12-2005

A Study of Associations between Third Grade Tennessee Comprehensive Assessment Program Scores and Subsequent Scores in a Rural Tennessee School District.

David Cloud
East Tennessee State University

A Study of Associations between Third Grade Tennessee Comprehensive Assessment Program Scores and Subsequent Scores in a Rural Tennessee School District

---

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
David Cloud
December, 2005

Dr. Terrence Tollefson, Chair
Dr. James Lampley
Dr. Elizabeth Ralston
Dr. Jasmine Renner

Keywords: Value-added, Assessment, Tennessee, Reading, Language Arts, Math, Rural, Socioeconomic, Free or reduced lunch, Gender, Elementary

ABSTRACT

A Study of Associations between Third Grade Tennessee

Comprehensive Assessment Program Scores and Subsequent

Scores in a Rural Tennessee School District

by

David Cloud

This study was designed to examine the associations and differences that exist in the NCE scores of students on the *Terranova* portion of the Tennessee Comprehensive Exam, as well as the possible influence of variables such as initial scores, grade level, gender, and free or reduced lunch status. The population consisted of a stable group of 49 students enrolled in a rural Tennessee school district reported to have taken the annual assessment at the same school as they progressed from 3rd to 8th grade during the 1999 to 2004 school years. The study focused on the performance of students over a six-year period. The analysis focused on eight research questions. The independent variables for the study were gender, socioeconomic status, and grade level (test score reporting year). The dependent variables consisted of *TerraNova* value-added scores (NRT) and proficiency scores (CRT) translated to Normal Curve Equivalent (NCE) scores on the Reading / Language Arts and Mathematics portion of the TCAP. A combination of *t* test for independent samples, examination of effect size using eta square (η^2), and an analysis of data to determine correlation coefficient using Pearson's product moment coefficients (*r*) were used in 50 hypotheses. Statistically significant results were discovered in the following instances: students' 3rd grade Math scores and the same students'4th, 5th, 6th, 7th, and 8th grade Math scores; 5th , 6th, 7th, and 8th grade Reading/Language Arts scores and free/reduced lunch status; 3rd , 4th , 5th, and 6th grade Math scores and free/reduced lunch status; 3rd grade Reading/Language Arts scores and the same students'3rd, 4th, 5th, 6th, 7th, and 8th grade Math scores; 3rd, 5th, 6th, 7th, and

8th grade Reading / Language Arts scores and student gender; 5th and 6th grade Math scores and student gender; 2002 and 2003 mean gain scores and student gender.

DEDICATION

This work is dedicated to my wife, Sylvia, and my sons, Rowdy, Jacob, Destry, and Caleb. Without your understanding and support, this personal milestone would not have been possible.

Thank you, Sylvia, for believing in me when I had my own doubts. I am looking forward to the times ahead when we can enjoy the fruits of our labors.

To my four boys, I am so proud of all your hard work and achievements during the years I have been immersed in my studies. I know there will be many more accomplishments for each of you in the years to come. Thanks guys, for being more like your mother and less like your dad.

To my mom and my dad for instilling in me the desire to always go farther, work harder, and dream bigger.

To Walter and Mary Duncan, thank you for your prayers and your daughter Sylvia.

To the road ahead, I pray that I will strive to accomplish what God has called me to do.

Finally, I would like to dedicate this work to some real American heroes, the classroom teacher. To all those who believe that teaching is a higher calling, may God bless you.

ACKNOWLEDGEMENTS

Recognition is due to many people who have aided in the completion of this dissertation. I would like to acknowledge the tireless assistance of my graduate committee members:

Dr. Terrence Tollefson, Chair

Dr. James Lampley

Dr. Elizabeth Ralston

Dr. Jasmine Renner

I was extremely fortunate to have shared the learning experience encompassed by the ELPA Cohort program with an exceptionally dedicated group of people. Thank you, Matthew, Julie, Angela, Paul, and Judy for the inspiration and motivation.

Thanks Dr. West, you will long be remembered.

I wish to especially thank another cohort member for his assistance in collecting, coding, and analyzing the data used in this study. Thank you, James Atkins, for your efforts.

Thank you Mrs. McCoy for watching my students all those times I had to leave early to go to class.

CONTENTS

LIST OF TABLES

Table Page

CHAPTER 1

INTRODUCTION TO THE STUDY

The American education system has longingly searched for a means to systematically check the effectiveness of educational programs, instructional strategies, and teacher effectiveness. Since the arguable inception of the concept of public education under the direction of Horace Mann in the early 1800s (Cremin, 1957), educators have continuously been challenged by public concerns to demonstrate, or prove, that the quality of education received by students was high. Over the past two centuries, those concerns have expanded to include equality of educational opportunities for the individual student, regardless of race, sex, or perceived physical or mental abilities (Knight, 1989).

Numerous federal and state mandates have been enacted in an effort to guarantee that policy makers and administrators at the state and district levels are following through on these ideals. The social conscience of America emerged as equal rights activists pursued reforms in an attempt to ensure equal access to all Americans, regardless of race, creed, or color, and in 1975 Public Law 94-142 was enacted in an effort to afford students with disabilities a free and appropriate education (Pulliam & Van Patten, 2003).

This first wave of reform during the 1960s had been inspired by a surge of social activism and a heightened awareness of the civil rights needs within this country and advocated that the educational process should be an open system with equal access to knowledge. Within this system, an education has been expected to be delivered via flexible instruction adapted by the educator to the individual and group needs of the students (Glickman, 1990).

In 1965, Congress passed the Elementary and Secondary Education Act. The passage of this bill was a direct result of the public's demand for a higher quality education that would better prepare students for postsecondary studies, whether technical or academic (Pulliam & Van Patten, 2003).

The publication of *A Nation at Risk* (1983) cited the gaps in preparedness that students were experiencing as a direct result of the ineffectiveness of the public elementary and secondary education system at that time. The public school system began a transformation as changes in the way our schools were organized and operated were strategically restructured and educational systems were held accountable for student performance (Kearns, 1988). Students at the high school level were required to demonstrate "minimum competency" on standardized tests as evidence to the quality of education being delivered by the system and the individual school (Salaganik, 1985).

This educational reform movement was not only responsible for the educational programs adopted by most statewide systems, but it also became a major plank in the platforms of both national political parties. The public perception of the need for massive educational reforms began as a whisper at the grassroots level of local systems, gaining voice through the governor's offices, until becoming a clanging mandate from the White House (Cunningham, 1991).

The idea or purpose of the reform policies was to promote excellence in education by creating organizational structures and institutional arrangements that could foster growth toward that end. The degree to which these school reforms were embraced within the educational system depended in large part upon the degree to which the purpose of the reforms could be

accomplished (Timar & Kirp, 1989). The majority of educational reform policies, whether enacted or merely proposed, usually have had the following three assumptions in common:

> Assumption 1: Public schools in this country are doing an inadequate job of ensuring that students have mastered the content and acquired the skills that students should upon graduation.
>
> Assumption 2: The poor performance of our schools can be corrected through the sort of structural changes proposed in educational reform plans.
>
> Assumption 3: Increasing the amount of testing, changing the structure of the tests used, was a necessary component of any educational reform plan (Cunningham, 1991, p. 238).

While the above-mentioned assumptions remained relatively constant amid the waves of reform movements that continued to rock the boat of public education, the perceived goals that fueled public sentiment seemed to shift from a focus upon open education to educational excellence (Pulliam & Van Patten, 2003). While a majority of systems were still trying to insure equity in the areas of educational opportunity, special education, and vocational education, the public began to express the fear that we as a nation were falling behind the rest of the world in the educational preparedness of our young people. The general public began to express the concern that, "the educational foundations of our society were being eroded by a rising tide of mediocrity that threatened our very future as a nation and a people" (Porter, 1988, p. 2). The growing concern about American student achievement scores that failed to match those of other nations was reflected in an outcry for increased academic productivity by our students and higher academic standards within our public school systems that might alter our perceived future inability to compete on an international basis (Ornstein, 1988).

Others within the educational research community viewed the findings in a different light. In his seventh report on the condition of public education, Bracey stated, "On questions where few American students got the right answer, few kids on any part of the globe got the right

answer. Maybe no one is learning the 'new basics,' but, when one finds commonality among fourth-graders in 26 countries, I have to suspect that something other than the low quality of American schools is at work."

