of the teacher in designing and facilitating high-quality learning activities is becoming increasingly valued.

Teacher Quality

In testimony before a U.S. House of Representatives subcommittee, Hanushek (1998b) said, "Considerable evidence shows that by far the largest differences in the impact of schools on student achievement relate to differences in the quality of teachers." His testimony illustrated the importance of quality teachers (and the overlapping of policy issues) when he continued, "Whether or not large-scale reductions in class sizes help or hurt will depend mostly on whether or not any new teachers are better or worse than the existing teachers." Monk and Rice (1999) reason that "any satisfactory attempt to grapple with the resource implications of high performance standards will need to deal explicitly with the existing knowledge about the available indicators of teacher quality and learning outcomes for students" (p. 116). They based this conclusion on the fact that (a) teacher resources represent a large proportion of the total resources committed to education, and consequently can have a disproportionate effect on the productivity of the enterprise, and (b) several of the elements affecting teacher quality, such as preservice teacher preparation and certification requirements, are particularly interesting from a policy perspective.

In addition to the importance of the huge resource investment in the teaching staff, there is also a huge importance on the outcomes side. Hanushek (1998a) said that his previous research showed that the differences in student achievement with a good versus a bad teacher could be more than 1½ grade levels of achievement within a single year. His work has also shown that variations in total teacher differences are much more significant in explaining variation in student achievement than other reform elements such as reducing class size (p. 35). Recent studies of teacher effects at the classroom level using the TVAAS and a similar database in Dallas have found that differential teacher effectiveness is a strong determinant of differences in student learning, far outweighing other factors such as class size (Sanders & Rivers, 1996; Wright, Horn, & Sanders, 1997).

Teacher Qualifications. A recent study has summarized previous and recent research concerning the ways in which teacher qualifications and other school inputs are related to student achievement across states. One finding was that among the variables assessing teacher quality, the percentage of teachers with full certification and a major in the field is a more powerful predictor of student achievement than teachers' education level. Measures of certification status are also strong correlates of student achievement (Darling-Hammond, 1999). The study also found:

> When aggregated at the state level, teacher quality variables appear to be more strongly related to student achievement than class sizes, overall spending levels, teacher salary levels (at least when unadjusted for cost of living differentials), or such factors as the statewide proportion of staff who are teachers. (p. 38)

The National Science Foundation reported "Proxy measures, such as an evaluation of undergraduate or graduate major or number of courses completed in the field of assignment, are one way to gauge how well teachers understand science and mathematics" (Weiss, Lawrenz, & Queitzsch, 1996, p. 50). According to the NSF, in 1993, less than 5% of elementary school science or mathematics teachers had majored in science or science education or mathematics or mathematics education at either the undergraduate or graduate level (p. 50). This figure is not considered surprising, given that most elementary teachers teach all or most academic subjects, rather than specialize in science or mathematics.

The NSF found that science and mathematics teachers in grades 9-12 are more likely to have majored in science or mathematics at the undergraduate or graduate levels than their elementary counterparts. However, nearly 30% of high school science teachers and 40% of high school mathematics teachers had neither an undergraduate nor a graduate major in science or science education or mathematics or mathematics education. Moreover, although more than 90% of high school science teachers had at least a minor in science or science education, only 81% of high school mathematics teachers had at least a minor in mathematics or mathematics education (Weiss et al., p.50)

Teacher skills and abilities influence student achievement. Greenwald, Hedges, and Laine (1996) reviewed a number of studies of the relationship between school inputs and student outcomes. School resource variables that described teacher ability, teacher education, and teacher experience were strongly related to student achievement. On the other hand, Hanushek's (1996) synthesis of research studies found mixed support for a relationship between school resources and achievement. Although his analysis did not detect a clear overall pattern, measures of teacher experience were more consistently related to achievement than measures of teacher education. Ashton (1996) noted that teachers with regular state certification receive higher supervisor ratings and student achievement than teachers who do not meet standards. Teachers without preparation have trouble anticipating and overcoming barriers to student learning and are likely to hold low expectations for low-income children. Ashton suggested that states that reduce certification requirements or permit the hiring of teachers who do not meet certification standards worsen inequities in the quality of education offered to low-income students.

Findings from the National Assessment of Educational Progress (NAEP) showed that, at Grade 8, students of teachers with a college major in mathematics outperformed students whose teachers had a college major in education or a field other than education, mathematics education, or mathematics. The experience of the teacher also was shown in this report to be significant. Students taught mathematics by teachers with more than five years of teaching experience were more likely to perform better than students taught by teachers with five or fewer years of experience (Hawkins, Stancavage, & Dossey, 1998, p.3). The same report showed Tennessee in the group of states and territories labeled “below the national average” in grade 8 teachers with a major in mathematics, with only Delaware, Guam, and Louisiana having lower percentages (p. 15). A similar low ranking for Tennessee was reported in the Grade 8 category “percentage of students whose teachers have mathematics teaching certificates” (p. 21).

In explaining the low performance of Grade 12 students from the United States on the Third International Mathematics and Science Study (TIMSS), Secretary of Education Richard W. Riley referenced the training of the teachers in U.S. high schools. Among other factors, he noted that 28% of high school math teachers and 55% of physics teachers had neither majors nor minors in their subjects (Viadero, 1998).

The equity in the distribution of teacher quality has been questioned. "Students in mathematics classes in low poverty public secondary schools are more likely to be taught by teachers who majored or minored in mathematics than were students in high poverty public secondary schools” (Young & Smith, 1997, p.23). The authors of this NCES report cited a 9-percentage point difference, 83.3% vs. 74.1%, between high and low poverty schools in the percentage of students taught mathematics by a teacher with a mathematics major or minor.

One policy concern is the equity issue concerning the distribution of capable, qualified teachers. Other issues concern the recruiting of sufficient numbers of teachers with quality credentials in an era of an aging teacher workforce and the professional development of the teachers already in our schools.

Professional Development. The National Education Commission on Time and Learning, among others, emphasizes the role of on-going professional development as a response to higher student expectations and emerging curriculum and standards implementations. The Commission reported in *Prisoners of Time*:

> One point cannot be restated too forcefully: professional development needs will be broad and massive. Indispensable to educated students are learned teachers in the classroom. An enormous change is at hand for the nation's 2.75 million teachers. To keep pace with changing content standards, teachers will need ongoing coursework in their disciplines *while they continue to teach their subjects*. (1994, p. 21)

The Commission's report pointed out that, in other countries such as Japan and Germany, there is a structure allowing much more time for teachers' inservice training and for collaboration with other (often more experienced) colleagues (p. 27).

Although 96% of public school teachers reported participating in some sort of professional development activity in 1993-94 (National Center for Educational Statistics, 1998b), there is a question whether such activities are meaningful or productive. Only 30% participated in professional development that involved in-depth study in a specific field, and only 15% received nine hours or more of this type of training (Darling-Hammond, 1997).

In a discussion of how to create school finance that facilitates new educational goals, Odden (1998) listed three district-level goals. Two of the three goals relate to teachers' professional development. One goal was that school sites should be provided with greater control of their resources. The other two goals, the ones that relate to professional development, are to reinvent teacher compensation and to provide school-based performance incentives. Professional development is linked to this vision of compensation in that salary increases would be based on knowledge, skills, and competencies teachers develop in the required school restructuring and the teaching of a more rigorous curriculum (p. 5). Often school-based performance incentive programs *encourage* but do not *enable* schools to improve the performance of all students.

Fuhrman (1999) said, “We have already seen that the mere imposition of a new accountability system–even ones...where the goals are thought to be clear and outcomes both sure and motivating–does not unleash some hidden capacity” (p. 9). Odden reported on research that shows that teachers and principals view the system’s setting performance improvement targets as a legitimate management strategy. Further, given appropriate assistance and *additional training* (emphasis added), most teachers believe they can produce improved student performance (p. 5). Schools in Kentucky that made significant improvements, moving from decline status to reward schools, attributed their success to the assistance of a Distinguished Educator who provided technical assistance (Kelley, 1998).

Professional Status. The report *Prisoners of Time* also stated that, in both Japan and Germany, “The Commission sensed considerably greater encouragement of teacher professionalism than is apparent in the United States” (p. 27). Ingersoll (1999) said that teachers in the U.S. are treated as semi-skilled workers and that teaching is largely treated as second-class work (p. 33). He said that because of the continued treatment and status of teaching as a *semi-profession*, “Teaching is plagued by problems of both recruitment and retention, and out-of-field teaching is not simply an emergency condition, but a common practice in this country” (p. 34).

One definition of teacher professionalism lists defining concepts (Schalock, Cowart, & Staebler, 1993). Among these defining concepts is “the engagement in continued growth and development as a professional” (as discussed earlier). Other such concepts are “working with colleagues, administrators, and others to improve one’s school as a workplace for students and teachers” and “the contribution of valued services to the community and profession.” Also listed are “reflection on one’s own effectiveness/productivity as a teacher, and improving it when necessary” and “the reflection of integrity and high ethical standards in all interactions with others” (p. 185).

In Tennessee, the Career Ladder program has been implemented as the state's version of several (real and proposed) incentive programs in which high-status professional educators are recognized and compensated with extra pay. Participation in the Career Ladder program's upper levels (II and III) is voluntary for eligible teachers. Tennessee's teachers identified as Level II or III have undergone an extensive process of self-report, observations, and structured interviews to formulate evidence of professionalism and productivity.

The 1995-96 *21st Century Schools Program Report Card* said that the Career Ladder system "is intended to reward our best public school teachers" and that "evaluations are based on teaching skills, knowledge, and service to students" (Tennessee Department of Education, 1996, p. 111). The report said that "participation by those eligible for Career Ladder II and III certificates has remained fairly stable for the last several years and now stands at 23%" (p. 111). Participation rates in Tennessee districts range from less than 10% to greater than 40%.

This range in participation rates may be due to a number of reasons. One might be the prevailing culture of the organization, either encouraging or discouraging a teacher's involvement in programs such as the Career Ladder program. Another might be a teacher's pattern of motivation. Odden and Kelly (1997) discussed the intrinsic motivation inherent in improved student achievement. Johnson (1986), Goodlad (1984), and Guskey and Passaro (1994) discussed the importance of professional efficacy. Extrinsic rewards such as salary, status, and promotional opportunities are also important motivators. Fuhrman (1999) said, "Our conclusion is that, contrary to the views offered by some commentators, money bonuses are valued by teachers and can be motivating" (p. 7).

<u>Teacher Salaries</u>. Teacher salaries are an important measure for the ability of schools to attract and retain high quality teachers. Odden and Kelly (1997) said, "Overall salary levels matter to teachers: lower salaries are associated with a decline in the quality of people attracted

to teaching, loss of teachers to higher paying districts (states) and loss of teachers to other professions” (p. 75).

In the 1993-94 school year, public school teachers in low poverty schools earned 28% more in total school earnings than their counterparts in high poverty schools (Young & Smith, 1997, p.25). This finding and similar data based on averages often overlook the fact that more than the official salary scale is involved when computing average teacher salaries. Two districts with identical salary scales would have different average teacher salaries if one district had, on average, teachers with more experience or with higher levels of training. Most salary scales compensate provide increased compensation based on levels of experience and training. Lower average salaries in high poverty schools could be the result of a lower salary scale or the result of the inability or unwillingness to attract and retain a more experienced, better-trained teaching force to these schools.

Higher salaries should not be viewed as motivators by themselves, in the view of Odden and Kelly (1997), but “rather as part of an overall set of education system strategies designed as a whole to advance teacher expertise and to educate students to high standards” (p. 75).

Summary

In this chapter a review of the literature concerning three specific categories of the Opportunity-to-Learn concept has been presented. These OTL categories include fiscal factors, educational process, and teacher quality. Additional external factors, or factors generally felt to be outside the realm of influence of the school, have been discussed. Accountability and measurement issues have also been addressed. These discussions are presented to frame the study of a policy initiative in the state of Tennessee--the use of end-of-course tests to provide accountability (for teachers, schools, school systems and, in the future, students).

In a discussion of how the new accountability systems (like Tennessee’s TVAAS) work, Fuhrman (1999) said:

> New accountability systems focus attention on performance, seeking to inspire school personnel to attend more to student achievement. What we don't yet know is the quality and duration of the instructional changes that occur in responding schools. We don't know whether such systems cause teachers and administrators to attend to the kind of capacity-building that will enable them to improve instruction and learning over time. (p. 6)

Fuhrman also said, "The key question for future research on the effects of external accountability systems is whether schools can, or will, respond by developing congruent internal expectations and accountability systems. And, perhaps more important, a related question is how these schools develop the capacity to generate these new internal norms and processes" (p. 8).

Tennessee has invested considerable time, energy, and resources in the development and refinement of an accountability system. With such a sizeable investment, it would seem a disservice to the plan if certain aspects of the policy were put in effect and other potential benefits not developed. It is not enough to focus attention and motivation for increased student achievement; it would be an important positive additional benefit if the accountability system contributed to the capacity-building ability of the schools and districts. Fuhrman (1999) said that, without explicit attention to capacity, new accountability systems are "insufficient approaches to improving student achievement" (p. 10).

In order for Tennessee's accountability system to add to the schools' ability to make the appropriate improvements, at least two things need to occur. One is that the accountability system should be understood by those most directly affected. The most technical aspects can be verified elsewhere, but the goals, performance standards, and methodology should be understood by teachers, administrators, parents and students and be congruent with school efforts. Second and most important, those on the "front lines" must get meaningful feedback and relevant research findings. As a companion to the generating of summary-type scores, the system should give insight into what works and what factors have potential to result in improvement. As Linn

(2000) said, “It is important to identify features of assessment and accountability systems that both influence the trustworthiness of the information provided and the likely impact of the systems on educational practices and student learning” (p. 4).

This study is a step in the direction of adding to the capacity of local improvement efforts. By looking at a variety of factors that influence learning, this study looks to find patterns and relationships from among existing measures and reports. If existing reports and measures do not reveal significant, useful research findings, then potential modifications will be explored. Chapter 3 outlines the procedures utilized in this study of existing measures and reports. Chapter 4 provides a description and interpretation of the results of the statistical analyses on the available measures described in this chapter. In Chapter 5 potential modifications of data gathering and analysis will be presented, as part of the conclusions from the study.

CHAPTER 3

RESEARCH METHODOLOGY

The purpose of this chapter is to present the methods and procedures used to study student achievement on mathematics end-of-course tests in the context of an Opportunity-To-Learn (OTL) framework of variables. In this chapter, the following elements of the research methodology will be discussed: data collection, research design, variables, hypotheses, and data analysis.

Data Collection

The data set used for this study was merged from subsets of three different Tennessee-specific data sets compiled at the state level. These data sets, compiled by the Tennessee Department of Education, include the 1995-96 end-of-course test results, the 1995-96 Report Card, and the Annual Statistical Report. After the data sets were merged, a subset of the cases (school systems) was determined. Of the 139 school systems in Tennessee in the 1995-96 school year, the 65 that had exactly one high school were selected for this study.

As a point of reference, there were 24 districts with two high schools, 13 districts with three, and 8 districts with four high schools. Larger districts include two districts with five, four districts with six, and one with seven high schools. The three largest school systems contained 11, 12, and 29 high schools. Also, some districts had no end-of-course test results because they were elementary only or some other exception to the traditional K-12 pattern.

The justification for this selection was that most of the accountability data were generated at the district level. District-level data would be problematic in districts with multiple high schools. It would be unrealistic to expect and impossible to prove that each of the high schools in such a district would be consistent in each of the demographic, fiscal, teacher, and other factors.

By selecting districts with exactly one high school, district-level information can more appropriately be included in the study.

Research Design

The research design for this study is based on correlation and multiple regression analysis. This approach (multiple regression) is versatile and appropriate for the analysis of diverse designs (Pedhazur, 1982). According to Pedhazur, in multiple regression "the overriding conception is that information from independent variables (continuous, categorical, or combinations of both types of variables) is brought to bear in attempts to explain variability of a dependent variable" (p. 7). The general approach incorporates the Opportunity-to Learn framework (discussed earlier) proposed by Odden (1996) and includes other, external influences. In general, student achievement is modeled as a function of several influences. These include fiscal characteristics of the district *(F),* characteristics of the educational process *(Ed),* qualifications of the teacher *(TQ),* and external influences *(Ext)* such as those of the community, peers, etc. That is,

$$A = f(F, Ed, TQ, Ext).$$

In their discussion of the types and quality of data needed for current cost-productivity research, Monk and Rice (1999) said that "Value-added test score measures may be the best that can be expected in the near term" (p. 135).

Variables

The variables included in this study are listed and described in the following tables. Two variables were selected from the 1995-96 Report Card data set to represent external factors, which are given in Table 3-1. These two factors are included to determine if student-population characteristics (not under the direct control of the school) have an influence on the dependent variables.

Table 3-1 Independent variables - external factors

Variable - Short Name	Variable - Long Name	Description	Source
FREEREDL	Free & Reduced Lunch	Percent of students participating in Free & Reduced Lunch program in 1995-96	21st Century Schools Report Card
SPED	Special Education	Percent of district's Average Daily Membership receiving Special Education services in 1995-96	21st Century Schools Report Card

The effects of fiscal factors *(F)* were explored through four relevant variables (see Table 3-2) obtained from the 1995-96 21st Century Schools Report Card. The Per Pupil Expenditure factor is a traditional measure, used in many studies of school finance. The next two factors,

Table 3-2 Independent variables - fiscal factors

Variable - Short Name	Variable - Long Name	Description	Source
PPEXP	Per Pupil Expenditure	Expenditure (in $) per student in 1995-96	21st Century Schools Report Card or Annual Statistical Report
ADMNSPND	Administrative spending	Percent of expenditures spend in Administrative category	21st Century Schools Report Card
RGEDSPND	Regular Education spending	Percent of expenditures spend in Regular Education category	21st Century Schools Report Card
LOCALREV	Local Revenue	Percent of district's total revenue from local sources	21st Century Schools Report Card

administrative spending and regular education spending, explore the school systems' allocation of available resources. Specifically, they reflect the percentage of expenditures allocated to administrative functions and to the regular education components of the district (as compared to special education, support services, extra-curricular activities, etc.). The remaining fiscal factor, local revenue, gives insight into both the capacity and *willingness* of the local community to contribute local resources for educational purposes. By design, Tennessee's funding formula forces local entities to contribute according to the local revenue-generating *capacity*. More affluent communities receive lower levels of state funding and are required to contribute more local (county and/or city) funds. Some communities also contribute local funds above required levels, reflecting a high local *willingness* to fund their schools.

The effects of factors pertaining to the educational process *(Ed)* were explored through 11 variables obtained (or calculated) from three of the data sets and from relevant literature (see Table 3-3). The attendance rate and dropout rate are key traditional measures of student involvement and schools' programs and policies. It could be argued that the attendance rate and the dropout rate are characteristics of the students and reflect the values of the students' social environment and family situation; under this scenario these measures would logically fit under the External Factors category described earlier. These two variables are included in this study in the category of educational process variables because, in most current accountability models, schools are measured on these factors. Under this policy logic, the schools are held accountable for enforcing mandatory attendance laws, designing school programs that meet the needs of the students, offering motivating environments for the students, and overcoming barriers to student success.

The district's eighth-grade achievement test math scores, the school's ACT Math test scores, and Competency test scores are included to give a more comprehensive view of the district's overall achievement level in mathematics.

Table 3-3 Independent variables - educational process factors

Variable - Short Name	Variable - Long Name	Description	Source
ATNDRATE	Attendance Rate	District's attendance rate (%) in Grades 7-12	21st Century Schools Report Card
DROUTRT	Dropout Rate	1995-96 one-year event rate of District's dropouts	21st Century Schools Report Card
MTHGR8NP	Math Gr.8 National Percentile	District's 1995-96 Median National Percentile score on TCAP Achievement test (Grade 8)	21st Century Schools Report Card
ACTMATH	ACT Math	District's graduates' Mean score on ACT test	21st Century Schools Report Card
CTMPCTP	Competency Test - percent passing	Percent of District's students passing the TCAP Competency Test (Mathematics section) in first year attempted (9th Grade)	21st Century Schools Report Card
PA_MPS	Pre-algebra Mean Predicted Score	Mean score of School's Pre-algebra students' predicted (via regression) scores, based on previous math achievement scores	1995-96 End-of-Course test results
A1_MPS	Algebra 1 Mean Predicted Score	Mean score of School's Algebra 1 students' predicted (via regression) scores, based on previous math achievement scores	1995-96 End-of-Course test results
A2_MPS	Algebra 2 Mean Predicted Score	Mean score of School's Algebra 2 students' predicted (via regression) scores, based on previous math achievement scores	1995-96 End-of-Course test results
G_MPS	Geometry Mean Predicted Score	Mean score of School's Geometry students' predicted (via regression) scores, based on previous math achievement scores	1995-96 End-of-Course test results
CTFAC	Course-taking factor	Factor showing relative distribution of test-takers in higher vs. lower mathematics courses	Calculated (for this study). ctfac = (# in A2 & Geom.) / (# in PA & A1)
BLSCH	Block Schedule School	Value Label 0 = no 1 = yes	Literature review (Univ. Of Memphis Report)

The four variables involving predicted scores are a product of the Tennessee Value-Added Assessment System (TVAAS). Conceptually, each student who completes one of the mathematics courses with an end-of-course test has a predicted score calculated by the TVAAS regression analysis. The mean of the participating students' predicted scores becomes the group's mean predicted score and serves as a comparison for the group's actual mean score. For this study, these four variables and the calculated course-taking factor, were included to gain insight into whether or not there is an optimum level of readiness for the various courses to maximize student achievement. Should schools structure their program for relatively small groups of well prepared students? Or, should they encourage more students into higher levels of courses, with a lower average level of preparation?

The remaining variable in the educational process category identifies those schools using a block schedule. The schools were identified from a report from a University of Memphis study (Smith & McNelis, 1996).

The effects of factors pertaining to teacher quality *(TQ)* were explored through six variables obtained from the 1995-96 Report Card and the Annual Statistical Report (see Table 3-4). There are two variables dealing with the participation of the school system's teachers in Tennessee's Career Ladder program. The variable dealing with eligibility in the program is a default measure of teacher experience; to be eligible for the upper Career Ladder levels one must have been a teacher for at least seven years. The Career Ladder Percent variable specifies the percent of a district's eligible teachers who successfully completed the assessment process for certification as a Career Level II or III teacher.

Three variables describe the degree levels attained by the district's teachers as part of their professional training. Another variable includes the average teacher's salary in the study; this variable separates the classroom teachers' salaries from administrators and other licensed employees. The average teacher's salary variable is influenced by the experience and degree levels of the teachers and by the local salary scale.

Table 3-4 Independent variables - teacher factors

Variable - Short Name	Variable - Long Name	Description	Source
CLPCT	Career Ladder Percent	Percent of eligible classroom teachers attaining Career Ladder Level II or III	21st Century Schools Report Card
CLEPCT	Career Ladder Eligible Percent	Percent of classroom teachers eligible for Career Ladder Level II or III	Calculated from Annual Statistical Report and 21st Century Schools Report Card data
BSPLUS	Bachelors plus	Percent of classroom teachers with training higher than a Bachelors degree	Calculated from Annual Statistical Report data
MSDEG	Masters degree	Percent of classroom teachers with a Masters degree (but no higher)	Calculated from Annual Statistical Report data
MSPLUS	Masters plus	Percent of classroom teachers with training higher than a Masters degree	Calculated from Annual Statistical Report data
AVGTS	Average Teacher's Salary	Average salary for classroom teachers (not all licensed personnel)	21st Century Schools Report Card or Annual Statistical Report

The eight dependent variables (described in Table 3-5) are of two types. One set of four pertains to the school effect. The school effect is the difference between the mean of the *predicted* scores and the mean of the *actual* scores of each of the mathematics courses. The predicted scores were calculated by regression analysis as part of the TVAAS methodology, using students' previous mathematics achievement as predictor variables. The other four, denoted with the phrase Mean Scale Score, reported the mean of the *actual* scores on the four end-of-course mathematics tests.

Data Analysis

The compiled data set for this study was imported into the statistics computer software program, Statistical Package for the Social Sciences (SPSS) 7.5 for Windows, for analysis. As a first step, descriptive statistics for each variable were computed and reported to provide a useful profile

Table 3-5 Dependent variables

Variable - Short Name	Variable - Long Name	Description	Source
PA_MSS	Pre-algebra Mean Scale Score	Mean score of School's Pre-algebra students' actual scale scores	1995-96 End-of-Course test results
PA_SCHEF	Pre-Algebra School Effect	Value-Added Pre-Algebra School Effect (difference between actual mean score and predicted mean score)	1995-96 End-of-Course test results
A1_MSS	Algebra 1 Mean Scale Score	Mean score of School's Algebra 1 students' actual scale scores	1995-96 End-of-Course test results
A1_SCHEF	Algebra 1 School Effect	Value-Added Algebra 1 School Effect (difference between actual mean score and predicted mean score)	1995-96 End-of-Course test results
A2_MSS	Algebra 2 Mean Scale Score	Mean score of School's Algebra 2 students' actual scale scores	1995-96 End-of-Course test results
A2_SCHEF	Algebra 2 School Effect	Value-Added Algebra 2 School Effect (difference between actual mean score and predicted mean score)	1995-96 End-of-Course test results
G_MSS	Geometry Mean Scale Score	Mean score of School's Geometry students' actual scale scores	1995-96 End-of-Course test results
G_SCHEFF	Geometry School Effect	Value-Added Geometry School Effect (difference between actual mean score and predicted mean score)	1995-96 End-of-Course test results

of the data.

To test the hypotheses and explore the relationships among the variables, a correlational analysis was used. The variables were grouped into the broad categories described in Tables 1-5 for the correlation analysis; this allowed the related variables to be studied in smaller sets and to determine which variables had the most promise to yield insight in the subsequent regression analysis.

The multiple regression analysis was performed with the variables introduced using the variety of procedures included in SPSS. The eight dependent variables included the four variables at the school level reporting the value-added *effect* for each of the mathematics subjects and the four variables reporting mean scale scores for each of the mathematics subjects. The independent variables selected for the regression analysis were chosen from among the larger group of independent variables included in the correlational analysis. Those independent variables "showing the most promise" in each of the categories were introduced into the eight regression equations.

Hypothesis Testing

Several hypotheses were tested to explore each of the four general research questions discussed in Chapter 1. The hypotheses are expressed in the null. In each case, each pertinent independent variable was compared to each of the eight dependent variables and a bivariate correlational statistic computed. A two-tailed test of significance was applied to each bivariate relationship. Additionally, in the single case of a dichotomous independent variable (identifying whether or not a given school utilized block scheduling), analysis of variance (ANOVA) was utilized to test the hypothesis.

Research Question 1. To address the question, "Is there a relationship between mathematics achievement and student population characteristics and/or community characteristics?", the following hypotheses were tested.

$H_0$1) There is no relationship between mathematics achievement and the percent of student population qualifying for free/reduced lunch. To test this hypothesis, the product-moment correlation (r) of the independent variable FREEREDL and each of the eight dependent variables was computed.

$H_0$2) There is no relationship between mathematics achievement and percent of students receiving special education services. To test this hypothesis, the product-moment correlation (r) of the independent variable SPED and each of the eight dependent variables was computed.

Research Question 2. To address the question, "Is there a relationship between mathematics achievement and fiscal factors?", the following hypotheses were tested.

$H_0$3) There is no relationship between mathematics achievement and per pupil expenditures. To test this hypothesis, the product-moment correlation (r) of the independent variable PPEXP and each of the eight dependent variables was computed.

$H_0$4) There is no relationship between mathematics achievement and percent of expenditures spent for *administrative* purposes. To test this hypothesis, the product-moment correlation (r) of the independent variable ADMNSPND and each of the eight dependent variables was computed.

$H_0$5) There is no relationship between mathematics achievement and percent of expenditures spent for *regular instruction* purposes. To test this hypothesis, the product-moment correlation (r) of the independent variable RGEDSPND and each of the eight dependent variables was computed.

$H_0$6) There is no relationship between mathematics achievement and percent of the budget revenues from local sources. To test this hypothesis, the product-moment correlation (r) of the independent variable LOCALREV and each of the eight dependent variables was computed.

Research Question 3. To address the question, "Is there a relationship between mathematics achievement and characteristics of the educational process?", the following hypotheses were tested.

$H_0$7) There is no significant difference in mathematics achievement associated with the schools' schedule design for delivery of instruction (block schedule vs. traditional schedule). To test this hypothesis, a one-way analysis of variance (ANOVA) involving the independent variable BLSCH and each of the eight dependent variables was computed.

$H_0$8) There is no relationship between this measure of mathematics achievement and other mathematics-related measures of student achievement. To test this hypothesis, the product-moment correlation (r) of the independent variables MTHGR8NP, ACTMATH, and CTMPCTP and each of the eight dependent variables was computed.

$H_0$9) There is no relationship between mathematics achievement and level of student readiness for related mathematics courses. To test this hypothesis, the product-moment correlation (r) of the independent variables PA_MPS, A1_MPS, A2_MPS, and G_MPS and each of the eight dependent variables was computed.

$H_0$10) There is no relationship between mathematics achievement and course-taking patterns. To test this hypothesis, the product-moment correlation (r) of the independent variable CTFAC and each of the eight dependent variables was computed.

$H_0$11) There is no relationship between mathematics achievement and drop-out rate. To test this hypothesis, the product-moment correlation (r) of the independent variable DROUTRT and each of the eight dependent variables was computed.

$H_0$12) There is no relationship between mathematics achievement and the attendance rate. To test this hypothesis, the product-moment correlation (r) of the independent variable ATNDRATE and each of the eight dependent variables was computed.

Research Question 4. To address the question, “Is there a relationship between mathematics achievement and measures of teacher quality?”, the following hypotheses were tested.

$H_0$13) There is no relationship between mathematics achievement and percent of the district’s teachers with advanced degrees. To test this hypothesis, the product-moment correlation (r) of the independent variables BSPLUS, MSDEG, and MSPLUS and each of the eight dependent variables was computed.

$H_0$14) There is no relationship between mathematics achievement and percent of the district's teachers at upper Career Ladder Levels (II or III). To test this hypothesis, the product-moment correlation (r) of the independent variables CLPCT and CLEPCT and each of the eight dependent variables was computed.

$H_0$15) There is no relationship between mathematics achievement and average teacher salary in the district. To test this hypothesis, the product-moment correlation (r) of the independent variable AVGTS and each of the eight dependent variables was computed.

<u>Regression analysis</u>

The next stage of the study was aimed at producing eight regression equations-- one equation for each of the dependent variables. As explained earlier, the independent variables were selected from among the larger group of independent variables included in the correlational analysis. Selected independent variables were those with a statistically significant correlation and a fit to the Opportunity-To-Learn model.

For four of the dependent variables (PA_MSS, A1_MSS, A2_MSS, and G_MSS) the multiple regression equations were represented as:

$$PA\_ MSS = b_1X_{ext} + b_2X_{fis} + b_3X_{edu} + b_4X_{tch} + b_0$$

$$A1\_ MSS = b_1X_{ext} + b_2X_{fis} + b_3X_{edu} + b_4X_{tch} + b_0$$

$$A2\_ MSS = b_1X_{ext} + b_2X_{fis} + b_3X_{edu} + b_4X_{tch} + b_0$$

$$G\_ MSS = b_1X_{ext} + b_2X_{fis} + b_3X_{edu} + b_4X_{tch} + b_0$$

where $b_{1,}$ b_2, b_3, and b_4 are regression weights and b_0 is a constant. X_{ext} is the predictor variable selected from the independent variables measuring external factors (in the OTL model). X_{fis} is the predictor variable selected from the independent variables measuring fiscal factors. X_{edu} is the predictor variable selected from the independent variables measuring educational process factors. X_{tch} is the predictor variable selected from the independent variables measuring teacher factors.

For the other four dependent variables (PA_SCHEF, A1_SCHEF, A2_SCHEF, and G_SCHEFF) the multiple regression equations were represented as:

$$PA\_SCHEF = b_1X_{ext} + b_2X_{fis} + b_3X_{edu} + b_4X_{tch} + b_0$$

$$A1\_SCHEF = b_1X_{ext} + b_2X_{fis} + b_3X_{edu} + b_4X_{tch} + b_0$$

$$A2\_SCHEF = b_1X_{ext} + b_2X_{fis} + b_3X_{edu} + b_4X_{tch} + b_0$$

$$G\_SCHEFF = b_1X_{ext} + b_2X_{fis} + b_3X_{edu} + b_4X_{tch} + b_0$$

where b_1, b_2, b_3, b_4, b_0, X_{ext}, X_{fis}, X_{edu}, and X_{tch} have the same meaning as above.

Summary

In this chapter the selection criteria and methods of data collection have been described. Variables for the study were described. An overview of the research design and data analysis was presented. More of the specifics of the data analysis and the results of the data analysis are presented in Chapter 4.

CHAPTER 4

PRESENTATION AND ANALYSIS OF DATA

Chapter 4 provides a description and interpretation of the results of the statistical analyses described in the previous chapter. In the first section of this chapter, a statistical profile of the schools and systems in the sample is presented. Next, a description of the findings from the testing of the hypotheses is presented. In the final section, the results from the regression analysis are described.

Description of the Schools/Systems in the Sample

As described in Chapter 3, this study examined a subset of the 139 school systems in Tennessee. The 65 school systems that had exactly one high school were selected for this study so that district-level information could more appropriately be examined. Table A-1 summarizes information about these 65 schools and systems.

The two variables used to examine Research Question 1 dealt with the percent of the district's students participating in the Free & Reduced Lunch program and the percent of the district's Average Daily Membership receiving Special Education services in 1995-96. As reported in Table A-1, the districts' percentage of free- and reduced- lunch students ranged from 18.7% to 81.5%, with a median of 39.7%. Student participation in special education programs ranged from a low of 10.9% to a high of 32.9% among the districts; the median rate of participation was 19.2%.

Four variables were used to examine Research Question 2. They are PPEXP (per pupil expenditure), ADMNSPND (administrative spending), RGEDSPND (regular education spending), and LOCALREV (local revenue). Table A-1 illustrates that the districts in this study had per pupil expenditures between $3,561 and $6,794 in the year studied; the median spending per pupil was $4,408.

The percentage of expenditures allocated to administrative functions ranged from 5.2% to 14.0% with a median of 8.7%. The percentage allocated to the regular education components of the district budget ranged from 44.4% to 64.4% with a median of 54.4%. The fourth factor, local revenue, revealed that contributions from local revenue sources ranged from 11.3% to 63.8%. The median local contribution was 52.5%.

Eleven variables were used in the six hypotheses formulated to examine Research Question 3, dealing with characteristics of the educational process. Table A-1 contains descriptive information about these eleven variables. Attendance rates ranged from 89.1% to 97.3% with a median of 93.7%. The median dropout rate of the districts in the study was 3.2%; the lowest rate was 0.0% and the highest rate 12.2%.

Three independent variables were mathematics-related measures of student achievement. These other measures of mathematics-related student achievement included the district's 1995-96 median national percentile score on the Grade 8 TCAP Achievement Test, the district's graduates' mean score on the mathematics section of the ACT test, and the percent of the district's 9th Grade students passing the Mathematics section of the TCAP Competency Test.

Among the districts, Grade 8 median Math scores ranged from the 28th to the 88th national percentiles, with the median at the 62nd percentile. ACT Math scores ranged from 15.5 to 21.9, with a median score of 18.3. The percent of 9th graders passing the Competency Test ranged from 37 to 99; the median percent passing was 75.

Table A-2 shows, that of the 65 schools in the study, 12 were identified as using block scheduling in their school program. Hypothesis 7 studies the relationship of block scheduling and mathematics achievement. Table A-1 shows that the median of the course-taking factors for the 65 schools is 1.03, implying that a slightly larger number of students in the study were enrolled in the non-required, more advanced courses of Algebra 2 and Geometry than in the Pre-Algebra and Algebra 1 courses. The course-taking ratios for the schools ranged from a low of .26 to a high of 2.75.

Table A-1 also reports information concerning the TVAAS calculation of predicted scores for the four subjects, studied in Hypothesis 9. This table reveals that Pre-Algebra, Algebra 1, and Algebra 2 have increasingly larger median scores for this variable, which illustrates their distribution along a common scale. The ranges for these three subjects were 44.1, 48.2, and 44.5, respectively. Geometry predicted scale scores, not on the common scale, ranged from 479.8 to 534.0, a range of 54.2.

Six variables were used in the three hypotheses used to examine Research Question 4, dealing with aspects of teacher quality. Table A-1 contains descriptive information about these variables. $H_0$13 examines the percent of the district's teachers with advanced degrees. The percentage of teachers in the districts with a masters degree as the highest earned degree ranged from 22.0 to 54.9, with a median of 35.7%. The percent of teachers with training higher than a masters degree (masters plus 45 quarter hours, educational specialist degree, or a doctorate) ranged from 2.4 to 34.0, with a median of 12.7%. The percentage of teachers with advanced degrees, defined as training higher than a bachelors degree, includes all the categories discussed above. Among the districts the percentage of teachers with advanced degrees ranged from 26.2 to 72.6, with a median of 48.1%.

The percent of classroom teachers eligible for Career Ladder Level II or III ranged from 57.1 to 90.8. The eligible classroom teachers who had attained Career Ladder Level II or III ranged from 5.4%to 50.0%. Median values showed 73.1% of classroom teachers eligible for upper levels of Career Ladder and 23.1% attaining this status. The median value for the variable Average Teacher's Salary was $29,152. District averages ranged from $26,419 to $41,453.