This new direction in educational reform brought with it a new emphasis on educational testing to evaluate curricula, to examine the relative educational progress being made at the district, state, and national levels, as well as to determine teacher effectiveness (Nickerson, 1989). Annual statewide standardized tests were introduced and administered at various grade levels such as, third, sixth, and ninth. Many states used these tests in a "gate-keeping role" and linked students' promotion to the next grade and occasional teacher rehire recommendations, to the students' test scores. Local schools began sending the test results out to parents in the form of a "school report card" and often those reports could be viewed in the local newspapers whereby parents could compare their children's school's performance to those of a neighboring school district's (Ornstein, 1988).

Federal Influence

The national mandate of school reform, proposed by President George W. Bush and passed into law by Congress, known as *No Child Left Behind* (NCLB) expanded on the groundwork laid by the *Elementary and Secondary Education Act* of 1965 (Pulliam & Van Patten, 2003). The sweeping changes called for in Bush's plan hinge heavily on making educators and administrators accountable for delivering a quality education to the children being served by the public school system. The checks and balances within the NCLB draw upon the model of the school systems within the Houston, Texas area and rely heavily upon standardized

testing and the proper analysis of the data generated from these tests to determine state, district, school, and teacher effectiveness.

A few states had already led the way with increased testing across all grade levels on a yearly basis and were tying the results of those tests to predetermined performance levels in an effort to determine adequate progress (California and Tennessee are two examples), in some instances, even testing students in kindergarten and 1st grade. The tests were scored and the results determined at the level of each state's department of education, then released to the general public. As public attention began to focus on the outcomes of the test scores, the comparisons between schools and districts based upon the results increased. School districts began to attempt policy, personnel, and curricular changes in an effort to improve student scores on these yearly tests. Because funding, public sentiment, and even job security began to be tied to the results of standardized test scores, the stakes were elevated for all involved. The public believed that a means was finally in place to conduct a legitimate check of school performance and hold all those involved with the operation of the schools accountable for the academic progress, or lack thereof, of the students involved. This new era in reform brought with it the promise to end the gridlock of the past where accountability consisted of different sectors of public education pointing accusing fingers at one another in a rhetorical and seldom enforced manner (Odden, 1990).

Statement of the Research Problem

The purpose of this study is to examine the associations between and among variables such as initial scale scores, NCE scores, socioeconomic levels, and the "value-added" scores

assigned to teachers as a result of the gain scores as reported on the Tennessee *TerraNova* portion of the Tennessee Comprehensive Assessment Program (TCAP).

Research Questions

The following research questions were formulated to guide the investigation.

1. To what extent, if any, are there relationships between students' 3rd grade TCAP scores in Reading / Language Arts and Math and the same students' scores as the same students progressed from the 3rd through the 8th grades?
2. To what extent, if any, are there differences between the 3rd, 4th, 5th, 6th, 7th, and 8th grade TCAP NCE scores of students on free or reduced lunch and students on paid lunch on the Reading / Language Arts portion of the *TerraNova*?
3. To what extent, if any, are there differences between the 3rd, 4th, 5th, 6th, 7th, and 8th grade TCAP NCE scores of students on free or reduced lunch and students on paid lunch on the Math portion of the *TerraNova*?
4. To what extent, if any, do relationships exist between a student's 3rd grade Reading / Language Arts TCAP score and the same student's 4th, 5th, 6th, 7th, and 8th grade TCAP scores on the Math portion of the TerraNova?
5. To what extent, if any, do differences between boys and girls exist in 3rd, 4th, 5th, 6th, 7th, and 8th grade TCAP NCE scores on the Reading / Language Arts portion of the *TerraNova*?
6. To what extent, if any, do differences between boys and girls exist in 3rd, 4th, 5th, 6th, 7th, and 8th grade TCAP NCE scores on the Math portion of the *TerraNova*?

7. To what extent, if any, do differences exist in TCAP NCE gain scores between students on free or reduced lunch and paid students for 2000, 2001, 2002, 2003, and 2004 on the Reading / Language Arts portion of the *TerraNova*?

8. To what extent, if any, do differences exist in TCAP NCE gain scores between students on free or reduced lunch and paid students for 2000, 2001, 2002, 2003, and 2004 on the Math portion of the *TerraNova*?

Significance of the Study

The Tennessee Value Added System (TVAAS) was developed and put into place to be used as an "accountability" tool. TVAAS refers to a statistical process designed for educational outcomes assessment and measures the amount of student learning over the course of a school year. The disaggregating of the data enables the public to compare the estimation of teacher, school, and school district statistical distribution. The individual teacher-effect data on the "value-added" scores are released only to administrators and individual teachers and are not intended to be used in teacher evaluations. In some school systems the confidentiality of the effect scores cannot be maintained due to the small teacher pool within grade levels. Many systems report a single teacher responsible for a subject across an entire grade level, making it impossible for the "value-added' scores reported for that grade not to be directly associated with a specific teacher. If other variables outside the realm of influence of the teacher in an educational setting exist that have a significant correlation to student achievement gains, these variables need to be factored into the equation rather than overlooked or omitted.

The idea that a student is also at least partially responsible for any learning that takes place within the classroom could be fostered through the examination of variables unique to the individual. The students and their parents, or guardian family, are responsible for many variables during the learning process that are difficult to measure but very evident in the effects on individual student achievement. Many of these variables are directly related to the home life and support systems used within the family structure.

Certain characteristics are common to healthy families: structure, support, encouragement, and security. The family unit and the student could be responsible and, therefore, accountable for a large part of the gains, or lack thereof, made during the learning process. This researcher hopes to discover if any of variables examined do indeed express a significant statistical relationship to annual student gains on the *TerraNova*.

Limitations, Delimitations, and Assumptions

The research in this study was delimited to one school within a four-school system in East Tennessee. The study was limited somewhat because of the small number of students used. For the purpose of this study, it was more feasible to select a single school population within the larger district rather than individual classes or cohort groups from a larger defined population. Because this study compared stable groups of students and the possible relationships they exhibit upon teacher effect scores, the efficacy and logic of accessing a student population within a single school seems appropriate

It was assumed that the *TerraNova* scores reported for the students were accurate and indicative of student achievement. Variables such as disruptive behavior, fear of failure, test

stress, teacher behaviors, and other extrinsic or intrinsic distractions were assumed to have been minimized by appropriate testing procedures.

Definitions of Terms

1. *Accountability Testing*. A program of testing designed to hold teachers and schools accountable for the amount of learning accomplished by their students. Standardized, multiple-choice, norm-referenced tests are most often used for this purpose.
2. *Basic Education Plan*. This plan was enacted as part of the Education Improvement Act of 1992 in an effort to equalize school funding in Tennessee. The General Assembly included accountability testing as a means of judging teacher effectiveness. The statistical model developed by Dr. William Sanders was selected to serve as the accountability instrument for this plan. The standardized tests use norm-referenced as well as criterion-referenced items (Baker, & Xu, 1995).
3. *Gain Score*. The numerical amount of improvement a student makes from one school year to the next on a standardized test.
4. *NRT*. Norm-Referenced Test, a test that is based on the standards at a national or normed level (Tennessee Department of Education, 2001).
5. *CRT*. Criterion-Referenced Test, a test that is based on the curriculum's criteria.
6. *Tennessee Comprehensive Assessment Program* (TCAP). The state's mandated system of tracking student achievement within the public school system. It consists of standardized norm-referenced and criterion-referenced testing in several subject areas in grades two through 12.

7. *Tennessee Value-Added Assessment System* (TVAAS). A procedure using a statistical model developed by William Sanders for estimating individual student achievement gains from one year to the next by comparing scale scores on the TerraNova achievement tests, using mixed model multivariate analysis to control for extraneous variables. TVAAS scores are reported as percentages of expected gains. While this system is used to compare groups of students, teachers, schools, and school systems, the public has access only to school-level data (Sanders, 1998).

8. *TerraNova Achievement Tests*. Norm-referenced and criterion-referenced batteries of objective tests published by CTB-McGraw Hill and administered to students on a yearly basis in grades two through eight under the guidelines of the TCAP (Tennessee Department of Education, 2001).

Organization of the Study

The study is comprised of five chapters. Chapter 1 is the introductory chapter containing the purpose of the study, the research questions that will guide the study, the significance of the study, delimitations, limitations, assumptions, and definition of terms. Chapter 2 contains the review of literature related to the study. Chapter 3 includes the research design, the population, the instrumentation, the method of data collection, and the methods of data analysis used in the study. Chapter 4 presents an analysis of the data and treatment of the results. Chapter 5 includes a summary of the findings, conclusions, and recommendations for practice and further study.