Table A-3 presents information relative to the mathematics achievement of the students in the sample's schools, from the perspective of *actual* achievement and of *value-added* (or *gain*) measures. The performance on the end-of-course mathematics tests is reflected by the eight dependent variables of this study. Table A-3 provides descriptive statistics for the dependent variables.

These descriptive statistics reveal a wide range of conditions in the 65 schools and systems in the study. Other findings reveal information about the relationship of the varied measures on these independent variables to the measures of mathematics achievement which form the basis of this study.

Hypothesis Testing

The following hypotheses were tested to explore each of the four general research questions discussed in Chapter 1. Each hypothesis was expressed in the null. In each case, each pertinent independent variable was compared to each of the eight dependent variables. A two-tailed test of significance was applied to each bivariate relationship, with the level of significance set at .05. Following is a discussion of each hypothesis with a concluding summary of the hypotheses.

Research Question 1

To address the question, "Is there a relationship between mathematics achievement and student population characteristics and/or community characteristics?", the following two hypotheses were tested. Detailed analyses are presented in Table A-4 and Table A-5.

$H_0$1. There is no relationship between mathematics achievement and the percent of student population qualifying for free/reduced lunch.

To test this hypothesis, the product-moment correlation (r) of the independent variable FREEREDL and each of the eight dependent variables was computed. The correlation coefficients are presented in Table 4-1 and Table 4-2. In the case of three of the dependent variables, the observed correlation coefficient was less than the critical value at the .05 level of significance, and the null hypothesis was not rejected.

Table 4-1 Correlations of external factors and mean scale scores

	Pearson Correlation			
	Pre-Algebra Mean Scale Score	Algebra 1 Mean Scale Score	Algebra 2 Mean Scale Score	Geometry Mean Scale Score
Free & Reduced Lunch	-.385**	-.301*	-.362**	-.352**
Special Education	.121	.031	-.071	.117

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

For the other five dependent variables, the null hypothesis was rejected. As shown in Table 4-1 and Table 4-2, statistically significant correlations were determined for the Mean Scale Scores for each of the four subjects and for the Pre-Algebra School Effect. The significant correlation coefficients ranged from -.385 ($p < 0.01$) for Pre-Algebra Mean Scale Score to -.301 ($p < 0.05$) for Algebra 1 Mean Scale Score.

$H_0$2. There is no relationship between mathematics achievement and percent of students receiving special education services. To test this hypothesis, the product-moment correlation (r) of the independent variable SPED and each of the eight dependent variables was computed. The correlation coefficients are presented in Table 4-1 and Table 4-2.

For each of the eight dependent variables, the observed correlation coefficient was less than the critical value at the .05 level of significance, and the null hypothesis was retained.

Table 4-2 Correlations of external factors and value-added scores

	Pearson Correlation			
	Pre-Algebra School Effect	Algebra 1 School Effect	Algebra 2 School Effect	Geometry School Effect
Free & Reduced Lunch	-.319*	-.221	-.215	-.076
Special Education	.006	.082	.001	.075

*. Correlation is significant at the 0.05 level (2-tailed).

Research Question 2

To address the question, “Is there a relationship between mathematics achievement and fiscal factors?”, the following four hypotheses were tested. Detailed analyses are presented in Table A-6 and Table A-7.

$H_0 3$. There is no relationship between mathematics achievement and per pupil expenditures.

To test this hypothesis, the product-moment correlation (r) of the independent variable PPEXP and each of the eight dependent variables was computed. The correlation coefficients are presented in Table 4-3 and Table 4-4. In the case of four of the dependent variables, the observed correlation coefficient was less than the critical value at the .05 level of significance, and the null hypothesis was not rejected.

For the other four dependent variables, the null hypothesis was rejected. As shown in Table 4-3 and Table 4-4, statistically significant correlations were determined for the Algebra 2 Mean Scale Score, the Geometry Mean Scale Score, and for the Algebra 1 and Algebra 2 School Effects. The significant correlation coefficients ranged from .337 ($p < 0.01$) for Algebra 2 Mean Scale Score to .263 ($p < 0.05$) for Geometry Mean Scale Score.

$H_0 4$. There is no relationship between mathematics achievement and percent of expenditures spent for *administrative* purposes.

Table 4-3 Correlations of fiscal factors and mean scale scores

	Pearson Correlation			
	Pre-Algebra Mean Scale Score	Algebra 1 Mean Scale Score	Algebra 2 Mean Scale Score	Geometry Mean Scale Score
Per Pupil Expenditure	.127	.194	.337**	.263*
Administrative spending	.174	.113	.046	.128
Regular Education spending	.135	.163	.390**	.192
Local Revenue	.229	.171	.478**	.351**

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 4-4 Correlations of fiscal factors and value-added scores

	Pearson Correlation			
	Pre-Algebra School Effect	Algebra 1 School Effect	Algebra 2 School Effect	Geometry School Effect
Per Pupil Expenditure	.238	.280*	.305*	.170
Administrative spending	.178	-.016	-.010	.026
Regular Education spending	.148	.115	.223	.086
Local Revenue	.321*	.207	.319**	.160

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

To test this hypothesis, the product-moment correlation (r) of the independent variable ADMNSPND and each of the eight dependent variables was computed. The correlation coefficients are presented in Table 4-3 and Table 4-4.

For each of the eight dependent variables, no statistically significant relationship was found; thus the null hypothesis was retained in each case.

$H_0$5. There is no relationship between mathematics achievement and percent of expenditures spent for *regular instruction* purposes.

To test this hypothesis, the product-moment correlation (r) of the independent variable RGEDSPND and each of the eight dependent variables was computed. The correlation coefficients are presented in Table 4-3 and Table 4-4.

A significant relationship involving this independent variable and the dependent variable Algebra 2 Mean Scale Score was determined. The correlation coefficient was .390 ($p < 0.01$); thus the null hypothesis was rejected.

For the other seven dependent variables, no statistically significant relationship was found; thus the null hypothesis was retained.

$H_0$6. There is no relationship between mathematics achievement and percent of the budget revenues from local sources.

To test this hypothesis, the product-moment correlation (r) of the independent variable LOCALREV and each of the eight dependent variables was computed. The correlation coefficients are presented in Table 4-3 and Table 4-4. In the case of four of the dependent variables, the observed correlation coefficient was less than the critical value at the .05 level of significance, and the null hypothesis was not rejected.

For the other four dependent variables, the null hypothesis was rejected. As shown in Table 4-3 and Table 4-4, statistically significant correlations were determined for the Algebra 2 Mean Scale Score, the Geometry Mean Scale Score, the Pre-Algebra School Effect, and Algebra 2 School Effect. The significant correlation coefficients ranged from .478 ($p < 0.01$) for Algebra 2 Mean Scale Score to .319 ($p < 0.01$) for Algebra 2 School Effect.

<u>Research Question 3</u>

To address the question, “Is there a relationship between mathematics achievement and characteristics of the educational process?”, the following six hypotheses were tested. Detailed analyses are presented in Table A-8, Table A-9, and Table A-10.

$H_0 7$. There is no significant difference in mathematics achievement associated with the schools’ schedule design for delivery of instruction (block schedule vs. traditional schedule).

To test this hypothesis, a one-way analysis of variance (ANOVA) involving the independent variable BLSCH and each of the eight dependent variables was computed. The one-way analysis of variance results (presented in Table A-10) shows that, for each of the eight dependent variables, no *F* approached a level of statistical significance. Therefore, for each of the eight dependent variables, the null hypothesis was retained. Comparisons of the ANOVA results to calculated Pearson’s product-moment correlation coefficients and Mann-Whitney *U* test statistics confirm the null hypothesis.

Table 4-5 Correlations of educational process factors and mean scale scores

	Pearson Correlation			
	Pre-Algebra Mean Scale Score	Algebra 1 Mean Scale Score	Algebra 2 Mean Scale Score	Geometry Mean Scale Score
Attendance Rate	.172	.292*	.145	.054
Dropout Rate	.005	-.102	-.094	-.109
Math Gr.8 National Percentile	.349**	.335**	.272*	.275*
ACT Math	.379**	.471**	.664**	.556**
Competency Test - percent passing	.464**	.320**	.149	.110
Pre-Algebra Mean Predicted Score	.645**	.283*	.143	.401**
Algebra 1 Mean Predicted Score	.340**	.696**	.098	.165
Algebra 2 Mean Predicted Score	.096	.243	.673**	.292*
Geometry Mean Predicted Score	.372**	.449**	.367**	.644**
Course-taking factor	-.015	-.024	.159	.015

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

$H_0 8$. There is no relationship between this measure of mathematics achievement and other mathematics-related measures of student achievement.

To test this hypothesis, the product-moment correlation (r) of the independent variables MTHGR8NP, ACTMATH, and CTMPCTP and each of the eight dependent variables was computed. The correlation coefficients are presented in Table 4-5 and Table 4-6.

For the independent variable MTHGR8NP, the observed correlation coefficient was less than the critical value at the .05 level of significance for each the four dependent variables concerned with school effects, and the null hypothesis was not rejected. For the other four dependent variables (dealing with mean scale scores), the null hypothesis was rejected. As shown in Table 4-5, significant correlation coefficients ranged from .349 ($p < 0.01$) for Pre-Algebra Mean Scale Score to .272 ($p < 0.05$) for Algebra 2 Mean Scale Score.

For the independent variable ACTMATH statistically significant correlations ($p < 0.01$) were determined for each of the eight dependent variables; the null hypothesis was rejected in each case.

Table 4-6 Correlations of educational process factors and value-added scores

	Pearson Correlation			
	Pre-Algebra School Effect	Algebra 1 School Effect	Algebra 2 School Effect	Geometry School Effect
Attendance Rate	.138	.139	-.003	-.111
Dropout Rate	.012	.052	.099	.028
Math Gr.8 National Percentile	.227	.160	.080	.016
ACT Math	.415**	.510**	.534**	.344**
Competency Test - percent passing	.404**	.106	.023	-.096
Pre-Algebra Mean Predicted Score	.198	.113	.059	.163
Algebra 1 Mean Predicted Score	.210	.184	-.064	-.093
Algebra 2 Mean Predicted Score	.008	.085	.276*	-.015
Geometry Mean Predicted Score	.168	.279*	.142	.125
Course-taking factor	.102	-.099	.064	-.114

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

For the independent variable CTMPCTP, the observed correlation coefficient was less than the critical value at the .05 level of significance in the case of five of the dependent variables. In these five cases the null hypothesis was not rejected. For the other three dependent variables, the null hypothesis was rejected. As shown in Table 4-5 and Table 4-6, statistically significant correlations were determined for Pre-Algebra Mean Scale Score, Algebra 1 Mean Scale Score, and Pre-Algebra School Effect. The significant correlation coefficients ranged from .464 ($p < 0.01$) for Pre-Algebra Mean Scale Score to .320 ($p < 0.01$) for Algebra 1 Mean Scale Score.

For each of the eight dependent variables, the null hypothesis was rejected at least once; at least one of the three independent variables was determined to have a statistically significant correlation.

$H_0 9$. There is no relationship between mathematics achievement and level of student readiness for related mathematics courses.

To test this hypothesis, the product-moment correlation (r) of the independent variables PA_MPS, A1_MPS, A2_MPS, and G_MPS and each of the eight dependent variables was computed. The correlation coefficients are presented in Table 4-5 and Table 4-6.

For the independent variable PA_MPS, the observed correlation coefficient was less than the critical value at the .05 level of significance for each the four dependent variables concerned with school effects, and the null hypothesis was not rejected. For three of the other four dependent variables (dealing with mean scale scores), the null hypothesis was rejected. As shown in Table 4-5, significant correlation coefficients ranged from .645 ($p < 0.01$) for Pre-Algebra Mean Scale Score to .283 ($p < 0.05$) for Algebra 1 Mean Scale Score.

For the independent variable A1_MPS statistically significant correlations were determined for two of the eight dependent variables; the null hypothesis was rejected for the dependent variables Pre-Algebra Mean Scale Score and Algebra 1 Mean Scale Score. The correlation coefficients were .340 ($p < 0.01$) and .696 ($p < 0.01$) respectively. For the other six dependent variables, no statistically significant relationship was found; thus the null hypothesis was retained.

For the independent variable A2_MPS, the observed correlation coefficient was less than the critical value at the .05 level of significance in the case of five of the dependent variables. In these five cases the null hypothesis was not rejected. For the other three dependent variables, the null hypothesis was rejected. As shown in Table 4-5 and Table 4-6, statistically significant correlations were determined for Algebra 2 Mean Scale Score, Geometry Mean Scale Score, and Algebra 2 School Effect. The significant correlation coefficients ranged from .673 ($p < 0.01$) for Algebra 2 Mean Scale Score to .276 ($p < 0.01$) for Algebra 2 School Effect.

For the independent variable G_MPS, the observed correlation coefficient was less than the critical value at the .05 level of significance in the case of three of the dependent variables. In these cases the null hypothesis was not rejected.

For the other five dependent variables, the null hypothesis was rejected. As shown in Table 4-5 and Table 4-6, statistically significant correlations were determined for the Mean Scale Score in all four courses and the Algebra 1 School Effect. The significant correlation coefficients ranged from .644 ($p < 0.01$) for Geometry Mean Scale Score to .279 ($p < 0.05$) for Algebra 1 School Effect.

$H_0$10. There is no relationship between mathematics achievement and course-taking patterns.

To test this hypothesis, the product-moment correlation (r) of the independent variable CTFAC and each of the eight dependent variables was computed. The correlation coefficients are presented in Table 4-5 and Table 4-6.

For each of the eight dependent variables, no statistically significant relationship was found; thus the null hypothesis was retained.

$H_0$11. There is no relationship between mathematics achievement and drop-out rate.

To test this hypothesis, the product-moment correlation (r) of the independent variable DROUTRT and each of the eight dependent variables was computed. The correlation coefficients are presented in Table 4-5 and Table 4-6.

For each of the eight dependent variables, no statistically significant relationship was found; thus the null hypothesis was retained.

$H_0$12. There is no relationship between mathematics achievement and the attendance rate.

To test this hypothesis, the product-moment correlation (r) of the independent variable ATNDRATE and each of the eight dependent variables was computed. The correlation coefficients are presented in Table 4-5 and Table 4-6.

For the dependent variable Algebra 1 Mean Scale Score a relationship with this independent variable was found. The correlation coefficient was .292 ($p < 0.05$); thus the null hypothesis was rejected. For the other seven dependent variables, no statistically significant relationship was found; thus the null hypothesis was retained.

Research Question 4

To address the question, "Is there a relationship between mathematics achievement and measures of teacher quality?", the following three hypotheses were tested. Detailed analyses are presented in Table A-11 and Table A-12.

$H_0$13. There is no relationship between mathematics achievement and percent of the district's teachers with advanced degrees.

To test this hypothesis, the product-moment correlation (r) of the independent variables BSPLUS, MSDEG, and MSPLUS and each of the eight dependent variables was computed. The correlation coefficients are presented in Table 4-7 and Table 4-8.

For six of the dependent variables, statistically significant relationships were found; thus the null hypothesis was rejected. The largest correlation coefficient was found for the relationship of percentage of teachers with a Masters degree and Algebra 2 Mean Scale Score; this coefficient was .497 ($p < 0.01$). The smallest significant correlation coefficient was found for the relationship of percentage of teachers with a Masters degree and Algebra 1 Mean Scale

Table 4-7 Correlations of teacher factors and mean scale scores

	Pearson Correlation			
	Pre-Algebra Mean Scale Score	Algebra 1 Mean Scale Score	Algebra 2 Mean Scale Score	Geometry Mean Scale Score
Career Ladder Percent	.216	.227	.248*	.165
Career Ladder Eligible Percent	.165	.201	.407**	.110
Bachelors plus	.195	.280*	.423**	.308*
Masters degree	.129	.262*	.497**	.309*
Masters plus	.132	.118	.062	.100
Average Teacher's Salary	.271*	.279*	.522**	.377**

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 4-8 Correlations of teacher factors and value-added scores

	Pearson Correlation			
	Pre-Algebra School Effect	Algebra 1 School Effect	Algebra 2 School Effect	Geometry School Effect
Career Ladder Percent	.250	.161	.126	.018
Career Ladder Eligible Percent	.212	.174	.237	-.113
Bachelors plus	.291*	.346**	.360**	.169
Masters degree	.209	.331**	.450**	.162
Masters plus	.179	.137	.022	.064
Average Teacher's Salary	.358**	.308*	.394**	.188

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Score; this coefficient was .262 ($p < 0.05$).

For the dependent variables Geometry School Effect and Pre-Algebra Mean Scale Score, no statistically significant relationship was found; thus the null hypothesis was retained.

$H_0$14. There is no relationship between mathematics achievement and percent of the district's teachers at upper Career Ladder Levels (II or III).

To test this hypothesis, the product-moment correlation (r) of the independent variables CLPCT and CLEPCT and each of the eight dependent variables was computed. The correlation coefficients are presented in Table 4-7 and Table 4-8.

For the dependent variable Algebra 2 Mean Scale Score a relationship with these independent variables was found. The correlation coefficient with CLPCT was .248 ($p < 0.05$) and with CLEPCT .407 ($p < 0.01$); thus the null hypothesis was rejected.

For the other seven dependent variables, no statistically significant relationship was found; thus the null hypothesis was retained.

$H_0$15. There is no relationship between mathematics achievement and average teacher salary in the district.

To test this hypothesis, the product-moment correlation (r) of the independent variable AVGTS and each of the eight dependent variables was computed. The correlation coefficients are presented in Table 4-7 and Table 4-8.

For the dependent variable Geometry School Effect, no statistically significant relationship was found; thus the null hypothesis was retained.

For the other seven dependent variables, statistically significant relationships were found; thus the null hypothesis was rejected. The significant correlation coefficients ranged from .522 ($p < 0.01$) for Algebra 2 Mean Scale Score to .271 ($p < 0.05$) for Pre-Algebra Mean Scale Score.

In this study, 15 hypotheses were tested to address the four major research questions. Some of the hypotheses used multiple independent variables and each hypothesis was tested for eight dependent variables. Table 4-9 summarizes the results of the hypothesis testing discussed in this section.