CHAPTER 2

REVIEW OF THE LITERATURE

Historical Perspective

Some scholars have stated that that the legislation enacted by the United States government in 1965, known as the *Elementary and Secondary Education Act* (ESEA), was the debut of the federal government into public education's efforts to equalize educational opportunities and levels of academic achievement. This legislation attempted to achieve equalization by extending federal resources to schools recognized as having high levels of poverty (Lansa & Potter, 1984).

> This measure provided funds for textbooks and other instructional materials and services in public and private elementary and secondary schools. The primary purpose was to ensure that children from low-income families had access to adequate materials. State and local governments, rather than federal agencies, had control of the funds. The act also included $100 million for research in the field of education to be administered by the United States Office of Education. This act was extended for four more years in 1966, at a cost of $12 billion (Pulliam & Van Patten, 2003, p. 207).

Following several years of relative calm on the public school front, modern American education received a wakeup call in 1983 with the release of the report *A Nation at Risk.* This report, compiled by the National Commission on Excellence in Education and submitted to the United States Secretary of Education, Terrell Bell, is considered to be responsible for the actions of many legislators and school boards across America. The focus of this report, as well as research reported by various other commissions, directed political attention to a deep public sentiment developing concerning the quality of education being provided by America's public education system. The warning was sounded on the academic deficiencies of American students in relation to their peers in Europe and Japan, specifically in the areas of mathematics and science (Gutek, 1988; Orfield & Wald, 2000).

The criticisms leveled at educational institutions grouped the decline of American students into two broad categories: first, the deteriorating moral and social values, and secondly, what the committee considered of greater importance, declining academic standards (Parkay & Hass, 2000). The view held by most critics blamed the continuing weakening of rigorous academic standards in mathematics and science, along with the gradual disappearance of broad educational standards and competencies indicated by the lowering of American students' scholastic achievement (Gutek, 1988). "Although achievement trends, for the most part , have been stable, academic and general knowledge have been at low levels for decades" (Steadman, 1993).

During this time, major changes in school policies and educational practices were put in place in an effort to counteract the public perception of the less than acceptable condition of public education (Walberg, 1986).

> With education among the electorate's top priorities, the phrase "higher standards" became ubiquitous in political campaigns across the country. Following the publication of *A Nation at Risk* during the Reagan administration, the standards-based school reform movement galvanized a broad coalition from right to left, and conservatives and business leaders were drawn to its pledge to improve the accountability of public school systems they saw as an entrenched bureaucracy. They also vowed to continue the goal of preparing a more globally competitive work force. The movement's underlying premise, that all children are able to learn at high levels, won over many liberals and civil rights advocates, who were concerned about teachers and schools lowering their expectations for poor and minority students (Orfield & Wald, 2000).

The complete perspective of recommendations that were presented by the National Commission on Excellence in Education can be found in their report: *The Imperative for Educational Reform* (Gutek, 1998).

The criticisms leveled at the outcomes of American educational institutions during this period resulted in a shift in policy. School systems, administrators, and teachers would now be "held accountable" for the educational gains (or lack thereof) of their students. According to

Orfield and Wald (2000), the reaction of policy makers ushered the era of high-stakes testing into the educational setting. This policy linked scores from one set of standardized tests to criteria such as high school graduation, grade promotion, and sometimes the scores were directly linked to decisions on teacher and administrator salaries.

High Stakes Testing

When the focus turns to high stakes testing, specifically through the application of standardized testing, educators and administrators line up in support for, or in defiance of, the process. Very few are undecided or noncommittal about the value of the results obtained from the administration of these tests in regard to the perceived downside. The term "teaching to the test" was in wide use by the latter part of the 20th century as researchers pointed to their results that indicated educators were spending more classroom time instructing students in a narrow scope of objectives found on the standardized tests and less time teaching students critical thinking and problem-solving skills. The argument also arose that some school systems buckled under the pressure to increase test scores by artificially inflating their test scores by excluding an ever-increasing percentage of students from the yearly tests. Administrators reportedly would assign lower performing students to remedial, bilingual, and special education classes in order to avoid including the scores of these students in the general population (Meier 1995).

A special report presented by the Education Commission of the States, while examining the pros and cons of annual testing, states: "The primary argument for testing each child each year is that it facilitates tracking year-to-year growth" (Education Commission of the States, 2001, p.5). Other arguable benefits of testing include: discerning state curriculum alignment with

state standards, identifying at-risk students to allow interventions, providing disaggregated date as "a tool to hold schools accountable for the progress of every student, every year" (p.5).

While some of the major criticisms of annual testing cite the costs involved as well as the time needed to administer the tests, critics also question the ability of state departments of education to manage the assessments in such a way as to provide results in a timely manner. Other major concerns argue that the tests:

> Are not yet fully aligned with states' content standards, and thus may test students on material they haven't been taught;
>
> Interfere with good teaching and learning by narrowing the curriculum and emphasizing rote memory;
>
> Do not provide accurate measures of student performance and potential;
>
> Are unfair to poor and minority students;
>
> May increase dropout rate;
>
> Could be unfair to students who lack test-taking skills. (Education Commission of the States, 2001, p.5)

One requirement of the *No Child Left Behind Act of 2001* is the annual administration of the National Assessment of Educational Progress (NAEP) test to all fourth- and eight-grade students in every state. The failure to make "adequate yearly progress" (AYP) on these assessments means that schools and districts will face corrective actions that can even include restructuring. This sweeping legislation requires that schools reach new levels of accountability by 2013-2014 (U.S. Department of Education, 2002, p. 17).

Many educators and administrators agree with the motive behind NCLB. Most would argue, however, with the implications that imply it was ever okay to leave a child behind. Houston (2005) stated, "No one wants to see a single child left behind. In fact, it is a tribute to the nation's teachers and administrators that, despite the inanity of the law's details, they are

working hard to implement it and make it work." He continues with the question, "Why would people enter a profession as pressurized and thankless as teaching if they wanted to prevent some children from reaching their dreams?" (p. 469)

TVAAS

The philosophical foundations upon which the Tennessee Value Added Assessment System (TVAAS) was built were formed as the officials of the Tennessee Department of Education reacted to the information revealed in 1983 by the release of *A Nation at Risk*. Statistical data were rapidly being presented across the state, providing Tennessee's educators with evidence that weak academic achievement existed at all levels within the educational system. Research findings were examined with the resulting claim that Tennessee had many mediocre schools. This information was coupled with the exposure of a large reading deficiency among students that filtered up from America's middle schools through the high schools (Cohen, 1995). According to Manno (1998), over one-third of America's high school seniors were unable to read proficiently.

While the purported downward spiral of academic achievement scores did not place either the teacher or student at fault, the trend did generate concern about the need for upgrading the entire educational process (Manno, 1998).

Basing their promotion for school reform on the 1983 report, advocates awakened public concern with legitimate questions concerning the academic preparation America's students were receiving within the public education system. Darling-Hammond and Archer (1991) reported that accountability was the central issue surrounding the surge for reform. "Accountability has

always been a basic concept in public education, although ideas about how to accomplish it have changed" (p. 1).

The resurgence of the public's demand for accountability in education prompted the actions of two Presidential administrations to include various funding incentives as well as federal financial support to states for developing higher academic standards (Holbein, 1998).

According to Sanders and Horn (1995), a response from the educational community (school systems, individual schools, and teachers) and in some instances, their legislative representatives, was often the reaction to publicly documented assessment results of their educational entities. As reported by Sanders (1998), Tennessee was at the national forefront in answering the public's call for schools to be held accountable and began using standardized tests to generate data for analysis to determine if, and to what degree, local school systems were successful.

In 1992, following a suit filed by small school districts, the State of Tennessee passed the Educational Improvement Act. The impetus behind this act was an attempt to equalize funding, and with it perceived educational opportunity, across the state. In order to generate the funding needed for this large reform attempt, an increase in sales tax was enacted. While the public was in basic support of the Educational Improvement Act, the demand for accountability increased significantly. The push for accountability fostered a high-stakes testing environment surrounding the assessment and measurement of student achievement and school effectiveness.

> There are a number of challenges in attempting to measure school performance and effectiveness. The school environment is an exceedingly complex milieu in which many forces are simultaneously at work impacting the school participants. There are several kinds of "effects" on student outcomes…..There are also contextual and environmental influences on students; for example it is clear that students with fewer resources and less opportunity have lower absolute levels of achievement (Stevens, Estrada, & Parkes, 2000, p. 4).

According to a report by Sanders and Horn (1998) the Tennessee Value Added System, or TVAAS, was designed to measure student achievement, dropout rates, attendance, and promotion in an attempt to determine the effectiveness of schools in increasing student achievement.

W. L. Sanders and R. A. McLean developed TVAAS while they were both statisticians at the University of Tennessee, Knoxville. This system uses a mixed-model method of statistical analysis. The data are fed into the system and generated on a yearly basis from the results reported by public schools statewide, based upon each student's performance scores on the Tennessee Comprehensive Assessment Program (TCAP). The standardized tests administered during TCAP are generated and validated by CBT McGraw-Hill (1996). TVAAS is then expected to provide unbiased reports based upon individual student gains reflected in the TCAP scores. The statistical estimation of gains would then "indicate the appropriateness of curriculum, school climate, and teacher effects on student learning" (Sanders, 1993).

> First, the mixed-model methodology used in TVAAS makes it possible to use all the data available on each child. Second, by using the longitudinal data, TVAAS is able to produce more reliable estimates of the school, system, and teacher effects on the academic gains of students than other assessment systems. Third, TVAAS contains methodology that ensures that no teacher will be misclassified as extremely good or extremely bad due to chance. Fourth, other assessment systems based on standardized testing depended on simple raw scores. TVAAS has dealt with the same evaluation problems by focusing on the measurement of academic progress. Fifth, experts in the field of educational statistics and highly respected theoretical statisticians, who have studied TVAAS, have found the process sound and appropriate for the assessment of educational effects (Bratton, Horn, & Wright, 1996, p.20)

It is worth noting that the overwhelming source of research reported to corroborate Sander's claims regarding the reliability of the methodology used within TVAAS stem directly from either Sanders himself or someone with a connection to Sanders' organization.

Other Educational Researchers have argued that additional evaluation of the TVAAS model must be undertaken by independent research teams in order to analyze the process and results in an effort to replicate the reported results. According to a report issued by Tennessee State Comptroller's office:

> The issues associated with the value-added assessment model are not merely statistical ones, but also include issues of educational measurement and public policy. Both statisticians and educational measurement experts need the opportunity to test the model. Without further evaluation, the state –and its educational practitioners-cannot determine the validity of the value-added model (Baker, et al., 1995).

The design of current accountability models, such as TVAAS, reflects the need to measure the change in student scores rather than the current status of the student. The concept behind these designs is an attempt to sift through the numerous variables affecting a student's performance and filter out the data indicating student growth.

> Use of prior achievement in longitudinal models also provides a crucially important degree of control over a wealth of confounding factors that complicate the evaluation of school effectiveness. Growth is less susceptible to background, intake, and other confounding factors. As a result, schools with lower ability students are likely to fare better when growth models are used if they are effective schools. Furthermore, use of rate of growth-type outcome measures places a focus and emphasis on the most important and relevant aspects of the educational process (Stevens, et al, 2000, p. 15).

Multiple sources of data are necessary to reach informed conclusions resulting from the analysis of that data. "Teachers and schools are held accountable for making sure that their students improve in scores from one test to the next, not for having their students meet some fixed standard minimum score" (Sanders & Horn, 1998, p. 250). Effective schools are generally distinguished from ineffective schools based on whether the students learn what is reportedly being taught. Sanders has agreed that his system is complex, and that even though he did not anticipate the majority of teachers having the ability to completely understand the complicated mathematical analysis behind the model, he claimed that the variations that could occur would

decrease over time (Hill, 2000; Sanders, 1998; Sanders & Horn, 1994). "The secret to obtaining consistent gains is to teach children from where they are when they enter the classroom" (Bratton, Horn, & Wright, 1998, p. 13).

Proponents of the Tennessee Value-Added Assessment System have pointed to the claim that TVAAS could statistically limit the influence of socioeconomic effects (SES) by taking into account an individual student's prior achievement record. Linn (2001) stated that "using prior achievement of students as a predictive factor in an accountability system has many advantages over systems that rely only on SES factors to adjust scores or to produce comparison bands of schools" (p. 17).

Those who support TVAAS cite the critical importance of identifying ineffective teachers for the sake of the students involved. In an article by Hill (2000), Sanders stated, "If any child catches two very weak teachers in a row, unless there is major intervention, that kid never recovers from it. And that's something that as a society we can't ignore" (p. 43). Sanders has summed up his stance by proclaiming that the influence of teachers "far overshadows classroom variables such as previous achievement level of students, differential class size, heterogeneity of students or the ethnic and socio-economic makeup of the classroom" (Rivers & Sanders, 2000, p. 4).

A major criticism of the Sander's model of "added-assessment" focuses on the degree to which those being assessed understand the process. Administrators and classroom teachers across the state of Tennessee appear to have formed into two camps in relationship to whether or not it is important for the individual to comprehend or simply trust the procedure. Many teachers echo the sentiments of Ballou (2002), "Virtually no one who is evaluated by these methods…will understand them. Thus, value-added systems that adjust for the unreliability of

raw test scores will fail one of the criteria that educators have deemed important for accountability: that they be *transparent*" (p. 12). Others argue that because the Sander's model is quite complex, a certain degree of trust should be invoked. "Statistical controls must be used if the assessments of teachers, schools, or programs are to be accurate…I do not require a transparent understanding of …the operating characteristics of my car; I trust the experts on the techniques. So must it be in educational evaluation" (Summers, 2002, p. 19).

Controversy

Some researchers have investigated the idea that norm-referenced tests or norm-referenced test items have been used to create standards-based interpretations. Popham (1998) found that norm-referenced test (NRT) items were not designed for use in standards-based assessment. The development process used in selecting items for a NRT naturally precludes these items for use in state systems referred to as "standards-based", "criterion-referenced", "aligned", or "linked" to state content standards due to issues in the validity and reliability of how well a student has mastered state content standards.

> There can hardly be anything as frustrating for a teacher as knowing that a student is growing and improving, yet is unable to show that growth on standardized tests and other traditional assessment measures. You scratch your head, baffled, and then you worry because you know this precious child has grown in ways these assessment measures cannot show (Nelson, 2000, p. 1)

Many educators have agreed that a structured test involving some type of comprehensive exam is needed to enable those involved to form a summative evaluation of student progress, whether it is in the form of a semester exam or simply at the end of a grading period (Hill, 2000). Some would argue that another component exists that is just as important as the bubble sheet

results rendered through standardized testing and that is affective measurement through attitudes and behaviors (Nitko, 1983).

Montouri (1994) reasoned that the suggestion that everyone learned the same and could therefore be tested in the same manner was as outdated as the bygone practice of using IQ scores in an educational placement system. The original IQ and standardized achievement tests were not designed to indicate specific problems that might randomly occur in the classroom, rather, the tests were set up to look at a students' general abilities on broad goals. Much misuse was discovered in regard to IQ tests as a means to determine if a child needed special education classes (Montouri).

In a dissertation centered on high school exit exams, J. Webb (2005) noted that students of different abilities achieved at much different levels. She stated that a "concern for the failure rate of Tennessee's students on the Gateway exam centers on socioeconomically disadvantaged students, African-American students, English Language Learners, and Special Education students" (p. 37).

One of the main criticisms of the current high-stakes testing permeating today's educational environment centers around the types of achievement tests being administered. Critics claim that norm-referenced tests (NRTs), which have been a testing mainstay for years, are far from perfect. Norm-referenced achievement tests do not reflect how well a student scored compared to everyone taking the test in a given year. NRTs compare an individual's score against the scores of a selected group of people who have already taken the same test. This group was previously selected as a sample of the target population (for example, fifth-grade students) and is supposed to fairly represent the entire target population. The scores of this

group, referred to as the "norming group", are then used to "rank-order" future test takers. These tests can also be biased in a way that favors one kind of student over another (FairTest).

> Non-school knowledge that is more commonly learned by middle or upper class children is often included in tests. To help make the bell curve, test makers usually eliminate questions that students with low overall scores might get right but those with high overall scores get wrong. The damage caused by using NRT's is far greater than any possible benefits the tests provide. The main purpose of NRT's is to rank and sort students, not to determine whether students have learned the material they have been taught. In the end, they provide a distorted view of learning that then causes damage to teaching and learning (FairTest, npn).