Regression Analysis

As described in Chapter 3, the next stage of the study was aimed at producing eight regression equations--one equation for each of the dependent variables. The multiple regression equations are represented as:

$$PA\_MSS = b_1X_{ext} + b_2X_{fis} + b_3X_{edu} + b_4X_{tch} + b_0$$

$$A1\_MSS = b_1X_{ext} + b_2X_{fis} + b_3X_{edu} + b_4X_{tch} + b_0$$

$$A2\_MSS = b_1X_{ext} + b_2X_{fis} + b_3X_{edu} + b_4X_{tch} + b_0$$

$$G\_MSS = b_1X_{ext} + b_2X_{fis} + b_3X_{edu} + b_4X_{tch} + b_0$$

$$PA\_SCHEF = b_1X_{ext} + b_2X_{fis} + b_3X_{edu} + b_4X_{tch} + b_0$$

$$A1\_SCHEF = b_1X_{ext} + b_2X_{fis} + b_3X_{edu} + b_4X_{tch} + b_0$$

$$A2\_SCHEF = b_1X_{ext} + b_2X_{fis} + b_3X_{edu} + b_4X_{tch} + b_0$$

$$G\_SCHEFF = b_1X_{ext} + b_2X_{fis} + b_3X_{edu} + b_4X_{tch} + b_0$$

where b_1, b_2, b_3, and b_4 are regression weights and b_0 is a constant. X_{ext} is the predictor variable selected from the independent variables measuring external factors (in the OTL model). X_{fis} is the predictor variable selected from the independent variables measuring fiscal factors. X_{edu} is the predictor variable selected from the independent variables measuring educational process factors. X_{tch} is the predictor variable selected from the independent variables measuring teacher factors.

Table 4-9 Summary of hypothesis testing

Research Question	1 External factors		2 Fiscal factors				3 Educ. Process factors						4 Teacher Quality factors		
Hypothesis #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Pre-Algebra Mean Scale	R							R	R*						R
Pre-Algebra School Effect	R					R		R*					R*		R
Algebra 1 Mean Scale Score	R							R	R*			R	R*		R
Algebra 1 School Effect			R					R*	R*				R*		R
Algebra 2 Mean Scale Score	R		R		R	R		R*	R*				R*	R	R
Algebra 2 School Effect			R			R		R*	R*				R*		R
Geometry Mean Scale Score	R		R			R		R*	R*				R*		R
Geometry School Effect								R*							

R = null hypothesis rejected
R* = null hypothesis rejected (for some of multiple independent variables)

The independent variables selected for this analysis (X_{ext}, X_{fis}, X_{edu}, and X_{tch}) were chosen from among the larger group of independent variables included in the correlational analysis. Selected independent variables were those with a statistically significant correlation and a fit to the Opportunity-to-Learn model. A summary of selected independent variables shown to have statistically significant correlations with each of the dependent variables is presented in Table 4-10. These independent variables were chosen to be potential candidates for the X_{ext}, X_{fis}, X_{edu}, and X_{tch} variables in the eight regression equations. The variables X_{ext}, X_{fis}, X_{edu}, and X_{tch} would not necessarily be the same in different equations.

Independent variables that reflect measurements occurring after the dependent variables' measurements were not included as potential predictor variables and are not included in Table 4-10. For instance, ACTMATH is not included as a potential independent variable because the ACT test is generally taken after a student has finished the math courses included in this study.

Attempts to fit the independent variables (which had proved to have statistically significant relationships with the dependent variables) into the eight complete regression equations were challenging. One area of difficulty is readily apparent in Table 4-10. Two of the regression equations did have an “eligible” candidate for each of the four predictor variables; however, some of the potential regression equations did not have a significant independent variable “candidate” for each of the four predictor variables. In the most extreme situation, only one independent variable was determined to have a significant correlation with the dependent variable, but it reflected a measurement occurring *after* the dependent variable’s measurement. Consequently, for the equation for Geometry School Effect, none of the predictor variables (X_{ext}, X_{fis}, X_{edu}, and X_{tch}) had “eligible” candidates from among the independent variables.

Another preliminary phase to the determination of the regression equations was the consideration of collinearity. Collinearity, sometimes called multicollinearity, exists when the independent variables being studied are statistically related to each other. Pedhazur has stated “Correlations among independent variables may lead to difficulties in the estimation of

Table 4-10 Independent variables with potential as predictor variables

Research Question	1 External factors	2 Fiscal factors	3 Educ. Process factors	4 Teacher Quality factors
Pre-Algebra Mean Scale Score	FREEREDL**		PA_MPS**, CTMPCTP**, MTHGR8NP**	AVGTS*
Pre-Algebra School Effect	FREEREDL*	LOCALREV*	CTMPCTP**	AVGTS**, BSPLUS*
Algebra 1 Mean Scale Score	FREEREDL*		A1_MPS**, MTHGR8NP**, CTMPCTP**, ATNDRATE*, PA_MPS*	BSPLUS*, AVGTS*, MSDEG*
Algebra 1 School Effect		PPEXP*		BSPLUS**, MSDEG**, AVGTS*
Algebra 2 Mean Scale Score	FREEREDL**	LOCALREV**, RGEDSPND**, PPEXP**	A2_MPS**, G_MPS**, MTHGR8NP*	AVGTS**, MSDEG**, BSPLUS**, CLEPCT**, CLPCT*
Algebra 2 School Effect		LOCALREV**, PPEXP*	A2_MPS*	MSDEG**, AVGTS**, BSPLUS**
Geometry Mean Scale Score	FREEREDL**	LOCALREV**, PPEXP*	G_MPS**, PA_MPS**, A2_MPS* , MTHGR8NP*	AVGTS**, MSDEG*, BSPLUS*
Geometry School Effect				

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

regression statistics. Problems emanating from intercorrelations among independent variables are generally discussed under the heading of multicollinearity" (1982, pp. 232-233). Pedhazur continued:

> There is ... no consensus about the meaning of this term. Some use it to refer to the existence of any correlations among the independent variables, whereas others reserve the

term to describe a situation in which the independent variables are highly correlated, although there is, understandably, no agreement what "high" means. (p. 233)

In the consideration of collinearity in this study, it was determined that the levels of correlation among the independent variables were not so strong as to preclude the determination of the regression equations. Table A-13 presents the correlation statistics of several of the independent variables that appeared to be "eligible" candidates for the predictor variables, based on their correlations to the dependent variables. Although certain statistical relationships were reported in this table, the pattern of collinearity among the independent variables was determined to be satisfactory for this study.

In determining the eight regression equations, a variety of potential predictor variables were entered in a variety of different sequences, using the full range of regression techniques available in SPSS. Table 4-11 presents the seven regression equations that resulted from this

Table 4-11 Regression equations in an OTL framework

Regression equation	R^2	adj. R^2	Standard error
PA_ MSS = -.13FREEREDL + 1.22PA_MPS + .00095AVGTS - 124.30	.491	.462	12.2
A1_ MSS = -.075FREEREDL + 1.19A1_MPS + .45BSPLUS - 117.57	.549	.527	12.8
A2_ MSS = -2.15FREEREDL + .082RGEDSPND + 1.20A2_MPS + .92MSDEG - 146.91	.534	.503	16.2
G_ MSS = -.0068FREEREDL + .0021PPEXP + 1.09G_MPS + .15MSDEG - 59.23	.430	.392	13.9
PA_SCHEF = -.056FREEREDL + .24CTMPCTP + .00066AVGTS - 36.10	.214	.169	10.3
A1_SCHEF = .0024PPEXP + .37BSPLUS - 29.75	.132	.104	12.2
A2_SCHEF = .0015PPEXP + .21A2_MPS + .82MSDEG - 151.03	.222	.184	14.9
G_SCHEFF [*no satisfactory equation resulted*]			

analysis. The equations represent the best effort to match the OTL framework, with predictors from each of the categories of variables (external, fiscal, educational process, and teacher factors) represented.

Other regression equations would have been possible if different combinations of independent variables had been entered into the analyses. In some instances, including two or more variables from the same OTL category would have produced different, perhaps better, regression situations. One observation that is immediately obvious from Table 4-11 is that the three *value-added* equations that were formulated have smaller R^2 values than the four equations involving *actual* scores. Other observations and discussion will be presented in Chapter 5.

Summary

This chapter has provided a description of the results of the statistical analyses described in Chapter 3. The schools and systems in the sample have been described, using the statistical profile summarized and presented in Tables A-1, A-2, and A-3. A description of the findings from the testing of the hypotheses has been presented and summarized. In this chapter's final section, the regression analysis was described. The seven regression equations (which were based on the theoretical model) were presented in Table 4-11.

In the next chapter, the discussion of the findings from these analyses will be continued. Chapter 5 will place an emphasis on an interpretation of the results and a discussion of their implications.

CHAPTER 5

DISCUSSION AND CONCLUSIONS

In Chapter 5, the four major research questions will form a basis for discussion of the research findings of this study. The implications of the results of the statistical analyses will be discussed. The relationship and roles of educational research and educational policy will be discussed. Finally, recommendations consistent with this study and the related discussion are presented.

Discussion

In this section, a further look at the findings will be presented. The findings will be discussed with a look at potential implications and the "fit" with other research and theory. Discussion is organized initially around the four research questions. This is followed by a look at the Opportunity-to-Learn framework which served as the basis of this study.

Research Question 1

This question dealt with the measures of student-population characteristics which are not considered to be under the direct control of the school. Two variables were selected from the 1995-96 21^{st} Century Schools Report Card data set to represent External factors in this study. One of the two variables, the percent of the district's students participating in the Free & Reduced Lunch program, yielded some interesting findings. First, the sheer range among the districts was noteworthy; the districts ranged from 18.7% to 81.5% in students in the lower income levels. Second, the income levels did prove to be significantly related to measures of mathematics achievement. The relationships exhibited were, in all cases, negative. In each measure of *actual* achievement among the dependent variables, there were significant negative correlation coefficients with this variable. In one measure of *value-added* achievement, Pre-Algebra School Effect, there was also a significant negative correlation .

This negative relationship also proved to exist among measures of student achievement contained among other independent variables. As discussed in Chapter 4, a trace of collinearity was found among the independent variables. Among these correlations is that the Free & Reduced Lunch independent variable is negatively correlated to ACT Math scores, -.589 ($p < 0.01$), the percent passing the Competency Test, -.488 ($p < 0.01$), and the district's median national percentile score on the Grade 8 TCAP Achievement Test, -.407 ($p < 0.01$). Nevertheless, the collinearity was not evident to a degree sufficient to cause concern in this study.

This study also showed evidence, consistent with findings cited in the review of literature in Chapter 2, that schools with poorer students tend to have less experienced teachers and less qualified teachers (as measured by level of degree attained). In this study, the correlation coefficient between the Free & Reduced Lunch independent variable and a measure of teacher experience, the independent variable measuring the percent of teachers eligible for upper levels of the Career Ladder (by years of experience), was -.317 ($p < 0.05$). The correlation coefficient between the Free & Reduced Lunch variable and the percent of teachers with a Masters degree or higher was -.346 ($p < 0.01$). The correlation with the Average Teacher's Salary was also negative and significant, -.482 ($p < 0.01$).

The other variable studied in this category was the percent of the district's Average Daily Membership receiving Special Education services. This variable did not appear to be a significant factor in any of the analyses. This would likely be due to the relatively small number of special education students with severe handicapping conditions actually enrolled in the courses analyzed in this study.

One potential factor that is reported in the state's 21st Century Report Card could have added additional information to this study. The per capita income is reported in the report card;

however, it is calculated on a county-wide basis. Considering the average income of the school district's residents would have given further insight into the characteristics of the population that are external factors, or factors beyond the control of the school. This factor was not included in the study because there is a recognized difference in population characteristics among the districts in counties with multiple school systems.

Research Question 2

The effects of fiscal factors, the focus of this research question, were explored through four relevant variables obtained from the 1995-96 Report Card. The Per Pupil Expenditure factor, a measure used in many educational studies, did prove to be significantly related to some of the measures of mathematics achievement in this study. The relationships that were exhibited were positive. They were most evident in the higher-level courses, with no relationships demonstrated for pre-algebra. The strongest positive relationship was exhibited in the Algebra 2 Mean Scale Score; the Geometry Mean Scale Score was also significantly positively correlated. Higher spending levels per pupil resulted in higher *actual* scores, for the more advanced courses. For *value-added* measures, higher spending levels per pupil resulted in better gains for Algebra 1 and Algebra 2 students.

The next two factors, administrative spending and regular education spending, explored the school systems' allocation of available resources. Specifically, they reflect the percentage of expenditures allocated to administrative functions and to the regular education components of the district (as compared to special education, support services, extra-curricular activities, etc.). The variable that focused on spending for administrative functions (ADMNSPND) exhibited no statistically significant relationship with any of the eight dependent variables.

It was only with one of the dependent variables (Algebra 2 Mean Scale Score) that a relationship with spending for the regular education program was found. At least two conclusions could be reached from this lack of demonstrated statistical correlation involving the distribution percentages of a district's fiscal resources. One is that how money is allocated within a district does not matter. A more logical explanation is that districts in Tennessee spend the available resources in similar patterns. As shown in Table A-1, administrative spending ranges only from 5.2% to 14.0%. This logic, as cited in Chapter 2, is that "School districts spend their money in remarkably consistent proportions, regardless of their wealth" (Mosburg, 1996, p. 13).

Regular education spending percentages (Table A-1) ranged from 44% to 64%. Although only one dependent variable was significantly and positively correlated with this factor, relationships were demonstrated with other independent variables. Associated with increased regular education spending were better-trained and more experienced teachers; correlation coefficients for regular education spending and number of teachers with Masters degrees and percent of teachers eligible for Career Ladder upper levels were .363 ($p < .01$) and .460 ($p < .01$) respectively. Since degree level and experience are the two primary factors which cause a teacher to "move up" any given district's salary scale, it follows that regular education spending was also shown to be correlated to the average teacher's salary ($r = .468$, $p < .01$).

The remaining fiscal factor, local revenue, gives insight into both the *capacity* and *willingness* of the local community to contribute local resources for educational purposes. The percent of the district's budget coming from local revenue sources was determined to have statistically significant correlations ($p < 0.01$) with the Algebra 2 Mean Scale Score ($r = .478$), Geometry Mean Scale Score ($r = .351$), Algebra 2 School Effect ($r = .319$), and with the Pre-Algebra School Effect ($r = .321$, $p < .05$) . With the exception of the Pre-Algebra School Effect, all the significant relationships were shown to be with the relatively higher-level courses.

The relationship of this factor (local revenue) with Algebra 2 and Geometry scores and the relationship of the regular education spending factor with the dependent variable Algebra 2 Mean Scale Score underscores the earlier observation that fiscal variables tend to show a relationship at the higher-level mathematics courses in this study, especially Algebra 2. In the discussion on per pupil expenditures it was noted that higher per-pupil spending levels resulted in higher *actual* scores in the more advanced courses of this study. Fiscal variables do not seem to affect the districts' student performance in lower-level mathematics courses, within the range of fiscal resources measured and studied here.

Whether this observation would hold true for even higher levels of mathematics courses, not examined in this study because of their exclusion from the TCAP battery of end-of-course tests, could not be determined. However, an examination of the ACT math scores, which logically would be positively influenced by participation in such courses as pre-calculus or Advanced Placement (AP) calculus, revealed other statistical relationships with these fiscal factors. As shown in Table A-13, ACT math scores are positively and significantly correlated with the local revenue, per pupil expenditure, and regular education spending variables.

As in the case of per capita income, which was discussed with the external factors, there are also additional fiscal factors that are reported in the state's 21st Century Report Card that could have provided valuable additional data to this study. Unfortunately, these fiscal data are also calculated at the county level, and thus inappropriate for this study. Two of these potential factors involve the tax base per pupil. The state reports the property tax per pupil and the sales tax per pupil; both of these factors could have been included in this study as measures of the capacity to generate revenue if they were more specifically defined by school district.

An even more potentially valuable measure involves the factor that reports the ratio of *actual* revenue generated locally for schools to the local *capacity* to generate revenue. This ratio is calculated by comparing the county's percent of statewide capacity to generate revenue to the percent of the statewide revenue actually generated. A ratio of less than 1.0 would indicate that the county allocated less money to the schools than their capacity would have indicated; this implies a weak local commitment to fund education. On the other hand, if a county elects to make funding for education a priority, and contributes a higher percentage than the statewide average, the calculated ratio would be greater than 1. The ratio of actual to capacity ranged from

0.6 to 1.6 in 1995-96, the year of this study, at the county level. This range could have been even greater if the measure were calculated and reported at the more localized, school-district level.

Research Question 3

The effects of factors pertaining to the educational process *(Ed)* were explored through 11 variables. Chapter 4 presented the findings of the analyses involving these variables.

The district's math scores from the eighth-grade achievement tests, the school's ACT Math test scores, and Competency test scores were included in this study to give a more comprehensive view of the district's overall achievement level in mathematics. These three measures were established at relatively different points in time, as compared to the end-of-course tests that were the primary measures of student achievement in the study. The district's eighth-grade scores obviously would have preceded the high school tests. The Competency test scores were calculated in the students' ninth-grade year, and would have preceded most, if not all, of the end-of-course test results.

The district's eighth-grade achievement test math scores were shown to be significantly correlated to the four dependent variables dealing with mean scale scores, or the *actual* scores. The strength of the correlational relationship decreased as the difficulty level of the course increased; the correlation was strongest for the Pre-Algebra Mean Scale Score and weaker, though still significant, for Algebra 2 Mean Scale Score.

The ninth-grade Competency test scores were shown to be significantly correlated to three dependent variables. The three measures correlated to Competency test scores, Pre-Algebra Mean Scale Score, Pre-Algebra School Effect, and Algebra 1 Mean Scale Score, are at the more basic level of the four courses studied. Since the ninth-grade Competency test is a minimum-competency test, it follows that performance on this test would be related to the performance of students in the lower-level classes, the students who would more likely be near the threshold or cut-off score for passing a basic-skills test.

Performance on the ACT mathematics test was shown to be significantly correlated to each of the eight dependent variables. In fact, ACTMATH was the only independent variable to do so. The correlations were relatively strong, the correlation coefficients were all significant at the 0.01 level. The districts' students who scored well on the mathematics portion of the ACT test had also previously scored well on the *actual* scores on all four courses; they also had good *value-added* scores.

The four variables involving predicted scores, PA_MPS, A1_MPS, A2_MPS, and G_MPS, yielded a mixed picture in this analysis. There was a fairly strong demonstrated relationship among these variables and the *actual* measures among the dependent variables. *Value-added* measures, the school effects, were only mildly related.

The four predicted scores are a product of the Tennessee Value-Added Assessment System (TVAAS), calculated by regression analysis from all available student scores on any previous TCAP tests. Among the relationships demonstrated by the *actual* measures, it should be no surprise that each predicted score was strongly correlated to the mean scale score for that same subject. In each of the four subjects these measures were strongly correlated, with a Pearson r coefficient greater than .6.

Other relationships demonstrated by the *actual* measures and the predicted scores illustrate a pattern whereby the actual scores and the predicted scores for any given subject are similar to the courses above and below the given course. In other words, a given level of performance is influenced by previous performance at a school; this performance, in turn, influences future performance.

For this study, these four variables and the calculated course-taking factor were included to gain insight into whether or not there is an optimum level of readiness for the various courses to maximize student achievement. Questions posed in Chapter 3 included whether or not schools should structure their program for relatively small groups of well prepared students. Or, should they encourage more students into higher levels of courses, with a lower average level of preparation?