According to P. Webb (2005) in a dissertation presented at East Tennessee State University, the main assumption of concern rests in the belief that the effects of outside variables can be eliminated through a statistical process. He stated, "At the heart of value-added assessment is the belief that comparing students' progress from one year to the next produces gain scores that are unaffected by variables such as socioeconomic status or race. Each student acts as his or her own control. Individual gains can then be aggregated at the state, district, school, grade, and teacher level for comparison to a norm"(p. 62). Webb followed this statement up with the following citation from Kupermintz (2003):

> TVAAS claims that by simply using the student's past achievement record as a starting point from which to measure progress, and then by keeping track of who teaches the student what, all of the possible influences on this student's learning can be filtered out or taken into account. No other educational assessment system has ever made such a bold claim. (p. 4)

Some opponents of the current educational accountability model believe there is a viable alternative to Sanders' TVAAS model. The use of real-life tasks to replace contrived test items is referred to as "authentic assessment".

> Authentic assessment is a way to evaluate all children, including children with special needs, in the learner-centered classroom. Authentic assessment considers all performance tasks, portfolios, student self-assessment surveys and probes, peer assessments, journals, logs, and many diverse types of projects (Robertson & Valentine, 1999, p.8).

One of the many ironies that thrive within the current educational climate involves how the teaching strategies being taught in college education courses and testing practices useed in school systems sometimes do not converge. Experts in the field of curriculum and instruction continue to recommend that new teachers use a variety of strategies to encourage the recall of subjects over extended periods of time which they were taught in college teacher education programs (Coleman, 2000, Cooper et al., 1990). "Learning to take reading and writing tests is not the same as learning to read and write, especially when test prep materials do not meet basic standards" (Orfield & Wald, 2000, p. 38). "Whenever it is critically important to master certain content, the knowledge that it will be tested produces a desirable concentration of effort. On the other hand, learning the answers to a set of questions is by no means the same as acquiring understanding of whatever topic the question represents" (Cronbach, 1963, p. 681).

Barriers to Student Achievement

An ample body of empirical research exists regarding evidence of the important role played by both teacher and pupil in accounting for gains in student academic achievement. Results indicate that a student's family characteristics and ethnicity can be overcome by the direct influence of the schools they attend and what their teachers know and do. Some of this research affirms the fact that teaching quality tends to be the most important factor influencing achievement gains by students (Darling-Hammond, 2000; Kaplan & Owings, 2001; Whitehurst, 2002). "Staffing all classrooms with highly qualified teachers…is a critical national concern" (Dosset & Munoz, 2003, p.12). Effective schools are generally distinguished from ineffective schools based on whether the students learn what is reportedly being taught. According to Sanders and Horn (1995), the main reason previous achievement data to determine whether a

student is learning or not has been used sparingly stems from the difficulty in separating school effects on learning from demographics.

A study by Good et al. (1983) linked several key factors to the amount of learning accomplished by students. According to their findings, increased learning occurred when teachers maximized allocated classroom time to academic endeavors, exposure to content, frequent application of what was learned, and student perception of the teacher's belief in a student's ability to succeed.

Lowering the number of students in a class and increasing access to technology has had a positive effect on struggling students. One-on-one tutoring and computer assisted instruction have been found to have a positive impact on disadvantaged students' academic achievement. The conclusions reached by Stringfield and Herman (1997) indicated that equality could be attained and achievement at high levels would be possible if students were provided with quality education programs. The most successful programs, such as Reading Recovery and Success For All, used very focused tutoring models. The researchers advised that these programs should begin at preschool and continue to be consistently implemented within educational structures that were considered to be appropriate.

A study of the Minneapolis Public Schools conducted by Du and Heistad (1999) noted several factors affecting student achievement. Among these variables, the students' prior reading scores and socioeconomic status (as defined by eligibility for free or reduced lunch) were found to be related to student achievement. The study found that more than two thirds of the variance in scores could be attributed to student background factors.

Some educational experts see the current trend toward increased use of high-stakes testing to evaluate the effectiveness of school systems as creating a landscape where those who

already possess the best chance to survive or excel because of life circumstance also have the greatest opportunites to learn (Fuchs & Fuchs, 1995; Sapon-Shavin, 1993).

Several of the variables that have been associated with the socioeconomically disadvantaged, or at-risk, students have also been correlated with student academic achievement. Researchers have studied several possible variables such as family background, student prior achievement, and socioeconomic status regarding their ability to predict future academic gains of students (Munoz & Dossett, 2001; Roeder, 2000; Smith & Meier, 1995, Chubb & Moe, 1990).

According to Berliner and Biddle (1995), "Schools bring little influence to bear on a child's achievement that is independent of his background and general social context; and that this very lack of independent effect means that the inequalities imposed on children by their home, neighborhood, and peer environment are carried along to become the inequalities with which they confront adult life at the end of school" (p. 71).

In his often cited 1966 report, Coleman examined the effects that characteristics of different schools had on students with differing socioeconomic and racial backgrounds. According to his findings, a school's facilities (libraries, science laboratories, etc.) exhibited a positive relationship to the achievement of minority students coming from a low socioeconomic background.

Roeder (2000) examined the relationship of selected social and academic variables with school performance. When the researcher controlled for several school and district variables, poverty emerged as the strongest determinant of school performance. A study by Munoz and Dosset (2001) recorded socioeconomic status (participation in free/reduced lunch program) as accounting for the highest percentage of explained variance in student achievement. An average

of 58% of the variance across all four regression models used in the study was directly linked to student participation in a free/reduced lunch program.

While some researchers have argued that the effect of a student's socioeconomic background is often overemphasized, other educational analysts point to the body of research compiled over a 20-year period as evidence to their claims. "Family background is the single best predictor of achievement. And students…arrive at school with significant cognitive differences that schools can't control" (Long & Cass, 2001).

Another variable to consider is student mobility. In a study by Wright (1999), the effect of student mobility on achievement gains was found to be less of a factor on outcomes than other socioeconomic variables.

> Mobility is generally subordinate in magnitude to other factors such as ethnicity, family income, and (in one comparison) gender. The results also provide an explanation for the somewhat counterintuitive observation that lower achievement often precedes mobility rather than following it; lower achievement is associated with other more powerful predictors than temporal mobility. …. The broad conclusion that may be drawn from the results is that student mobility is subordinate in its effects on achievement to the risk factors for ethnic minority status and low family income (Wright, 1999, July/August, p. 352)

When the composition of schools and students vary greatly within a district or community, an understanding of the schooling process is important. It is important to understand the complexity of variables present: parental involvement, peer attitudes, school environment, school atmosphere, disciplinary problems, and socioeconomic factors. The composition of a school and the environment encountered by the student has been shown to have substantial effects on achievement beyond the effects associated with the individual students social class or ability (Willms, 1986).

Prior Achievement

The notion that previous academic performance can be a strong predictor of future academic gains was supported by Smith and Meier (1995). In their study of school systems at the district level, schools that did well in the past continued to perform at a similar success rate. Brown (1994) affirmed that "…a particularly important development over the last decade has arisen from the large measure of agreement that account must be taken of differences among pupils when they arrive at school, and that the unit of analysis has to be the individual pupil rather than some average measure across pupils" (p. 56).

Inherent weaknesses accompany the most commonly used educational outcomes indicators, such as median test scores, gain indicators (change in scores exhibited by a cohort of students as they progress from grade to grade), and proficiency level indicators (the proportion of students who score above a determined score deemed as a proficiency level). These frequently used performance indicators tend to be biased against schools that serve high proportions of at-risk students with high mobility (Meyers 1996, 2000). Value-added assessment models have been designed to control or minimize the effect of outside variables associated with at-risk students in an attempt to use these indicators to isolate and measure the effect a school setting has upon the educational gains of students.

Student Accountability

The concept that teachers are totally responsible for their students' academic gains has also drawn some criticism. One of the major questions surrounding TVAAS relates directly to student accountability. The statistical model conceived by Sanders and used by TVAAS assumes

that all learning takes place in the classroom environment and does not take into consideration any learning that takes place in other educational settings or at home (Hill, 2000).

"The model seems to assume that all gain (or lack thereof) is purely teacher-related, while it has not provided adequate evidence to support this contention" (Baker & Xu, 1995). Research conducted by Chubb and Moe (1990) used initial student scores to predict future gain. The researchers determined student ability or initial achievement to be the strongest predictor of achievement gains.