There was, however, no relationship shown between mathematics achievement and course-taking patterns, which compared participation in higher-level and lower-level courses. Similarly, very little was proven concerning the relationship of predicted scores, or level of readiness, and *value-added* scores.

Other variables in the educational process category were shown to have no relationship with the variables measuring student performance. No difference was found between the group of 12 schools using a block schedule and the group of 53 that used a traditional schedule. There was also no exhibited relationship between mathematics achievement and drop-out rate for the school.

It was only for one dependent variable, the Algebra 1 Mean Scale Score, that a relationship with the district’s attendance rate was found. Although no stronger relationship was proven concerning the attendance rate’s impact on the learning of mathematics in general, this single finding may warrant further study. As the Algebra 1 end-of-course test is phased in as a graduation requirement, any finding concerning improved performance on the Algebra 1 test will be especially important to students and to the district (as a measure of accountability).

As in the case of the two previous research questions, there are many potential variables relating to the educational process that could have been included in this study if the data had been available. These potential factors will be discussed later, in the Recommendations section.

Research Question 4

Six variables were analyzed to determine the effects of factors pertaining to teacher quality. Two variables dealt with the participation of the school system’s teachers in Tennessee’s Career Ladder program. Three variables described the degree levels attained by the district’s teachers as part of their professional training. Another variable in the study reported the average teacher's salary.

One relationship that was exhibited in this study, the relationship between mathematics achievement and percent of the district's teachers at upper Career Ladder Levels (II or III), revealed a pattern described earlier concerning fiscal factors. The independent variables CLPCT and CLEPCT demonstrated a relationship with the dependent variable Algebra 2 Mean Scale Score. These two factors, like the fiscal factors, only became an influencing factor at the highest level of mathematics courses included in this study. No significant relationship was found between Career Ladder status and any of the other seven dependent variables, whether *actual* or *value-added* measures.

The variable CLEPCT dealt with eligibility in the program is a default measure of teacher experience; to be eligible for the upper Career Ladder levels, one must have been a teacher for at least seven years. The Career Ladder Percent variable specifies the percent of a district's eligible teachers who successfully completed the assessment process for certification as a Career Level II or III teacher. The finding described here indicates that a more experienced teaching force and one with a higher degree of successful involvement in the Career Ladder evaluation process "pays off" at the higher levels of course-taking. It is quite possible that districts with less-experienced and less-involved teachers take care of the teaching resources for lower-level courses first, the ones that are more basic and affect more students. The districts and schools with a more robust supply of experienced, involved teachers can also provide adequate teaching resources for the upper-level courses, with a higher rate of success.

The relationship between mathematics achievement and percent of the district's teachers with advanced degrees was tested using the three independent variables BSPLUS, MSDEG, and MSPLUS.

For six of the dependent variables, statistically significant relationships were found; thus the null hypothesis was rejected. The largest correlation coefficient was found for the relationship of percentage of teachers with a Masters degree and Algebra 2 Mean Scale Score; this coefficient was .497 ($p < 0.01$). The smallest significant correlation coefficient was found for the relationship of percentage of teachers with a masters degree and Algebra 1 Mean Scale Score; this coefficient was .262 ($p < 0.05$).

For the dependent variables Geometry School Effect and Pre-Algebra Mean Scale Score, no statistically significant relationship was found; thus the null hypothesis was retained. The educational level of the district's teachers, as measured in this study, had no relationship with these two measures of student achievement in mathematics.

Among the three measures of teacher training, the variable concerned with number of teachers with a degree level higher than a masters degree (MSPLUS) added very little to an understanding of the relationship of this measure to mathematics achievement. The number of such teachers was relatively small and there is no indication that training above the masters level added to student performance.

The other two measures were responsible for the positive relationships to mathematics achievement that were exhibited. As would be expected by their definitions, the two variables bachelors plus and masters degrees (BSPLUS and MSDEG) are highly correlated to each other and exhibit similar patterns of relationships.

This study showed a relationship between mathematics achievement and average teacher salary in the district. For seven of the dependent variables, this study found statistically significant relationships. The significant correlation coefficients ranged from .522 ($p < 0.01$) for Algebra 2 Mean Scale Score to .271 ($p < 0.05$) for Pre-Algebra Mean Scale Score.

For the dependent variable Geometry School Effect, no statistically significant relationship was found with the average teacher salary. This lack of relationship and the scarcity of demonstrated relationships with the factors discussed earlier for the subject of Geometry illustrate the different nature of the course. Geometry does not fit on the common scale of measurement, nor does it exhibit the continuity of algebraic content and logic that characterizes the other three courses. In this study, student performance in Geometry “behaves differently.”

The average teacher’s salary variable is influenced by the experience and degree levels of the teachers and by the local salary scale. This statement is confirmed in this study by the correlation of this and two other independent variables, the variable measuring the percent of teachers eligible for the upper levels of the Career Ladder program and the percent of teachers with educational degree attainment higher than a bachelors degree. The correlation of average teacher’s salary and these two variables is .682 ($p < .01$) and .688 ($p < .01$) respectively.

This study also revealed that the average teacher’s salary variable is correlated with per pupil expenditures ($r = .774$, $p < .01$) and the percent of revenue from local revenue sources ($r = .884$, $p < .01$). Average teacher’s salary is also positively related to the ACT math independent variable ($r = .678$, $p < .01$), a measure of student achievement in mathematics. This positive correlation is in addition to the correlation with seven of the eight dependent variables mentioned earlier. These combinations of relationships among the study's variables lead to a discussion of the Opportunity-to-Learn model that served as a framework of the study.

Opportunity-to-Learn Framework

The framework of variables used in this study was comprised of three specific OTL categories: fiscal, educational process, and teacher quality (Odden, 1996); additionally, this study included a category of factors considered to be external to the control of the school and school system. An OTL framework was chosen for this study because establishing that Tennessee students have adequate “opportunity to learn” will become ethically, and perhaps legally, more important as the end-of-course tests of this study are replaced by Gateway tests (tests that are graduation requirements for all students). As the stakes increase, so do the technical requirements for the assessments and accountability systems.

As was mentioned in Chapter 2, there are alternatives to Odden's framework of OTL variables (Herman & Klein, 1997; Porter, 1995a) and other frameworks which make use of a similar wide range of variables, in frameworks not oriented to OTL. Any such combination of variables is likely to contribute to difficulties in determining optimal multiple linear regression equations, especially if the variables are characterized by a definition at a more global level (as opposed to a specific, direct measure), as was exhibited in the independent variables of this study. Independent variables measured at the district level (or, even worse, at the county level) are not precise enough to yield maximum research results.

If the independent variables of any framework exhibit collinearity, this would add to the difficulties of determining satisfactory regression equations. The correlational relationships among the independent variables of this study underscore problems associated with interpreting production function analyses and applying the findings in the policy arena. As mentioned in Chapter 2, most studies aimed at determining "Does money matter in education?" have been production function studies with linear regression (or other correlational techniques) as the standard analysis technique. Attributing credit or blame to any of the multiple factors at play in a complex environment such as the public education of students is problematic and most conclusions drawn from such analyses are challengeable.

This study has provided additional evidence to support the claim that certain factors are related to student achievement in mathematics. In this study it has been shown, for instance, that a district's average teacher's salary, student performance on other measures of mathematics achievement, the poverty rate of the students, the percent of revenue generated at the local level, the per pupil expenditure, and the degree level of the teachers are related to several of the measures associated with the TCAP end-of-course mathematics tests. Looked at in isolation, any of these (and other) factors could have statistically significant relationships. When examined collectively, a different picture emerges, because some variables may represent proxies for a general construct.

In this study, the regression analysis combined significant factors to determine relative influences from among these factors. No clear-cut patterns could be established. One factor that emerged as significantly related in most of the attempts of regression analysis was the ACT math score. This factor could justifiably be viewed as a summary score of mathematics achievement and one that takes place near the end of students' high school careers. It could also illustrate the overlapping of school, community, and student influences.

One way the ACT score reflects on the students involved is that those students who intend to pursue higher education are more likely to be motivated to enroll in higher-level classes and to study hard. The ACTMATH variable was significantly correlated to the course-taking factor ($r = .288$, $p < .05$). It was significantly correlated to the free- and reduced-lunch percentage ($r = -.589$, $p < .01$). It could be argued that students whose family income levels are low do not aspire to higher education, at least at the highly regarded institutions whose strict entrance requirements foster competitive pressure to excel on entrance exams.

Similarly, community expectations and the willingness to fund to the level of expectations revealed relationships that are pertinent to this discussion. Local revenue and per pupil expenditures are both significantly, positively correlated to the ACT math variable. Local revenue has a stronger correlation ($r = .633$, $p < .01$) than per pupil expenditures ($r = .420$, $p < .01$), however, indicating that the *willingness* and effort to fund locally is perhaps more important than the *amount* of funding.

Other confounding cause-and-effect reflections concern the role of the factor of the district's average teacher's salary. This factor has been shown here to be positively related to many of the measures of mathematics achievement. It is also related to local revenue, per pupil expenditure, and the enrollment of students in higher-level courses. It is often assumed that teachers with experience and advanced degrees are attracted to districts where they can earn a higher salary. It is also feasible that these teachers are attracted to districts with a motivated, ambitious student body, where supplies and equipment are provided appropriately, and that are located in communities that are attractive places to live and raise a family. Just how much of an influence teacher salary is, relative to these other types of issues, was beyond the scope of this study. This issue is debatable and illustrates the dynamic tension between research and policy.

Recommendations

Tennessee's assessment system should have as a goal the enabling of teachers and administrators to attend to the kind of capacity-building that will enable them to improve instruction and learning over time. Current practice among the state's efforts places an emphasis on accountability and reporting, with very little effort to inform practice. Policy decisions for the state should be based on well-designed research tailored to the needs and conditions in the state. The following recommendations are offered as possible approaches to reach these goals.

Research and Policy

Tennessee should initiate an intensive conversation among researchers, educators, and policymakers. This conversation should be designed to lead to an ambitious link between research and practice. An American Educational Research Association (AERA) policy analysis statement pointed out the absence of mutual accountability between research developers and (potential) research users. This AERA-ANALYSIS stated, "When researchers take little or no responsibility for making things work and practitioners eschew the development of explanatory systems for how and why things work or not, neither research nor practice benefits" (1999, paragraph 19).

The conversation among the groups should assess the quality of existing research, determine gaps in the knowledge base, and identify barriers to the dissemination and understanding of the research. There should be, as a result of the period of conversation, the establishment of a formalized partnership among the researchers and the practice and policymaking communities. The partnership should design, implement, monitor, and evaluate a large-scale research effort in Tennessee that truly has the ability to improve student learning and achievement. The research effort should not only be different in scope and quality; it should also use a variety of analytical approaches. Inductive and deductive approaches should be incorporated into designs that include multi-level analyses. Determining how to measure what needs to be measured, in an accurate, efficient manner, would be one of the challenges to be addressed.

Analysis of Relevant Factors

Policy makers and assessment designers should try to increase the number of variables measured and increase the specificity of the variables. This study examined patterns and relationships from among existing measures and reports. It revealed weaknesses among the measures and identified areas in which additional measures would be advantageous. Pedhazur (1982) discussed two types of errors, in addition to the multicollinearity discussed earlier, that are relevant to improving the measurement of variables in regression studies. Specification errors include the omission of relevant variables and the inclusion of irrelevant variables. Measurement errors include conceptual errors; an example is when a proxy variable is used for the variable of interest. Proxy variables are used, for convenience or availability, instead of the more properly specified and measured variable.

Additional Variables . A comprehensive study of the type proposed here would certainly include variables not available for this study. Especially valuable in this type of proposed study would be variables related to the educational process. Porter (1993) listed five categories of educational process variables. These categories included time spent for classroom instruction, high-school course-taking patterns, college entrance requirements (and Carnegie unit requirements), the enacted curriculum, the type of pedagogy and instruction used to teach a curriculum. Some of these are policy issues; such as the graduation and college entrance requirements. Others reflect the academic “press”, or push toward excellence, that is often mutually determined by the school, student, and community; measures of the course-taking pattern fit this characterization. However, most of the potential variables depend on getting additional information concerning actual practice in the classroom.

How time is allocated in the classroom, how the curriculum is shaped and enacted by a given teacher, and what types of instructional activities are actually used with students are factors which are not available, in current practice, in Tennessee assessment efforts. In most research efforts gathering this information is problematic. There are examples, however, that give insight into how this information could be gathered.

The National Assessment of Educational Progress (NAEP) has for some time made use of teacher and student surveys to gather information on classroom practices. The 1998 College Board study of the impact of block scheduling on AP scores made use of background questions. Research and development efforts are ongoing to improve the availability and accuracy of information on classroom practice. Aschbacher (1999) has described a national project designed to develop indicators of classroom practice to monitor and support school reform.

More Specific Variables. Other variables used in this study would have been more informative if they were more specifically related to the situation studied, or more focused. For instance, measures of teacher experience and degree level of the teachers were district-wide averages. If they reflected the experience and training of the teachers actually involved in the high school mathematics classrooms, additional (and perhaps better) information could have been analyzed.

The fact that certain fiscal measures that are reported by the state were not appropriate for this study has been discussed earlier in this chapter. Because measures of tax base per student and average income per resident are currently available only at the county level, they were not included in this study. It would be advantageous if these measures were more focused to the unit of analysis. In some states, there is interest in moving the unit of fiscal analysis from the district to the school level (Olson, 1998). Others have advocated the creation of a national framework for a student-level resource measure (Berne & Stiefel, 1997).

Conclusion

Tennessee has invested in the development and refinement of an accountability system. Next steps for the state include improvement in the design and collection of relevant data and a re-thinking of the roles of policymakers, educators, and researchers. A comprehensive system needs to be designed to meet several needs with the common goal of providing a better educational program. The assessment/research/accountability system should insure that policies imposed on the schools are based on sound research. It should insure that teachers and administrators engage in the kind of capacity-building that will enable them to improve instruction and learning. Proper communication channels should be established so that the different constituent groups can, and do, continue to work in a coordinated effort to improve in all areas. Turf battles and political issues should be subordinated to the greater issue of improving education. Dorn (1998) said that student performance should be the starting point of educational politics, not an occasion for political opportunism or crude comparisons and that accountability should encourage deeper discussion of educational problems. This “deeper discussion” must involve a coordinated effort with many viewpoints and approaches represented.

REFERENCES

Airasian, P. W., & Madaus, G. F. 1983. Linking testing and instruction: Policy issues. *Journal of Educational Measurement*, 20 (2), Summer: 103-118.

Aitkin, M., & Longford, N. 1986. Statistical modeling issues in school effectiveness studies (with discussion). *Journal of the Royal Statistical Society, Series A*, 149: 1-43. (Cited by Bock and Wolfe).

American Educational Research Association. 1999.*Gorillas in our midst: Emerging themes on how to improve education research.* AERA-ANALYSIS. August/September. 36 paragraphs. (Available at http://www.aera.net/gov/archive/n0899-06.htm).

Arnold, M. L. 1998. Three kinds of equity. *The American School Board Journal*, 185 (5), May: 34-36.

Aschbacher, P. R. 1999. *Developing indicators of classroom practice to monitor and support school reform*. CSE Technical Report 513. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST) Center for the Study of Evaluation (CSE). 91 p.

Ashton, P. 1996. Improving the preparation of teachers. *Educational Researcher*, 25 (9): 21-22.

Baglin, R. F. 1986. A problem in calculating group scores on norm-referenced tests. *Journal of Educational Measurement*, 23 (1), Spring: 57-68.

Barr, R., & Dreeben, R. 1977. Instruction in education. In L. S. Shulman (Ed.), *Review of research in education*. Vol. 5: 89-162. Itasco, IL: F. E. Peacock.

Bateson, D. J. 1990. Science achievement in semester and all-year courses. *Journal of Research in Science Teaching*, 27: 233-240.

Becker, D. F., & Forsyth, R. A. 1992. An empirical investigation of Thurston and IRT methods of scaling achievement tests. *Journal of Educational Measurement*, 29, Winter: 341-354.

Berne, R., & Stiefel, L. 1997. Student-level school resource measures. NCES 97-536. In W. J. Fowler, Jr. (Ed.), *Selected papers in school finance, 1995* : 62-79. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Bock, R., & Wolfe, R. 1996. Audit and review of the Tennessee Value-Added Assessment System (TVAAS): Final report. In R. Bock, R. Wolfe & T. Fisher (Eds.), *A review and analysis of the Tennessee Value-Added Assessment System*. Nashville: TN: Comptroller of the Treasury, Office of Education Accountability.

Bowles, S., & Levin, H. M. 1968. The determinants of scholastic achievement: An appraisal of some recent evidence. *Journal of Human Resources*, 3: 3-24.

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

Buckman, D. C., King, B. B., & Ryan, S. 1995. Block scheduling: A means to improve school climate. *NASSP Bulletin,* 79, May: 9-18.

Camilli, G. 1996. Standard errors in educational assessment: A policy analysis perspective. *Educational Evaluation and Policy Analysis*, 4 (4), March 15. (Available at http://olam.ed.asu.edu/epaa).

Carroll, J. B. 1963. A model of school learning. *Teachers College Record*, 64: 723-733.

Carroll, J. M. 1994. Organizing time to support learning. *The School Administrator*, 51 (3), March: 26-33.

Centra, J. A., & Potter, D. A. 1980. School and teacher effects: An interrelational model. *Review of Educational Research*, 50, Summer: 273-291.

Chaney, B., Burgdorf, K., & Atash, N. 1997. Influencing achievement through high school graduation requirements. *Educational Evaluation and Policy Analysis*, 19: 229-244.

Clune, W. 1994a. The cost and management of program adequacy: An emerging issue in education policy and finance. *Educational Policy*, 8: 365-375.

Clune, W. 1994b. The shift from equity to adequacy in school finance. *Educational Policy*, 8: 376-394.

Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfield, F., & York, R. 1966. *Equality of educational opportunity*. Washington, DC: U.S. Department of Health, Education, and Welfare; Office of Education.

The College Board. 1998. *Block schedules and student performance on AP examinations*. Research News, RN-03 . New York: College Entrance Examination Board. 12 p.

Cooley, W. 1993. The difficulty of the educational task: Implications for comparing student achievement in states, school districts, and schools. *ERS Spectrum*, 11 (3), Summer: 27-31.

Cronbach, L. J. 1963. Course improvement through evaluation. *Teachers College Record*, 64: 672-683.

Cronbach, L. J., & Furby, L. 1970. How do we measure "change" - or should we? *Psychological Bulletin*, 74: 68-80.

Darling-Hammond, L. 1992. Educational indicators and enlightened policy. *Educational Policy*, 6, September: 235-65.

Darling-Hammond, L. 1997. *Doing what matters most: Investing in quality teaching* . New York: National Commission on Teaching and America's Future.