A number of educational experts have sounded a warning that by attributing learning growth trends merely to teacher effectiveness, a huge risk is incurred. Critics claim that by focusing solely on the perceived effectiveness of the teacher, educators are ignoring the role students play in their own educational growth. Natural aptitude and a student's individual motivation can play a crucial role in how well, or how poorly, that student performs (Kupermintz, 2002). "More variability in teacher effectiveness exists in the higher elementary grades than in the lower grades. As the grade level increases, teacher variability increases (Pipho, 1998, p. 342).

If certain factors outside a teacher's control are not taken into consideration, such as whether certain teachers are assigned higher- or lower-ability students at a consistent rate, "it becomes increasingly difficult to differentiate between learning gains that should be attributed to teachers and those that reflect the superior aptitude of their students" (Kupermintz, 2002, p. 8).

Summary

This chapter has presented a review of literature that focused on research findings and scholarly literature relevant to the topics of study. These topics include an historical perspective

of school accountability, an overview of high stakes testing, the Tennessee Value-Added Assessment System (TVAAS), the controversy surrounding value-added assessment models, some barriers to student achievement, the importance of prior achievement as a predictive variable, and student accountability.

CHAPTER 3

RESEARCH METHODOLOGY

The purpose of this study was to examine the relationships between variables such as initial scale scores, socioeconomic levels, gender, and gain scores reported on the Tennessee *TerraNova* portion of the Tennessee Comprehensive Assessment Program. This chapter describes the research design, population, instrumentation, data collection methods, and methods of analysis to be used in the study.

Research Design

The causal-comparative, quasi-experimental quantitative approach to exploring possible associations and comparisons of scores was employed in this study. The purpose of this study was to examine the initial scale scores and normal-curve-equivalent scores of students who took the Tennessee *TerraNova* portion of the Tennessee Comprehensive Assessment Program as 3rd grade students within the Grainger County School system. While this study used 3rd grade scores as a starting point, the resulting scores of these same students for grades four, five, six, seven, and eight were included for analysis. The scores were analyzed to determine whether relationships exist between these scores and variables such as the socioeconomic levels of students, the gender of students, and the possible predictive nature of initial scores as their group passes through a rural public school system. The initial scale scores and normal curve equivalent scores of a cohort of students were compared to the scale scores, NCE, and gain scores of these same students during the testing period of their third- through eighth-grade years. These same scores were then analyzed to determine if any relationships exist among the aforementioned

variables. This research design featured the study and analysis of archival data based on causes that were examined after they have exerted their effect on another variable. This method is often referred to as *ex post facto* research (Gall, M. D., Borg, W. R., & Gall, J. P, 1996). Even though this design is not a true experimental design and, therefore, does not provide for a direct test of causation, it will provide information that will support or refute causal explanations. In this case, achievement test scores, demographic information, and gain scores were collected from the Grainger County district records and comparisons were made between the scores and determined variables.

Population

The population for this study was chosen in large part to access and availability of individual student achievement data. While the availability of data was a driving force behind the selection of Grainger County Schools for inclusion in this study, the relatively high percentage of students considered to be economically disadvantaged within the system was also a major consideration. The population consisted of elementary school students within the Grainger County School District. The Grainger County School District is located in the eastern portion of Tennessee and is comprised of six individual school sites. The Local Education Authority (LEA) in Grainger County directs and maintains two K-8 elementary schools (Joppa and Bean Station), one K-2 (Rutledge Elementary), one 3-8 (Rutledge Middle), one K-12 (Washburn), and one 9-12 (Rutledge High). The total number of students attending school within the Grainger County School District during the 2003-2004 school year was 3,335, with over 98% reported as white, just over 17% listed as students with disabilities, and almost 58% of the school population considered to be economically disadvantaged. The subjects of this study were drawn from the

entire student population of Grainger County and they include all of those students who completed the *TerraNova* portion of the Tennessee Comprehensive Assessment Program (TCAP) as 3rd grade students during the 1999-2000 school year as well as subsequent years through 2003-2004. The demographics of individual schools within the Grainger County School District by school site were: Joppa K-8, 618 total students, 99% white, and 46% economically disadvantaged; Rutledge Middle 3-8, 398 total students, 97% white, 57% economically disadvantaged; Bean Station K-8, 732 total students, 99% white, 64% economically disadvantaged (2004 Tennessee Report Card). The cohort consisted of the 49 students attending a local school as 3rd grade students during the 1999-2000 school year and continuing through the 8th grade, 2003-2004 school year. The school used as the cohort was chosen, with permission from the Director of Schools, from one of the four elementary schools within the Grainger County School system and individual students remained anonymous to prevent possible recognition of subjects. No individual students or teachers were identified within the scope of this study.

Instrumentation

Academic achievement for the population being studied was compared through the use of the *TerraNova* Comprehensive Test of Basic Skills (CBT McGraw-Hill, 1996). Each spring, students in Tennessee schools in grades three through eight are mandated to take an achievement test as part of the Tennessee Comprehensive Assessment Program (TCAP). The goal of this test was to provide an accurate measure of basic academic skills. Content knowledge, as well as the application of such knowledge, was assessed using a multiple-choice question and answer format with set time limits. The questions were purported to go beyond workbook drill and practice and

evaluate students' higher order thinking skills. This format was similar to that used on the National Assessment of Educational Progress (NAEP) test (Tennessee Department of Education, 1999).

The *TerraNova*, published by CTB/McGraw-Hill (1996), provides both norm-referenced and criterion-referenced test items. The test used the most recently available national norms. Norm-referenced data allowed for the comparison of the achievement level of a group of students to the performance of a national sample of students on the same test. Summary reports provided results expressed in national percentiles. Median national percentile data were provided for reading, language, mathematics, science, and social studies. Criterion referenced test items provided an opportunity to compare student achievement against a pre-determined level of performance.

Third-grade students were permitted to "bubble" their answer choices directly into the test booklet. Students in grades four through eight were required to use a separate answer sheet (Tennessee Department of Education, 1999). The CTBS reportedly have been found to be both reliable and valid. The public school samples were stratified by region, community type, size, and Orshansky percentile, which is an indicator of a districts socioeconomic status. Standardization and norming procedures, as well as research studies addressing reliability issues are reported in the Tennessee Coordinator's Handbook (CTB McGraw-Hill, 1997).

Data Collection

Approval to initiate this study was obtained from the Institutional Review Board at East Tennessee State University prior to any data collection. Data to be used for this study were completely archival. Written permission to conduct this study was obtained from the Director of

Schools, Grainger County, Tennessee. Reports provided by the testing service were obtained from official cumulative records. Coded identities for the names and schools were used to protect the privacy of all students and teachers.

The major source of data for comparison were gain scores and the Normal Curve Equivalent scores (NCE). These scores were used for comparison and to calculate gains from one test to the next. The NCE is an equal-interval score that can be treated arithmetically (Cannon, 2000). The scores from Total Reading, Total Language, and Total Battery were used to make comparisons for statistically significant differences.

Data Analysis

Descriptive statistics were performed as the initial step in data analysis and were used to provide a profile of the population to be studied. The Statistical Program for the Social Sciences (SPSS) was used to analyze data. In order to answer questions 1 and 4, a statistical analysis was performed utilizing student NCE scores to determine a correlation coefficient. The interval scale correlation coefficient used in this case was the Pearson product-moment coefficient (r). A correlation coefficient (r) of 0 indicated that the two variables being tested exhibited no relationship with one another, or were considered to be independent of one another. Third-grade scores, fourth-grade scores, fifth-grade scores, sixth-grade scores, seventh-grade scores, and eighth-grade scores on the *TerraNova* were collected for these comparisons. Questions 2, 3, 5, 6, 7, and 8 were analyzed using a t-test for independent means to determine any differences that might have existed between the initial third grade NCE and gain scores of the students and later 4^{th} through 8^{th} grade scores of the same students. All statistical tests were conducted using a preset alpha level (a) of .05 to determine if statistically significant results occurred in Total

Reading, Total Language, and Total Battery scores of students. The effect size of the findings was indicated by computing eta square (η^2). The description of small versus large η^2 is dependent on the area of investigation. An η^2 of .01, .06, and .14 are interpreted as small, medium, and large effect sizes (Green & Salkind, 2004).

Summary

Chapter 3 presented the methodology and procedures to be used in this study. The causal-comparative research method was chosen and explained. The population and selection methods were described. *TerraNova* CTBS along with its reliability and validity were presented. The methods of data collection and data analysis were detailed. Results of the analysis of data research are presented in Chapter 4.