Darling-Hammond, L. 1999. *Teacher quality and student achievement: A review of state policy evidence*. Document R-99-1. University of Washington, Center for the Study of Teaching and Policy.

Davenport, E. C., Jr., Davison, M. L., Kuang, H., Ding, S., Kim, S., & Kwak, N. 1998. High school mathematics course-taking by gender and ethnicity. *American Educational Research Journal*, 35, Fall: 497-514.

Dorn, S. 1998. The political legacy of school accountability systems. *Education Policy Analysis Archives,* 6 (1), January 2. (Available at http://olam.ed.asu.edu/epaa).

Epstein, J. 1984. Effects of parent involvement on change in student achievement in reading and math. Annual Meeting of the American Educational Research Association. New Orleans, LA.

Fetler, M. 1994. Carrot or stick? How do school performance reports work? *Education Policy Analysis Archives*, 2 (13), October 7. (Available at http://olam.ed.asu.edu/epaa).

Fisher, T. 1996. A review and analysis of the Tennessee Value-Added Assessment System. In R. Bock, R. Wolfe & T. Fisher (Eds.), *A review and analysis of the Tennessee Value-Added Assessment System* . Nashville: Tennessee Comptroller of the Treasury, Office of Education Accountability.

Fortune, J. C. 1993. Why production function analysis is irrelevant in policy deliberations concerning educational funding equity. *Education Policy Analysis Archives*, 1 (11), November 2. (Available at http://olam.ed.asu.edu/epaa).

Fuhrman, S. H. 1999. The new accountability. *CPRE Policy Briefs*, RB-27, January. Philadelphia: University of Pennsylvania, Consortium for Policy Research in Education.

Fullan, M. 1991. *The new meaning of educational change, 2nd ed.* . New York: Teachers College Press.

Garnier, H., Stein, J., & Jacobs, J. 1997. The process of dropping out of high school: A 19-year perspective. *American Educational Research Journal*, 34, Summer: 395-419.

Gintis, H. 1995. Review of Eric A. Hanushek's, Making Schools Work:. *Education Policy Analysis Archives*, 3 (7), March 31. (Available at http://olam.ed.asu.edu/epaa).

Goldstein, H. I. 1987. *Multilevel models in educational and social research*. London: Oxford University Press. (Cited by Bock and Wolfe).

Goodlad, J. 1984. *A place called school*. New York: McGraw-Hill.

Greenwald, R., Hedges, L. V., & Laine, R. D. 1996. The effect of school resources on student achievement. *Review of Educational Research*, 66, Fall: 361-396.

Grissmer, D. W., Kirby, S. N., Berends, M., & Williamson, S. 1994. *Student achievement and the changing American family*. MR-488-LE. Santa Monica, CA: RAND Institute on Education and Training. 131 pp.

Guskey, T., & Passaro, P. 1994. Teacher efficacy: A study of construct dimensions. *American Educational Research Journal*, 31: 627-643.

Haertel, G. D., Walberg, H. J., & Weinstein, T. 1983. Psychological models of educational performance: A theoretical synthesis of constructs. *Review of Educational Research*, 53: 75-92.

Hall, G. E., & Hord, S. M. 1987. *Change in schools: Facilitating the process*. Albany: State University of New York Press.

Hanushek, E. A. 1979. Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources*, 14, Summer: 351-88.

Hanushek, E. A. 1989. The impact of differential expenditures on school performance. *Educational Researcher*, 18, May: 45-51, 62.

Hanushek, E. A. 1994. Money might matter somewhere: A response to Hedges, Laine, and Greenwald. *Educational Researcher*, 23 (4): 5-8.

Hanushek, E. A. 1996. A more complete picture of school resource policies. *Review of Educational Research*, 66: 397-409.

Hanushek, E. A. 1998a. The evidence on class size. Occasional Paper Number 98-1. W. Allen Wallis Institute of Political Economy, University of Rochester. 40 pages.

Hanushek, E. A. 1998b. The evidence on class size: Summary of testimony. Hearings of: Subcommittee on Early Childhood, Youth, and Families; Committee on Education and the Workforce; U.S. House of Representatives.

Hawkins, E. F., Stancavage, F., & Dossey, J. A. 1998. *School policies affecting instruction in mathematics*. NCES 98-495. Washington, D.C.: National Center for Education Statistics.

Hedges, L. V., Laine, R. D., & Greenwald, R. 1994. Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher*, 23 (3), April: 5-14.

Herman, J. L., & Klein, D. C., D. 1997. *Assessing opportunity to learn: A California example*. CSE Technical Report 453 . Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST) Center for the Study of Evaluation (CSE). 19 p.

Hilton, T. L. 1971. *A study of intellectual growth and vocational development*. Washington, DC: Office of Education, U.S. Department of Health, Education, and Welfare.

Hirth, M. A., & Mitchell, R. C. 1995. An Indiana investigation of the impact of expenditures & socioeconomic status on student achievement: Does money matter? ERIC # ED389056. Annual Meeting of the National Council of Professors of Educational Administration. . Williamsburg, VA.

Hoffer, T., Moore, W., Quinn, P., & Suter, L. E. 1996. *High school seniors' instructional experiences in science and mathematics*. NCES 95-278. National Education Longitudinal Study of 1988. Washington, D. C.: National Center for Education Statistics.

Hoffer, T., Rasinski, K., & Moore, W. 1995. *Social background differences in high school mathematics and science course-taking and achievement*. NCES 95-206. Statistics in Brief. Washington, D. C.: National Center for Education Statistics. 17 pages.

Huberman, M., & Miles, M. 1984. *Innovation up close*. New York: Plenum.

Ingersoll, R. M. 1999. The problem of underqualified teachers in American secondary schools. *Educational Researcher*, 28 (2), March: 26-37.

Johnson, S. M. 1986. Incentives for teachers: What motivates, what matters. *Educational Administration Quarterly*, 22 (3): 54-79.

Jones, L. V. 1996. A history of the National Assessment of Educational Progress and some questions about its future. *Educational Researcher*, 25 (7), October: 15-22.

Kelley, C. 1998. The Kentucky school-based performance award program: School-level effects. *Educational Policy*, 12: 305-324.

Kolen, M. J. 1981. Comparison of traditional and item response theory methods for equating tests. *Journal of Educational Measurement*, 18 (1), Spring: 1-11.

Kramer, S. L. 1996. Block scheduling and high school mathematics instruction. *The Mathematics Teacher*, 89, December: 758-768.

Kramer, S. L. 1997. What we know about block scheduling and its effects on math instruction, Part I. *NASSP Bulletin*, 81 (586), February: 18-42.

Lamdin, D. J. 1996. Evidence of student attendance as an independent variable in education production functions. *Journal of Educational Research*, 89 (3), Jan-Feb 1996: 155-62.

Lara, A. V., & Medley, D. M. 1987. Effective teacher behavior as a function of learner ability. *Professional School Psychology*, 2 (1): 15-33.

Levin, B. 1993. Students and educational productivity. *Education Policy Analysis Archives*, 1 (5), May 4. (Available at http://olam.ed.asu.edu/epaa).

Levin, B. 1998. Criticizing the schools: Then and now. *Education Policy Analysis Archives*, 6 (16), August 29. (Available at http://olam.ed.asu.edu/epaa).

Linn, R. L. 1983. Testing and instruction: Links and distinctions. *Journal of Educational Measurement*, 20 (2), Summer: 179-189.

Linn, R. L. 1990. Has item response theory increased the validity of achievement test scores? *Applied Measurement in Education*, 3: 115-141.

Linn, R. L. 1993. Linking results of district assessments. *Applied Measurement in Education*, 6: 83-102.

Linn, R. L. 1994. Assessment-based reform: Challenges to educational measurement. First annual. William H. Angoff Memorial Lecture Series. Princeton, NJ: Educational Testing Service.

Linn, R. L. 2000. Assessments and accountability. *Educational Researcher*, 29 (2), March: 4-16.

Lockwood, S. L. 1995. Semesterizing the high school schedule: The impact on student achievement in algebra and geometry. *NASSP Bulletin*, 79 (575), December: 102-110.

Lord, F. 1956. The measurement of growth. *Educational and Psychological Measurement*, 16: 421-437.

LoVette, O. K. 1995. You ask, 'Why have school costs increased so much in the last 20 years?' *Phi Delta Kappan*, October: 169-172.

Mandeville, G. K., & Anderson, L. W. 1987. The stability of school effectiveness indices across grade levels and subject areas. *Journal of Educational Measurement*, 24 (3): 20-24.

Marsh, H., & Yeung, A. 1997. Coursework selection: Relations to academic self-concept and achievement. *American Educational Research Journal*, 34, Winter: 691-720.

McPartland, J. M., Epstein, J. L., Karweit, N., & Slavin, R. E. 1976. *Productivity of schools: Conceptual and methodological frameworks for research*. Baltimore: The Johns Hopkins University.

Mislevy, R. J. 1992. *Linking educational assessments: Concepts, issues, methods, and prospects*. ETS/CRESST technical report. Princeton, NJ: Educational Testing Service.

Mislevy, R. J. 1994. *Test theory reconceived*. CSE Technical Report 376. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Teaching (CRESST).

Moles, O. 1982. Syntheses of recent research on parent participation in children's education. *Educational Leadership*, 40: 44-47.

Monk, D. H. 1990. *Educational finance: An economic approach*. New York: McGraw-Hill.

Monk, D. H., & Rice, J. K. 1999. Modern education productivity research: Emerging implications for the financing of education. NCES 1999-334. In W. J. Fowler, Jr. (Ed.), *Selected papers in school finance, 1997-99*: 111-139. Washington, D.C.: U.S. Department of Education, National Center for Education Statistics.

Mosburg, S. 1996. *How money matters to school performance: Four points policymakers should know*. Northwest Regional Educational Laboratory Program Report. Portland, OR: Northwest Regional Educational Laboratory. 29 pages.

Muthén, B. O., Huang, L., Jo, B., Khoo, S., Goff, G. N., Novak, J., & Shih, J. 1995. *Opportunity-to-learn effects on achievement: Analytical aspects*. CSE Technical Report 407 . Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Teaching (CRESST).

Muthén, B. O., & Nelson, G. 1992. *Latent variable models for analysis of growth*. CSE Technical Report 345 . Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Teaching (CRESST).

National Center for Education Statistics. 1998a. *Revenues and expenditures for public elementary and secondary education: School year 1995-96*. NCES 98-205. Statistics in Brief. Washington, DC: U. S. Department of Education, Office of Educational Research and Improvement. 8 p.

National Center for Education Statistics. 1998b. *Toward better teaching: Professional development in 1993-94*. Washington, D.C.: U.S. Department of Education.

National Education Commission on Time and Learning. 1994. *Prisoners of time*. Washington, DC. 56 p.

Odden, A., & Kelley, C. 1997. *Paying teachers for what they know and do: New and smarter compensation strategies to improve schools.* Thousand Oaks, CA: Corwin Press.

Odden, A. 1996. Costs of measuring and providing opportunity to learn: Preliminary thoughts. In W. J. Fowler, Jr. (Ed.), *Selected papers in school finance, 1994*: 115-143. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Odden, A. 1998. Creating school finance policies that facilitate new goals. *CPRE Policy Briefs*, RB-26, September. Philadelphia: University of Pennsylvania, Consortium for Policy Research in Education.

Okagaki, L., & Frensch, P. A. 1998. Parenting and children's school achievement: A multi-ethnic perspective. *American Educational Research Journal*, 3 (1), Spring: 123-144.

Olson, L. 1998. Oregon to spend $2.9 million to track district spending. *Education Week*, February 25: 15.

O'Neil, H. F., Jr., Sugrue, B., Abedi, J., Baker, E. L., & Golan, S. 1997. *Final report of experimental studies on motivation and NAEP test performance*. CSE Technical Report 427 . Los Angeles: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing. 171 p.

O'Neil, J., & And Others. 1994. The states where money has been shown to make a difference in school achievement. ERIC # ED 3844099. Annual Meeting of the Mid-South Educational Research Association. . Nashville, TN.

Parrish, T. B., Matsumoto, C. S., & Fowler, W. J., Jr. 1995. *Disparities in public school district spending 1989-90*. NCES 95-300. Washington, DC: National Center for Education Statistics.

Pedhazur, E. J. 1982. *Multiple regression in behavioral research: Explanation and prediction* (2nd edition). Fort Worth, TX: Holt, Rhinehart and Winston.

Peterson, D. 1989. *Parental involvement in hte educational process*. Urbana, IL: ERIC Clearinghouse on Educational Management, University of Illinois.

Porter, A. C. 1993. Defining and measuring opportunity to learn. In *The debate on opportunity-to-learn standards*. Washington, DC: National Governors' Association.

Porter, A. C. 1995a. Defining and measuring opportunity to learn. In W. J. Fowler, Jr. (Ed.), *Developments in school finance*. Fiscal proceedings from the 1993 Annual NCES State Data Conference: 51-78. Washington, DC: National Center for Education Statistics.

Porter, A. C. 1995b. The uses and misuses of opportunity-to-learn standards. *Educational Researcher*, 24 (1): 21-27.

Raphael, D., & Wahlstrom, M. W. 1986. The semestered secondary school and student achievement: Results from the Second Ontario International Science Study. *Canadian Journal of Education*, 11 (2), Spring (May): 180-183.

Raphael, D., Wahlstrom, M. W., & McLean, L. D. 1986. Debunking the semestering myth. *Canadian Journal of Education*, 11 (1), Winter (March): 36-52.

Reid, W. M., Hierck, T., & Veregin, L. 1994. Measurable gains of block scheduling. *The School Administrator*, 51 (3), March: 32-33.

Rettig, M. D., & Canady, R. L. 1996. All around the block: The benefits and challenges of a non-traditional school schedule. *The School Administrator*, 53 (8), September.

Rock, D. A., & Pollack, J. M. 1995. *Mathematics course taking and gains in mathematics achievement*. Statistics in Brief; NCES 95-714. National Education Longitudinal Study of 1988. Washington, DC: National Center for Education Statistics. 15 pages.

Rock, D. A., Pollack, J. M., & Quinn, P. 1995. *Psychometric report for the NELS:88 base year through second follow-up*. NCES 95-382. National Education Longitudinal Study of 1988. Washington, DC: National Center for Education Statistics.

Rogosa, D. R., & Willett, J. B. 1983. Demonstrating the reliability of the difference score in the measurement of change. *Journal of Educational Measurement*, 20, Winter: 335-343.

Rothstein, R., & Miles, K. H. 1996. *Where's the money gone?* . Washington, DC: Economic Policy Institute (Available at http://epn.org/epi/epwtmg.html).

Sadowski, M. 1995. The numbers game yields simplistic answers on the link between spending and outcomes. *The Harvard Education Letter*, 11 (2), March/April: 1-4.

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

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

Schalock, M. D., Cowart, B., & Staebler, B. 1993. Teacher productivity revisited: Definition, theory, measurement, and application. *Journal of Personnel Evaluation in Education*, 7: 179-196.

Schalock, H. D., Schalock, M. D., Cowart, B., & Myton, D. 1993. Extending teacher assessment beyond knowledge and skills: An emerging focus on teacher accomplishments. *Journal of Personnel Evaluation in Education*, 7: 105-133.

Schmidt, W. H. 1883. Content biases in achievement tests. *Journal of Educational Measurement*, 20, Summer: 165-178.

Scott, L. A., Rock, D. A., Pollack, J. M., & Ingels, S. J. 1995. *Two years later: Cognitive gains and school transitions of NELS:88 eighth graders*. NCES 95-436. National Education Longitudinal Study of 1988. Washington, DC: National Center for Education Statistics.

Signer, B., Beasley, T. M., & Bauer, E. 1997. Interaction of ethnicity, mathematics achievement level, socioeconomic status, and gender among high school students' mathematics self-concepts. *Journal of Education for Students Placed At Risk*, 2: 377-393.

Sizer, T. R. 1992. *Horace's school: Redesigning the American high school*. Boston: Houghton Mifflin.

Smith, D. L., & McNelis, M. J. 1996. *A status report on alternative scheduling in Tennessee high schools*. Policy Practice Brief 9602. Memphis, TN: Center for Research in Educational Policy, The University of Memphis.

Spady, W. G. 1976. The impact of school resources on students. In W. H. Sewell, R. M. Hauser & D. L. Featherman (Eds.), *Schooling and achievement in American society*. New York: Academic Press.

Stake, R. 1998. Some comments on assessment in U.S. education. *Education Policy Analysis Archives,* 6 (14), July 21. (Available at http://olam.ed.asu.edu/epaa).

Stecher, B., & Barron, S. 1999. Test-based accountability: The perverse consequences of milepost testing. Annual Meeting of the American Educational Research Association. . Montreal, Canada.

Stohr-Hunt, P. M. 1996. An analysis of frequency of hands-on experience and science achievement. *Journal of Research in Science Teaching*, 33 (1), January: 101-109.

Tennessee Department of Education. 1996. *21st century schools program report card* . Nashville: Author.

Tennessee Department of Education. 1997. *Annual statistical report: For the scholastic year ending June 30, 1996* . Nashville: Tennessee Department of Education.

Tennessee State Board of Education, & Tennessee Higher Education Commission. 1997. *Annual joint report on kindergarten through higher education in Tennessee*. Nashville.

Tomlinson, T. M. 1994. Afterwords: Research and reform in the United States. In T. M. Tomlinson & A. C. Tuijnman (Eds.), *Education research and reform: An international perspective*: 173-181. Washington, DC: U.S. Department of Education.

U.S. General Accounting Office. 1995a. *School finance: Three states' experiences with equity in school funding*. GAO/HEHS-96-39. Washington, DC: U.S. General Accounting Office. 56 pages.

U.S. General Accounting Office. 1995b. *School finance: Trends in U. S. education spending*. GAO/HEHS-95-235. Washington, DC: U.S. General Accounting Office. 55 pages.

Viadero, D. 1998. U.S. seniors near bottom in world test. *Education Week*, March 4.

Walberg, H. 1984. Families as partners in educational productivity. *Phi Delta Kappan*, 65: 397-400.

Wang, M. C., Haertel, G. D., & Walberg, H. J. 1993. Toward a knowledge base for school learning. *Review of Educational Research*, 63 (3).

Weiss, I., Lawrenz, F., & Queitzsch, M. L. 1996. The elementary and secondary learning environment. In L. E. Suter (Ed.), *Indicators of science and mathematics education 1995*. NSF 96-52. REC Indicators Series. Arlington, VA: Division of Research, Evaluation, and Communication, Directorate for Education and Human Resources, National Science Foundation.

Wenglinsky, H. 1997. *When money matters: How educational expenditures improve student performance and how they don't*. Princeton, NJ: Policy Information Center, Educational Testing Service.

Wenglinsky, H. 1998. School district expenditures, school resources and student achievement: Modeling the production function. In W. J. Fowler, Jr. (Ed.), *Developments in school finance, 1997*: 99-120. Washington, DC: National Center for Education Statistics.