CHAPTER 4

ANALYSIS OF DATA

As a result of state and federal government intervention in the form of more rigorous accountability measures, schools are relying on programs to increase standardized test scores (Scanlon, 1998). This sharpened focus on improving school performance on yearly assessment programs has been accompanied by an increased scrutiny of possible effects teachers have on individual student test scores. Tennessee has led the way in determining a teachers' "value-added" effect on student test scores with the implementation of the TVAAS, long before the current accountability requirements were mandated by NCLB. Under Tennessee's current accountability system, teachers are assumed to be the single most influential variable in a child's educational setting (Sanders & Horn, 1998).

The purpose of this study was to examine the relationship between and among variables unique to individual students, such as initial student scores on the Reading / Language Arts and Math sections of the *TerraNova* portion of the TCAP, subsequent scores of the same students in following years, participation in the school's free or reduced lunch program, and student gender.

The students participating in the study were third-, fourth-, fifth-, sixth-, seventh-, and eighth-grade students who were enrolled in Grainger County Schools during the 1999-2004 school years and who took the *TerraNova* for six consecutive years between 1999 and 2004. Of the 58 students who were third-, fourth-, fifth-, sixth-, seventh-, and eighth-graders during 1999-2004, 9 were excluded because they did not take the *TerraNova* all six years of the study. The resulting population was 49. This chapter contains an analysis of data collected over a six-year period, involving14 teachers, 49 students, and a single elementary school.

Eight research questions were constructed to guide the investigation. The data were used to test 50 null hypotheses. Computer data analysis was performed using the Statistical Package

for the Social Sciences (SPSS). Demographic characteristics of the population included gender and free or reduced lunch status. Table 1 shows the demographic profile of the sample.

Table 1

Demographic Profile of the Sample

Characteristic	*N*	*%*
Gender		
Female	26	53
Male	23	47
Total	49	100
Socioeconomic Status		
Paid	19	39
Free/Reduced	30	61
Total	49	100
Gender/SES Status		
Female		
Paid	11	22
Free/Reduced	15	31
Total	26	53
Male		
Paid	8	16
Free/Reduced	15	31
Total	23	47

As shown in Table 1, the sample consists of 49 students reported taking the TCAP exam at the same school during the 1999-2004 school years. Of these students, a total of 26 (53%) are female and 23 (47%) are male. Thirty (61%) of the students were identified as receiving free or reduced lunch. Of the students identified as receiving free or reduced lunch, 31 percent (15 male, 15 female) of both the male and female populations were equally represented.

The data consisted of both CRT and NRT components. The CRT data contain raw scores, scale scores, and proficiency levels. However, the NRT component is more complicated. In 2004, the Tennessee Department of Education changed the way value-added was calculated. Prior to 2004, value-added was calculated by taking the net gain/loss in scale scores from the previous to the current year divided by the USA norm. This value was multiplied by 100 to convert the decimal to a percent (Atkins, 2005).

In 2004, the process of determining value-added scores in Tennessee changed. Tennessee statute authorizes the State Commissioner of Education to set state growth standards for grades four through eight. The state growth rates in 1998 were used as the growth standard in 2004 for grades four thorough eight. Elementary and middle schools in Tennessee show progress on CRT tests in "State NCE Scores." The state NCE scores are based on 1998 growth standards. All previous *TerraNova* NRT scores were mapped to CRT scores using concordance tables. The Tennessee Department of Education used the equi-percentile method for single group design to map the scores. The state growth standards replaced the USA norms. Results greater than zero indicate more progress than the growth standard or the state average in 1998. Therefore, value-added scores greater than zero represent gains greater than Tennessee's average scores in 1998. These changes reflected a shift from Tennessee's previous accountability system to the federal system as a result of NCLB. The change necessitated a transition between the two different types of tests (Park, 2005). As a result, the tests were equated using the NRT - CRT - NCE concordance tables for the sixth, seventh, and eighth grades.

To provide history about the transition from NRT to CRT scores in Tennessee, Tennessee state law required NRT scores be used for value-added calculations. However, NCLB required the use of CRT scores. This resulted in double testing in 2004 to satisfy both Tennessee and NCLB requirements. The conversion of NRT to CRT equated scores avoids double testing (Atkins, 2005). Concordance was used to map the scores from one scale to another. Concordance refers to the process in which methods are used to link scores on tests that are built to different specifications. Concordance can be used when tests are measuring similar construct

and scores are highly correlated (Park, 2005). Equi-percentile scaling and linear scaling are two basic statistical methods used to produce concordance tables. The equi-percentile method sets equal the scores on each test having the same percentile ranks. For example, the score at the 65th percentile on the *TerraNova* NRT score distribution would correspond to the score at the 65th percentile of the *TerraNova* CRT score distribution. However, equity cannot be achieved even for scores measuring the same thing unless the two scores are parallel. If equity cannot be achieved for measuring the same thing, it likewise cannot be achieved for scores measuring different things. Scores are referred to as closely equable if they are parallel (Hanson, B., Harris, D., Pommerich, M., Sconing, J., & Yi, Q.,2001).

Concordance tables may be based on equating or scaling. To support scaling, the correlation must be high. If the correlation is too low, then concordance becomes merely a predictor. Researchers refer to a "reduction of uncertainty" as $1 - \sqrt{(1-r^2)}$, where r is the correlation (Dorans, 2000). Tennessee uses a 50% reduction of uncertainty. Therefore, the correlation must be a minimum of .866 between the scores being mapped to reduce the uncertainty by at least 50%. "If a predictor cannot reduce uncertainty by at least 50%, it is unlikely that it can serve as a valid surrogate for the score you want to predict" (Dorans, p. 3). Four prerequisites for equating include: (a) the two tests must measure the same construct, (b) the equating must achieve equity, (c) the equating transformation should be symmetric, and (d) the transformation should be invariant across subpopulations (Dorans).

Research Question 1

To what extent, if any, are there relationships between students' 3rd grade TCAP scores in Reading / Language Arts and Math and the same students' scores as the students progressed from the 3rd through the 8th grades?

From Research Question 1, the following hypotheses were developed and tested:

$Ho1_1$: There is no relationship between students' 3rd grade TCAP scores in Reading / Language Arts and the same students' scores on the 4th grade Reading / Language Arts TCAP.

$Ho1_2$: There is no relationship between students' 3rd grade TCAP scores in Math and the same students' scores on the 4th grade Math TCAP.

$Ho1_3$: There is no relationship between students' 3rd grade TCAP scores in Reading / Language Arts and the same students' scores on the 5th grade Reading / Language Arts TCAP.

$Ho1_4$: There is no relationship between students' 3rd grade TCAP scores in Math and the same students' scores on the 5th grade Math TCAP.

$Ho1_5$: There is no relationship between students' 3rd grade TCAP scores in Reading / Language Arts and the same students' scores on the 6th grade Reading / Language Arts TCAP.

$Ho1_6$: There is no relationship between students' 3rd grade TCAP scores in Math and the same students' scores on the 6th grade Math TCAP.

$Ho1_7$: There is no relationship between students' 3rd grade TCAP scores in Reading / Language Arts and the same students' scores on the 7th grade Reading / Language Arts TCAP.

$Ho1_8$: There is no relationship between students' 3rd grade TCAP scores in Math and the same students' scores on the 7th grade Math TCAP.

$Ho1_9$: There is no relationship between students' 3rd grade TCAP scores in Reading / Language Arts and the same students' scores on the 8th grade Reading / Language Arts TCAP.

$Ho1_{10}$: There is no relationship between students' 3rd grade TCAP scores in Math and the same students' scores on the 8th grade Math TCAP.

In order to test $Ho1_1$, $Ho1_3$ $Ho1_5$, $Ho1_7$, and $Ho1_9$, a statistical analysis was performed to determine a correlation coefficient. The interval scale correlation coefficient used in this case

was the Pearson product-moment coefficient (r). The predetermined alpha (α) is .05. Table 2 shows the results of the data analysis.