Williamson, G. L., Appelbaum, M., & Epanchin, A. 1991. Longitudinal analyses of academic achievement. *Journal of Educational Measurement*, 28 (1), Spring: 61-76.

Willms, J. D. 1992. *Monitoring school performance: A guide for educators* . Washington, DC: Falmer.

Willms, J. D., & Raudenbush, S. W. 1989. A longitudinal hierarchical model for estimating school effects and their stability. *Journal of Educational Measurement*, 26, Fall: 209-232.

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

Yen, W. M. 1986. The choice of scale for educational measurement: An IRT perspective. *Journal of Educational Measurement*, 23: 299-325.

Yoon, B., & Resnick, L. B. 1998. *Instructional validity, opportunity to learn and equity: New Standards Examinations for the California Mathematics Renaissance*. CSE Technical Report 484 . Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST) Center for the Study of Evaluation (CSE). 27 p.

Young, B. A., & Smith, T. M. 1997. *The social context of education*. NCES 97-981. Findings from the Condition of Education 1997, vol. 10. Washington, DC: U. S. Department of Education, National Center for Educational Statistics. 38 p.

Zimmerman, D. W., & Williams, R. 1982. Gain scores in research can be highly reliable. *Journal of Educational Measurement*, 19: 149-154.

Table A-1 Descriptive statistics for independent variables

	N		Mean	Median	Std. Deviation	Range	Minimum	Maximum
	Valid	Missing						
Free & Reduced Lunch	65	0	43.390%	39.730%	13.780%	62.8%	18.7%	81.5%
Special Education	65	0	20.093%	19.199%	4.542%	21.9%	10.9%	32.9%
Per Pupil Expenditure	65	0	$4,530.51	$4,408.12	$713.31	$3,233	$3,561	$6,794
Administrative spending	65	0	8.928%	8.745%	1.528%	8.8%	5.2%	14.0%
Regular Education spending	65	0	54.906%	54.404%	4.273%	19.9%	44.4%	64.4%
Local Revenue	65	0	31.153%	27.067%	11.897%	52.5%	11.3%	63.8%
Attendance Rate	65	0	93.394%	93.685%	1.809%	8.2%	89.1%	97.3%
Dropout Rate	65	0	3.603%	3.238%	2.193%	12.2%	.0%	12.2%
Math Gr.8 National Percentile	65	0	60.51	62.00	11.34	60	28	88
ACT Math	65	0	18.546	18.300	1.483	6.4	15.5	21.9
Competency Test - percent passing	65	0	74.83%	75.00%	13.10%	62%	37%	99%
Pre-Algebra Mean Predicted Score	57	8	457.879	457.700	8.352	44.1	439.5	483.6
Algebra 1 Mean Predicted Score	65	0	503.354	502.800	10.444	48.2	477.5	525.7
Algebra 2 Mean Predicted Score	65	0	536.591	536.900	10.645	44.5	513.3	557.8
Geometry Mean Predicted Score	64	1	503.892	502.450	9.805	54.2	479.8	534.0
Course-taking factor	65	0	1.0584	1.0269	.4176	2.49	.26	2.75
Career Ladder Percent	65	0	23.443%	23.106%	8.552%	44.6%	5.4%	50.0%
Career Ladder Eligible Percent	65	0	72.436%	73.116%	7.668%	33.7%	57.1%	90.8%
Bachelors plus	65	0	49.6802%	48.1308%	9.6528%	46.38%	26.19%	72.57%
Masters degree	65	0	36.0097%	35.7143%	7.4122%	32.93%	21.98%	54.91%
Masters plus	65	0	13.6705%	12.7072%	6.4879%	31.66%	2.38%	34.04%
Average Teacher's Salary	65	0	$30,537	$29,152	$3,511.51	$15,035	$26,419	$41,453

Table A-2 Block schedule school

	Frequency	Percent
No	53	81.5
Yes	12	18.5
Total	65	100.0

Table A-3 Descriptive statistics for dependent variables

	N							
	Valid	Missing	Mean	Median	Std. Deviation	Range	Minimum	Maximum
Pre-Algebra Mean Scale Score	57	8	457.489	457.600	16.688	83.6	429.0	512.6
Pre-Algebra School Effect	57	8	-.304	-1.400	11.344	53.4	-28.4	25.0
Algebra 1 Mean Scale Score	65	0	502.515	501.800	18.581	85.5	459.7	545.2
Algebra 1 School Effect	65	0	-.798	-1.200	12.861	60.0	-26.3	33.7
Algebra 2 Mean Scale Score	65	0	535.365	533.500	22.918	114.1	484.3	598.4
Algebra 2 School Effect	65	0	-1.086	-4.100	16.501	68.3	-27.6	40.7
Geometry Mean Scale Score	64	1	506.303	504.250	17.832	82.5	464.3	546.8
Geometry School Effect	64	1	2.300	1.050	12.609	62.5	-25.4	37.1

Table A-4 Correlations of external factors and mean scale scores

		Free & Reduced Lunch	Special Education	Pre-Algebra Mean Scale Score	Algebra 1 Mean Scale Score	Algebra 2 Mean Scale Score	Geometry Mean Scale Score
Pearson Correlation	Free & Reduced Lunch	1.000	.056	-.385**	-.301*	-.362**	-.352**
	Special Education	.056	1.000	.121	.031	-.071	.117
	Pre-Algebra Mean Scale Score	-.385**	.121	1.000	.605**	.336*	.487**
	Algebra 1 Mean Scale Score	-.301*	.031	.605**	1.000	.372**	.503**
	Algebra 2 Mean Scale Score	-.362**	-.071	.336*	.372**	1.000	.367**
	Geometry Mean Scale Score	-.352**	.117	.487**	.503**	.367**	1.000
Sig. (2-tailed)	Free & Reduced Lunch	.	.657	.003	.015	.003	.004
	Special Education	.657	.	.371	.806	.575	.357
	Pre-Algebra Mean Scale Score	.003	.371	.	.000	.011	.000
	Algebra 1 Mean Scale Score	.015	.806	.000	.	.002	.000
	Algebra 2 Mean Scale Score	.003	.575	.011	.002	.	.003
	Geometry Mean Scale Score	.004	.357	.000	.000	.003	.
N	Free & Reduced Lunch	65	65	57	65	65	64
	Special Education	65	65	57	65	65	64
	Pre-Algebra Mean Scale Score	57	57	57	57	57	57
	Algebra 1 Mean Scale Score	65	65	57	65	65	64
	Algebra 2 Mean Scale Score	65	65	57	65	65	64
	Geometry Mean Scale Score	64	64	57	64	64	64

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table A-5 Correlations of external factors and value-added scores

		Free & Reduced Lunch	Special Education	Pre-Algebra School Effect	Algebra 1 School Effect	Algebra 2 School Effect	Geometry School Effect
Pearson Correlation	Free & Reduced Lunch	1.000	.056	-.319*	-.221	-.215	-.076
	Special Education	.056	1.000	.006	.082	.001	.075
	Pre-Algebra School Effect	-.319*	.006	1.000	.609**	.432**	.375**
	Algebra 1 School Effect	-.221	.082	.609**	1.000	.509**	.530**
	Algebra 2 School Effect	-.215	.001	.432**	.509**	1.000	.284*
	Geometry School Effect	-.076	.075	.375**	.530**	.284*	1.000
Sig. (2-tailed)	Free & Reduced Lunch	.	.657	.015	.077	.085	.552
	Special Education	.657	.	.965	.518	.991	.554
	Pre-Algebra School Effect	.015	.965	.	.000	.001	.004
	Algebra 1 School Effect	.077	.518	.000	.	.000	.000
	Algebra 2 School Effect	.085	.991	.001	.000	.	.023
	Geometry School Effect	.552	.554	.004	.000	.023	.
N	Free & Reduced Lunch	65	65	57	65	65	64
	Special Education	65	65	57	65	65	64
	Pre-Algebra School Effect	57	57	57	57	57	57
	Algebra 1 School Effect	65	65	57	65	65	64
	Algebra 2 School Effect	65	65	57	65	65	64
	Geometry School Effect	64	64	57	64	64	64

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Table A-6 Correlations of fiscal factors and mean scale scores

		Per Pupil Expenditure	Administrative spending	Regular Education spending	Local Revenue	Pre-Algebra Mean Scale Score	Algebra 1 Mean Scale Score	Algebra 2 Mean Scale Score	Geometry Mean Scale Score
Pearson Correlation	Per Pupil Expenditure	1.000	.084	.110	.648**	.127	.194	.337**	.263*
	Administrative spending	.084	1.000	.128	.281*	.174	.113	.046	.128
	Regular Education spending	.110	.128	1.000	.441**	.135	.163	.390**	.192
	Local Revenue	.648**	.281*	.441**	1.000	.229	.171	.478**	.351**
	Pre-Algebra Mean Scale Score	.127	.174	.135	.229	1.000	.605**	.336*	.487**
	Algebra 1 Mean Scale Score	.194	.113	.163	.171	.605**	1.000	.372**	.503**
	Algebra 2 Mean Scale Score	.337**	.046	.390**	.478**	.336*	.372**	1.000	.367**
	Geometry Mean Scale Score	.263*	.128	.192	.351**	.487**	.503**	.367**	1.000
Sig. (2-tailed)	Per Pupil Expenditure	.	.508	.384	.000	.348	.122	.006	.035
	Administrative spending	.508	.	.309	.023	.195	.368	.716	.314
	Regular Education spending	.384	.309	.	.000	.318	.194	.001	.128
	Local Revenue	.000	.023	.000	.	.087	.173	.000	.004
	Pre-Algebra Mean Scale Score	.348	.195	.318	.087	.	.000	.011	.000
	Algebra 1 Mean Scale Score	.122	.368	.194	.173	.000	.	.002	.000
	Algebra 2 Mean Scale Score	.006	.716	.001	.000	.011	.002	.	.003
	Geometry Mean Scale Score	.035	.314	.128	.004	.000	.000	.003	.
N	Per Pupil Expenditure	65	65	65	65	57	65	65	64
	Administrative spending	65	65	65	65	57	65	65	64
	Regular Education spending	65	65	65	65	57	65	65	64
	Local Revenue	65	65	65	65	57	65	65	64
	Pre-Algebra Mean Scale Score	57	57	57	57	57	57	57	57
	Algebra 1 Mean Scale Score	65	65	65	65	57	65	65	64
	Algebra 2 Mean Scale Score	65	65	65	65	57	65	65	64
	Geometry Mean Scale Score	64	64	64	64	57	64	64	64

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table A-7 Correlations of fiscal factors and value-added scores

		Per Pupil Expenditure	Administrative spending	Regular Education spending	Local Revenue	Pre-Algebra School Effect	Algebra 1 School Effect	Algebra 2 School Effect	Geometry School Effect
Pearson Correlation	Per Pupil Expenditure	1.000	.084	.110	.648 **	.238	.280 *	.305 *	.170
	Administrative spending	.084	1.000	.128	.281 *	.178	-.016	-.010	.026
	Regular Education spending	.110	.128	1.000	.441 **	.148	.115	.223	.086
	Local Revenue	.648 **	.281 *	.441 **	1.000	.321 *	.207	.319 **	.160
	Pre-Algebra School Effect	.238	.178	.148	.321 *	1.000	.609 **	.432 **	.375 **
	Algebra 1 School Effect	.280 *	-.016	.115	.207	.609 **	1.000	.509 **	.530 **
	Algebra 2 School Effect	.305 *	-.010	.223	.319 **	.432 **	.509 **	1.000	.284 *
	Geometry School Effect	.170	.026	.086	.160	.375 **	.530 **	.284 *	1.000
Sig. (2-tailed)	Per Pupil Expenditure	.	.508	.384	.000	.075	.024	.014	.179
	Administrative spending	.508	.	.309	.023	.185	.902	.936	.838
	Regular Education spending	.384	.309	.	.000	.271	.360	.074	.499
	Local Revenue	.000	.023	.000	.	.015	.098	.010	.206
	Pre-Algebra School Effect	.075	.185	.271	.015	.	.000	.001	.004
	Algebra 1 School Effect	.024	.902	.360	.098	.000	.	.000	.000
	Algebra 2 School Effect	.014	.936	.074	.010	.001	.000	.	.023
	Geometry School Effect	.179	.838	.499	.206	.004	.000	.023	.
N	Per Pupil Expenditure	65	65	65	65	57	65	65	64
	Administrative spending	65	65	65	65	57	65	65	64
	Regular Education spending	65	65	65	65	57	65	65	64
	Local Revenue	65	65	65	65	57	65	65	64
	Pre-Algebra School Effect	57	57	57	57	57	57	57	57
	Algebra 1 School Effect	65	65	65	65	57	65	65	64
	Algebra 2 School Effect	65	65	65	65	57	65	65	64
	Geometry School Effect	64	64	64	64	57	64	64	64

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table A-8 Correlations of educational process factors and mean scale scores

		Attendance Rate	Dropout Rate	Math Gr.8 National Percentile	ACT Math	Competency Test - percent passing	Pre-Algebra Mean Predicted Score	Algebra 1 Mean Predicted Score	Algebra 2 Mean Predicted Score	Geometry Mean Predicted Score	Course-taking factor	Pre-Algebra Mean Scale Score	Algebra 1 Mean Scale Score	Algebra 2 Mean Scale Score	Geometry Mean Scale Score
Pearson Correlation	Attendance Rate	1.000	-.379**	.430**	.270*	.476**	.127	.338**	.317**	.267*	.214	.172	.292*	.145	.054
	Dropout Rate	-.379**	1.000	-.334**	-.255*	-.340**	-.030	-.253*	-.380**	-.277*	-.195	.005	-.102	-.094	-.109
	Math Gr.8 National Percentile	.430**	-.334**	1.000	.530**	.568**	.368**	.385**	.467**	.483**	.279*	.349**	.335**	.272*	.275*
	ACT Math	.270*	-.255*	.530**	1.000	.349**	.142	.171	.544**	.529**	.288*	.379**	.471**	.664**	.556**
	Competency Test - percent passing	.476**	-.340**	.568**	.349**	1.000	.307*	.438**	.288*	.350**	.255*	.464**	.320**	.149	.110
	Pre-Algebra Mean Predicted Score	.127	-.030	.368**	.142	.307*	1.000	.359**	.202	.499**	-.190	.645**	.283*	.143	.401**
	Algebra 1 Mean Predicted Score	.338**	-.253*	.385**	.171	.438**	.359**	1.000	.328**	.433**	.080	.340**	.696**	.098	.165
	Algebra 2 Mean Predicted Score	.317**	-.380**	.467**	.544**	.288*	.202	.328**	1.000	.569**	.257*	.096	.243	.673**	.292*
	Geometry Mean Predicted Score	.267*	-.277*	.483**	.529**	.350**	.499**	.433**	.569**	1.000	.185	.372**	.449**	.367**	.644**
	Course-taking factor	.214	-.195	.279*	.288*	.255*	-.190	.080	.257*	.185	1.000	-.015	-.024	.159	.015
	Pre-Algebra Mean Scale Score	.172	.005	.349**	.379**	.464**	.645**	.340**	.096	.372**	-.015	1.000	.605**	.336*	.487**
	Algebra 1 Mean Scale Score	.292*	-.102	.335**	.471**	.320**	.283*	.696**	.243	.449**	-.024	.605**	1.000	.372**	.503**
	Algebra 2 Mean Scale Score	.145	-.094	.272*	.664**	.149	.143	.098	.673**	.367**	.159	.336*	.372**	1.000	.367**
	Geometry Mean Scale Score	.054	-.109	.275*	.556**	.110	.401**	.165	.292*	.644**	.015	.487**	.503**	.367**	1.000
Sig. (2-tailed)	Attendance Rate	.	.002	.000	.030	.000	.346	.006	.010	.033	.087	.200	.018	.250	.670
	Dropout Rate	.002	.	.007	.040	.006	.827	.042	.002	.027	.119	.972	.417	.457	.391
	Math Gr.8 National Percentile	.000	.007	.	.000	.000	.005	.002	.000	.000	.024	.008	.006	.029	.028
	ACT Math	.030	.040	.000	.	.004	.291	.173	.000	.000	.020	.004	.000	.000	.000
	Competency Test - percent passing	.000	.006	.000	.004	.	.020	.000	.020	.005	.040	.000	.009	.238	.386
	Pre-Algebra Mean Predicted Score	.346	.827	.005	.291	.020	.	.006	.131	.000	.157	.000	.033	.289	.002
	Algebra 1 Mean Predicted Score	.006	.042	.002	.173	.000	.006	.	.008	.000	.526	.010	.000	.438	.193
	Algebra 2 Mean Predicted Score	.010	.002	.000	.000	.020	.131	.008	.	.000	.039	.478	.051	.000	.019
	Geometry Mean Predicted Score	.033	.027	.000	.000	.005	.000	.000	.000	.	.144	.004	.000	.003	.000
	Course-taking factor	.087	.119	.024	.020	.040	.157	.526	.039	.144	.	.911	.852	.204	.906
	Pre-Algebra Mean Scale Score	.200	.972	.008	.004	.000	.000	.010	.478	.004	.911	.	.000	.011	.000
	Algebra 1 Mean Scale Score	.018	.417	.006	.000	.009	.033	.000	.051	.000	.852	.000	.	.002	.000
	Algebra 2 Mean Scale Score	.250	.457	.029	.000	.238	.289	.438	.000	.003	.204	.011	.002	.	.003
	Geometry Mean Scale Score	.670	.391	.028	.000	.386	.002	.193	.019	.000	.906	.000	.000	.003	.
N	Attendance Rate	65	65	65	65	65	57	65	65	64	65	57	65	65	64
	Dropout Rate	65	65	65	65	65	57	65	65	64	65	57	65	65	64
	Math Gr.8 National Percentile	65	65	65	65	65	57	65	65	64	65	57	65	65	64
	ACT Math	65	65	65	65	65	57	65	65	64	65	57	65	65	64
	Competency Test - percent passing	65	65	65	65	65	57	65	65	64	65	57	65	65	64
	Pre-Algebra Mean Predicted Score	57	57	57	57	57	57	57	57	57	57	57	57	57	57
	Algebra 1 Mean Predicted Score	65	65	65	65	65	57	65	65	64	65	57	65	65	64
	Algebra 2 Mean Predicted Score	65	65	65	65	65	57	65	65	64	65	57	65	65	64
	Geometry Mean Predicted Score	64	64	64	64	64	57	64	64	64	64	57	64	64	64
	Course-taking factor	65	65	65	65	65	57	65	65	64	65	57	65	65	64
	Pre-Algebra Mean Scale Score	57	57	57	57	57	57	57	57	57	57	57	57	57	57
	Algebra 1 Mean Scale Score	65	65	65	65	65	57	65	65	64	65	57	65	65	64
	Algebra 2 Mean Scale Score	65	65	65	65	65	57	65	65	64	65	57	65	65	64
	Geometry Mean Scale Score	64	64	64	64	64	57	64	64	64	64	57	64	64	64

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table A-9 Correlations of educational process factors and value-added scores

		Attendance Rate	Dropout Rate	Math Gr.8 National Percentile	ACT Math	Competency Test - percent passing	Pre-Algebra Mean Predicted Score	Algebra 1 Mean Predicted Score	Algebra 2 Mean Predicted Score	Geometry Mean Predicted Score	Course-taking factor	Pre-Algebra School Effect	Algebra 1 School Effect	Algebra 2 School Effect	Geometry School Effect
Pearson Correlation	Attendance Rate	1.000	-.379**	.430**	.270*	.476**	.127	.338**	.317**	.267*	.214	.138	.139	-.003	-.111
	Dropout Rate	-.379**	1.000	-.334**	-.255*	-.340**	-.030	-.253*	-.380**	-.277*	-.195	.012	.052	.099	.028
	Math Gr.8 National Percentile	.430**	-.334**	1.000	.530**	.568**	.368**	.385**	.467**	.483**	.279*	.227	.160	.080	.016
	ACT Math	.270*	-.255*	.530**	1.000	.349**	.142	.171	.544**	.529**	.288*	.415**	.510**	.534**	.344**
	Competency Test - percent passing	.476**	-.340**	.568**	.349**	1.000	.307*	.438**	.288*	.350**	.255*	.404**	.106	.023	-.096
	Pre-Algebra Mean Predicted Score	.127	-.030	.368**	.142	.307*	1.000	.359**	.202	.499**	-.190	.198	.113	.059	.163
	Algebra 1 Mean Predicted Score	.338**	-.253*	.385**	.171	.438**	.359**	1.000	.328**	.433**	.080	.210	.184	-.064	-.093
	Algebra 2 Mean Predicted Score	.317**	-.380**	.467**	.544**	.288*	.202	.328**	1.000	.569**	.257*	.008	.085	.276*	-.015
	Geometry Mean Predicted Score	.267*	-.277*	.483**	.529**	.350**	.499**	.433**	.569**	1.000	.185	.168	.279*	.142	.125
	Course-taking factor	.214	-.195	.279*	.288*	.255*	-.190	.080	.257*	.185	1.000	.102	-.099	.064	-.114
	Pre-Algebra School Effect	.138	.012	.227	.415**	.404**	.198	.210	.008	.168	.102	1.000	.609**	.432**	.375**
	Algebra 1 School Effect	.139	.052	.160	.510**	.106	.113	.184	.085	.279*	-.099	.609**	1.000	.509**	.530**
	Algebra 2 School Effect	-.003	.099	.080	.534**	.023	.059	-.064	.276*	.142	.064	.432**	.509**	1.000	.284*
	Geometry School Effect	-.111	.028	.016	.344**	-.096	.163	-.093	-.015	.125	-.114	.375**	.530**	.284*	1.000
Sig. (2-tailed)	Attendance Rate	.	.002	.000	.030	.000	.346	.006	.010	.033	.087	.305	.269	.984	.385
	Dropout Rate	.002	.	.007	.040	.006	.827	.042	.002	.027	.119	.931	.681	.431	.825
	Math Gr.8 National Percentile	.000	.007	.	.000	.000	.005	.002	.000	.000	.024	.090	.202	.527	.899
	ACT Math	.030	.040	.000	.	.004	.291	.173	.000	.000	.020	.001	.000	.000	.005
	Competency Test - percent passing	.000	.006	.000	.004	.	.020	.000	.020	.005	.040	.002	.403	.856	.449
	Pre-Algebra Mean Predicted Score	.346	.827	.005	.291	.020	.	.006	.131	.000	.157	.140	.405	.661	.225
	Algebra 1 Mean Predicted Score	.006	.042	.002	.173	.000	.006	.	.008	.000	.526	.117	.143	.614	.466
	Algebra 2 Mean Predicted Score	.010	.002	.000	.000	.020	.131	.008	.	.000	.039	.952	.502	.026	.905
	Geometry Mean Predicted Score	.033	.027	.000	.000	.005	.000	.000	.000	.	.144	.212	.025	.262	.324
	Course-taking factor	.087	.119	.024	.020	.040	.157	.526	.039	.144	.	.448	.431	.611	.371
	Pre-Algebra School Effect	.305	.931	.090	.001	.002	.140	.117	.952	.212	.448	.	.000	.001	.004
	Algebra 1 School Effect	.269	.681	.202	.000	.403	.405	.143	.502	.025	.431	.000	.	.000	.000
	Algebra 2 School Effect	.984	.431	.527	.000	.856	.661	.614	.026	.262	.611	.001	.000	.	.023
	Geometry School Effect	.385	.825	.899	.005	.449	.225	.466	.905	.324	.371	.004	.000	.023	.
N	Attendance Rate	65	65	65	65	65	57	65	65	64	65	57	65	65	64
	Dropout Rate	65	65	65	65	65	57	65	65	64	65	57	65	65	64
	Math Gr.8 National Percentile	65	65	65	65	65	57	65	65	64	65	57	65	65	64
	ACT Math	65	65	65	65	65	57	65	65	64	65	57	65	65	64
	Competency Test - percent passing	65	65	65	65	65	57	65	65	64	65	57	65	65	64
	Pre-Algebra Mean Predicted Score	57	57	57	57	57	57	57	57	57	57	57	57	57	57
	Algebra 1 Mean Predicted Score	65	65	65	65	65	57	65	65	64	65	57	65	65	64
	Algebra 2 Mean Predicted Score	65	65	65	65	65	57	65	65	64	65	57	65	65	64
	Geometry Mean Predicted Score	64	64	64	64	64	57	64	64	64	64	57	64	64	64
	Course-taking factor	65	65	65	65	65	57	65	65	64	65	57	65	65	64
	Pre-Algebra School Effect	57	57	57	57	57	57	57	57	57	57	57	57	57	57
	Algebra 1 School Effect	65	65	65	65	65	57	65	65	64	65	57	65	65	64
	Algebra 2 School Effect	65	65	65	65	65	57	65	65	64	65	57	65	65	64
	Geometry School Effect	64	64	64	64	64	57	64	64	64	64	57	64	64	64

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table A-10 ANOVA (block schedule)

		Sum of Squares	df	Mean Square	F	Sig.
Pre-Algebra Mean Scale Score	Between Groups	.752	1	.752	.003	.959
	Within Groups	15594.861	55	283.543		
	Total	15595.614	56			
Pre-Algebra School Effect	Between Groups	53.725	1	53.725	.413	.523
	Within Groups	7152.695	55	130.049		
	Total	7206.419	56			
Algebra 1 Mean Scale Score	Between Groups	864.742	1	864.742	2.566	.114
	Within Groups	21230.462	63	336.991		
	Total	22095.205	64			
Algebra 1 School Effect	Between Groups	298.223	1	298.223	1.826	.181
	Within Groups	10287.307	63	163.291		
	Total	10585.530	64			
Algebra 2 Mean Scale Score	Between Groups	12.422	1	12.422	.023	.879
	Within Groups	33602.367	63	533.371		
	Total	33614.789	64			
Algebra 2 School Effect	Between Groups	5.056	1	5.056	.018	.893
	Within Groups	17420.501	63	276.516		
	Total	17425.558	64			
Geometry Mean Scale Score	Between Groups	87.076	1	87.076	.271	.605
	Within Groups	19946.003	62	321.710		
	Total	20033.079	63			
Geometry School Effect	Between Groups	23.386	1	23.386	.145	.705
	Within Groups	9992.594	62	161.171		
	Total	10015.980	63			

Table A-11 Correlations of teacher factors and mean scale scores

		Career Ladder Percent	Career Ladder Eligible Percent	Bachelors plus	Masters degree	Masters plus	Average Teacher's Salary	Pre-Algebra Mean Scale Score	Algebra 1 Mean Scale Score	Algebra 2 Mean Scale Score	Geometry Mean Scale Score
Pearson Correlation	Career Ladder Percent	1.000	.438**	.412**	.206	.377**	.487**	.216	.227	.248*	.165
	Career Ladder Eligible Percent	.438**	1.000	.540**	.445**	.295*	.682**	.165	.201	.407**	.110
	Bachelors plus	.412**	.540**	1.000	.741**	.641**	.688**	.195	.280*	.423**	.308*
	Masters degree	.206	.445**	.741**	1.000	-.040	.668**	.129	.262*	.497**	.309*
	Masters plus	.377**	.295*	.641**	-.040	1.000	.260*	.132	.118	.062	.100
	Average Teacher's Salary	.487**	.682**	.688**	.668**	.260*	1.000	.271*	.279*	.522**	.377**
	Pre-Algebra Mean Scale Score	.216	.165	.195	.129	.132	.271*	1.000	.605**	.336*	.487**
	Algebra 1 Mean Scale Score	.227	.201	.280*	.262*	.118	.279*	.605**	1.000	.372**	.503**
	Algebra 2 Mean Scale Score	.248*	.407**	.423**	.497**	.062	.522**	.336*	.372**	1.000	.367**
	Geometry Mean Scale Score	.165	.110	.308*	.309*	.100	.377**	.487**	.503**	.367**	1.000
Sig. (2-tailed)	Career Ladder Percent	.	.000	.001	.099	.002	.000	.107	.069	.047	.191
	Career Ladder Eligible Percent	.000	.	.000	.000	.017	.000	.220	.109	.001	.387
	Bachelors plus	.001	.000	.	.000	.000	.000	.147	.024	.000	.013
	Masters degree	.099	.000	.000	.	.751	.000	.338	.035	.000	.013
	Masters plus	.002	.017	.000	.751	.	.036	.327	.349	.626	.434
	Average Teacher's Salary	.000	.000	.000	.000	.036	.	.042	.025	.000	.002
	Pre-Algebra Mean Scale Score	.107	.220	.147	.338	.327	.042	.	.000	.011	.000
	Algebra 1 Mean Scale Score	.069	.109	.024	.035	.349	.025	.000	.	.002	.000
	Algebra 2 Mean Scale Score	.047	.001	.000	.000	.626	.000	.011	.002	.	.003
	Geometry Mean Scale Score	.191	.387	.013	.013	.434	.002	.000	.000	.003	.
N	Career Ladder Percent	65	65	65	65	65	65	57	65	65	64
	Career Ladder Eligible Percent	65	65	65	65	65	65	57	65	65	64
	Bachelors plus	65	65	65	65	65	65	57	65	65	64
	Masters degree	65	65	65	65	65	65	57	65	65	64
	Masters plus	65	65	65	65	65	65	57	65	65	64
	Average Teacher's Salary	65	65	65	65	65	65	57	65	65	64
	Pre-Algebra Mean Scale Score	57	57	57	57	57	57	57	57	57	57
	Algebra 1 Mean Scale Score	65	65	65	65	65	65	57	65	65	64
	Algebra 2 Mean Scale Score	65	65	65	65	65	65	57	65	65	64
	Geometry Mean Scale Score	64	64	64	64	64	64	57	64	64	64

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table A-12 Correlations of teacher factors and value-added scores

		Career Ladder Percent	Career Ladder Eligible Percent	Bachelors plus	Masters degree	Masters plus	Average Teacher's Salary	Pre-Algebra School Effect	Algebra 1 School Effect	Algebra 2 School Effect	Geometry School Effect
Pearson Correlation	Career Ladder Percent	1.000	.438**	.412**	.206	.377**	.487**	.250	.161	.126	.018
	Career Ladder Eligible Percent	.438**	1.000	.540**	.445**	.295*	.682**	.212	.174	.237	-.113
	Bachelors plus	.412**	.540**	1.000	.741**	.641**	.688**	.291*	.346**	.360**	.169
	Masters degree	.206	.445**	.741**	1.000	-.040	.668**	.209	.331**	.450**	.162
	Masters plus	.377**	.295*	.641**	-.040	1.000	.260*	.179	.137	.022	.064
	Average Teacher's Salary	.487**	.682**	.688**	.668**	.260*	1.000	.358**	.308*	.394**	.188
	Pre-Algebra School Effect	.250	.212	.291*	.209	.179	.358**	1.000	.609**	.432**	.375**
	Algebra 1 School Effect	.161	.174	.346**	.331**	.137	.308*	.609**	1.000	.509**	.530**
	Algebra 2 School Effect	.126	.237	.360**	.450**	.022	.394**	.432**	.509**	1.000	.284*
	Geometry School Effect	.018	-.113	.169	.162	.064	.188	.375**	.530**	.284*	1.000
Sig. (2-tailed)	Career Ladder Percent	.	.000	.001	.099	.002	.000	.061	.201	.316	.887
	Career Ladder Eligible Percent	.000	.	.000	.000	.017	.000	.113	.165	.058	.375
	Bachelors plus	.001	.000	.	.000	.000	.000	.028	.005	.003	.181
	Masters degree	.099	.000	.000	.	.751	.000	.118	.007	.000	.201
	Masters plus	.002	.017	.000	.751	.	.036	.184	.275	.865	.618
	Average Teacher's Salary	.000	.000	.000	.000	.036	.	.006	.013	.001	.137
	Pre-Algebra School Effect	.061	.113	.028	.118	.184	.006	.	.000	.001	.004
	Algebra 1 School Effect	.201	.165	.005	.007	.275	.013	.000	.	.000	.000
	Algebra 2 School Effect	.316	.058	.003	.000	.865	.001	.001	.000	.	.023
	Geometry School Effect	.887	.375	.181	.201	.618	.137	.004	.000	.023	.
N	Career Ladder Percent	65	65	65	65	65	65	57	65	65	64
	Career Ladder Eligible Percent	65	65	65	65	65	65	57	65	65	64
	Bachelors plus	65	65	65	65	65	65	57	65	65	64
	Masters degree	65	65	65	65	65	65	57	65	65	64
	Masters plus	65	65	65	65	65	65	57	65	65	64
	Average Teacher's Salary	65	65	65	65	65	65	57	65	65	64
	Pre-Algebra School Effect	57	57	57	57	57	57	57	57	57	57
	Algebra 1 School Effect	65	65	65	65	65	65	57	65	65	64
	Algebra 2 School Effect	65	65	65	65	65	65	57	65	65	64
	Geometry School Effect	64	64	64	64	64	64	57	64	64	64

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table A-13 Correlations among selected independent variables

		Free & Reduced Lunch	Local Revenue	Per Pupil Expenditure	Regular Education spending	ACT Math	Competency Test - percent passing	Math Gr.8 National Percentile	Pre-Algebra Mean Predicted Score	Algebra 1 Mean Predicted Score	Algebra 2 Mean Predicted Score	Geometry Mean Predicted Score	Average Teacher's Salary	Bachelors plus	Masters degree	Career Ladder Eligible Percent	Career Ladder Percent
Pearson Correlation	Free & Reduced Lunch	1.000	-.526**	-.029	-.252*	-.589**	-.488**	-.407**	-.287*	-.257*	-.434**	-.532**	-.482**	-.315*	-.346**	-.317*	-.367**
	Local Revenue	-.526**	1.000	.648**	.441**	.633**	.323**	.374**	-.008	.047	.518**	.417**	.884**	.568**	.617**	.591**	.412**
	Per Pupil Expenditure	-.029	.648**	1.000	.110	.420**	.104	.162	-.078	-.016	.232	.239	.774**	.541**	.565**	.399**	.166
	Regular Education spending	-.252*	.441**	.110	1.000	.459**	.245*	.270*	.057	.146	.471**	.235	.468**	.412**	.363**	.460**	.342**
	ACT Math	-.589**	.633**	.420**	.459**	1.000	.349**	.530**	.142	.171	.544**	.529**	.678**	.609**	.534**	.435**	.405**
	Competency Test - percent passing	-.488**	.323**	.104	.245*	.349**	1.000	.568**	.307*	.438**	.288*	.350**	.365**	.252*	.163	.399**	.244
	Math Gr.8 National Percentile	-.407**	.374**	.162	.270*	.530**	.568**	1.000	.368**	.385**	.467**	.483**	.309*	.326**	.235	.225	.285*
	Pre-Algebra Mean Predicted Score	-.287*	-.008	-.078	.057	.142	.307*	.368**	1.000	.359**	.202	.499**	.017	-.056	-.060	.018	.052
	Algebra 1 Mean Predicted Score	-.257*	.047	-.016	.146	.171	.438**	.385**	.359**	1.000	.328**	.433**	.103	.044	.039	.133	.204
	Algebra 2 Mean Predicted Score	-.434**	-.434**	.232	.471**	.544**	.288*	.467**	.202	.328**	1.000	.569**	.486**	.316*	.340**	.483**	.335**
	Geometry Mean Predicted Score	-.532**	.417**	.239	.235	.529**	.350**	.483**	.499**	.433**	.569**	1.000	.430**	.321**	.329**	.366**	.276*
	Average Teacher's Salary	-.482**	.884**	.774**	.468**	.678**	.365**	.309*	.017	.103	.486**	.430**	1.000	.688**	.668**	.682**	.487**
	Bachelors plus	-.315*	.568**	.541**	.412**	.609**	.252*	.326**	-.056	.044	.316*	.321**	.688**	1.000	.741**	.540**	.412**
	Masters degree	-.346**	.617**	.565**	.363**	.534**	.163	.235	-.060	.039	.340**	.329**	.668**	.741**	1.000	.445**	.206
	Career Ladder Eligible Percent	-.317*	.591**	.399**	.460**	.435**	.399**	.225	.018	.133	.483**	.366**	.682**	.540**	.445**	1.000	.438**
	Career Ladder Percent	-.367**	.412**	.166	.342**	.405**	.244	.285*	.052	.204	.335**	.276*	.487**	.412**	.206	.438**	1.000

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

VITA

RICHARD KITZMILLER

Personal Data: Date of Birth: May 11, 1949
Place of Birth: Kingsport, Tennessee
Marital Status: Married

Education: Holston High School, Sullivan County, Tennessee
University of Tennessee, Knoxville, Tennessee;
Mathematics and Related Sciences, B.S., 1972
University of Tennessee, Knoxville, Tennessee;
Curriculum, M.S., 1973
East Tennessee State University, Johnson City, Tennessee;
School Leadership concentration, Educational Leadership and Policy Analysis, Ed.D., 2001

Professional Experience: Teacher/Coach, Bethel High School; Tipp City, Ohio, 1973-1974
Teacher/Coach, Halls Middle School; Knoxville, Tennessee, 1974-1978
Teacher/Coach, Robinson Middle School; Kingsport, Tennessee, 1978-1988
Staff Development Coordinator, Kingsport City Schools; Kingsport, Tennessee, 1988-1991
Assistant Director - Lifelong Learning, Kingsport City Schools; Kingsport, Tennessee, 1991-1992
Assessment Coordinator, Kingsport City Schools; Kingsport, Tennessee, 1992-1994
Special Assistant for Assessment and Research, Kingsport City Schools; Kingsport, Tennessee, 1994-2001