Table 2

Correlation of Reading / Language Arts Scores

		3rd grade Reading /Language	
	N	*r*	*p*
4th grade Reading/Language	49	.87	<.01
5th grade Reading/Language	49	.83	<.01
6th grade Reading/Language	49	.78	<.01
7th grade Reading/Language	49	.75	<.01
8th grade Reading/Language	49	.64	<.01

$\alpha = .05$

As shown in Table 2, there was a significant relationship between students'3rd grade Reading/Language Arts NCE scores and the same students' 4th, 5th, 6th, 7th, and 8th grade Reading/Language Arts NCE scores. Because the *p*-value (<.01) is less than the predetermined alpha (.05), I rejected the null hypotheses ($Ho1_1$, $Ho1_3$ $Ho1_5$, $Ho1_7$, and $Ho1_9$). The Pearson Correlation Coefficient between 3rd grade Reading/Language Arts scores and the 4th (r= .87) and 5th (r= .83) grade NCE scores indicated a strong positive relationship existed.

Also shown in Table 2, the Pearson Correlation Coefficient between 3rd grade Reading/Language and the NCE scores for 6th (r= .78) and 7th (r= .75) grades exhibited a

medium-strength, positive relationship. The comparison of 3rd and 8th grade NCE scores indicated a low-medium strength, positive relationship (r= .64).

In order to test $Ho1_{2,}$ $Ho1_{4}$ $Ho1_{6,}$ $Ho1_{8,}$ and $Ho1_{10,}$ a statistical analysis was performed to determine a correlation coefficient. The interval scale correlation coefficient used in this case was the Pearson product-moment coefficient (r). The predetermined alpha (α) was .05. Table 3 shows the results of the data analysis.

Table 3

Correlation of 3rd through 8th Grade Math Scores

		3rd grade Math	
	N	*r*	*p*
4th grade Math	49	.85	<.01
5th grade Math	49	.80	<.01
6th grade Math	49	.79	<.01
7th grade Math	49	.79	<.01
8th grade Math	49	.83	<.01

$\alpha = .05$

As shown in Table 3, a significant relationship existed between students' 3rd grade State Math NCE scores and the same students'4th, 5th, 6th, 7th, and 8th grade Math NCE scores. Because the *p*-value in the tested hypotheses, $Ho1_{2,}$ $Ho1_{4}$ $Ho1_{6,}$ $Ho1_{8,}$ and $Ho1_{10}$ (<.01), is less than the predetermined alpha (.05), I rejected the null hypotheses ($Ho1_{2,}$ $Ho1_{4}$ $Ho1_{6,}$ $Ho1_{8,}$ and $Ho1_{10}$). Because the Pearson Correlation Coefficient (*r*) of the tested hypotheses resulted in an *r* equal to or greater than .79 (*r=> .79),* a strong positive relationship was indicated.

Research Question 2

To what extent, if any, are there differences between the 3rd, 4th, 5th, 6th, 7th, and 8th grade TCAP NCE scores of students on free or reduced lunch and students on paid lunch on the Reading / Language Arts portion of the *TerraNova*.

From Research Question 2, the following hypotheses were developed and tested:

$Ho2_1$: There is no difference in the 3rd grade TCAP NCE scores of students on free or reduced lunch and students on paid lunch on the Reading / Language Arts portion of the *TerraNova.*

$Ho2_2$: There is no difference in the 4th grade TCAP NCE scores of students on free or reduced lunch and students on paid lunch on the Reading / Language Arts portion of the *TerraNova.*

$Ho2_3$: There is no difference in the 5th grade TCAP NCE scores of students on free or reduced lunch and students on paid lunch on the Reading / Language Arts portion of the *TerraNova.*

$Ho2_4$: There is no difference in the 6th grade TCAP NCE scores of students on free or reduced lunch and students on paid lunch on the Reading / Language Arts portion of the *TerraNova.*

$Ho2_5$: There is no difference in the 7th grade TCAP NCE scores of students on free or reduced lunch and students on paid lunch on the Reading / Language Arts portion of the *TerraNova.*

$Ho2_6$: There is no difference in the 8th grade TCAP NCE scores of students on free or reduced lunch and students on paid lunch on the Reading / Language Arts portion of the *TerraNova.*

An independent-samples t test was conducted to evaluate whether the mean TCAP NCE score on the Reading/Language Arts portion of the *TerraNova* would differ depending upon the free/reduced or paid lunch status of a student. The NCE score was the test variable and the grouping variable was whether a student was classified as receiving free or reduced lunch or whether the student paid full lunch price. Table 4 shows the results of the data analysis.

Table 4

Comparison of Reading / Language Arts Scores by Socioeconomic Status

Subtest	*N*	*M*	*t*	*df*	*p*
3rd grade Reading /Language					
Paid	19	57.58	1.49	47	.14
Free/Reduced	30	48.27			
4th grade Reading /Language					
Paid	19	59.53	2.71	47	.01
Free/Reduced	30	46.07			
5th grade Reading /Language					
Paid	19	55.26	2.18	47	.04
Free/Reduced	30	43.67			
6th grade Reading /Language					
Paid	19	68.53	3.84	47	<.01
Free/Reduced	30	46.47			
7th grade Reading /Language					
Paid	19	60.89	3.99	47	<.01
Free/Reduced	30	45.17			
8th grade Reading /Language					
Paid	19	59.68	2.64	47	.01
Free/Reduced	30	47.83			

*critical value of $t = 1.68$ $\alpha = .05$

As shown in Table 4, the findings were varied. The results were not significant, $t(47) = 1.49$, $p = .14$, for $Ho2_1$. I failed to reject the null ($Ho2_1$) because t (1.49) is less than the critical value of t (1.68) and concluded that there is no difference in the 3rd grade Reading / Language Arts NCE scores of students on free or reduced lunch and students on paid lunch. The test was significant for 4th, 5th, 6th, 7th, and 8th grade scores.

The results for 4th grade, $t(47) = 2.71$, $p = .01$, indicated that a statistical significance existed and I rejected the null hypothesis ($Ho2_2$). The η^2 (.07) indicated a medium effect size.

The results for 5th grade, $t(47) = 2.18$, $p = .04$, indicated that a statistical significance existed and I rejected the null hypothesis ($Ho2_3$). The η^2 (.05) indicated a low/medium effect size.

The results for 6th grade, $t(47) = 3.84$, $p = <.01$, indicated that a statistical significance existed and I rejected the null hypothesis ($Ho2_4$). The η^2 (.14) indicated a large effect size.

The results for 7th grade, $t(47) = 3.99$, $p = <.01$, indicated that a statistical significance existed and I rejected the null hypothesis ($Ho2_5$). The η^2 (.15) indicated a large effect size.

The results for 8th grade, $t(47) = 2.64$, $p = .01$, indicated that a statistical significance existed and I rejected the null hypothesis ($Ho2_6$). The η^2 (.07) indicated a medium effect size.

……*Research Question 3*

To what extent, if any, are there differences between the 3rd, 4th, 5th, 6th, 7th, and 8th grade TCAP NCE scores of students on free or reduced lunch and students on paid lunch on the Math portion of the *TerraNova*.

From Research Question 3, the following hypotheses were developed and tested:

$Ho3_1$: There is no difference in the 3rd grade TCAP NCE scores of students on free or reduced lunch and students on paid lunch on the Math portion of the *TerraNova.*

$Ho3_2$: There is no difference in the 4th grade TCAP NCE scores of students on free or reduced lunch and students on paid lunch on the Math portion of the *TerraNova.*

$Ho3_3$: There is no difference in the 5th grade TCAP NCE scores of students on free or reduced lunch and students on paid lunch on the Math portion of the *TerraNova.*

$Ho3_4$: There is no difference in the 6th grade TCAP NCE scores of students on free or reduced lunch and students on paid lunch on the Math Arts portion of the *TerraNova.*

$Ho3_5$: There is no difference in the 7th grade TCAP NCE scores of students on free or reduced lunch and students on paid lunch on the Math portion of the *TerraNova.*

$Ho3_6$: There is no difference in the 8th grade TCAP NCE scores of students on free or reduced lunch and students on paid lunch on the Math portion of the *TerraNova.*

An independent-samples t test was conducted to evaluate whether the mean TCAP NCE score on the Math portion of the *TerraNova* would differ depending upon the free/reduced or paid lunch status of a student. The NCE score was the test variable and the grouping variable was whether a student was classified as receiving free or reduced lunch or whether the student paid full lunch price. Table 5 shows the results of the data analysis.

Table 5

Comparison of Math Scores by Socioeconomic Status

Subtest	*N*	*M*	*t*	*df*	*p*
3rd grade Math					
Paid	19	61.89	2.24	47	.03
Free/Reduced	30	48.40			
4th grade Math					
Paid	19	58.53	2.85	47	.01
Free/Reduced	30	43.63			
5th grade Math					
Paid	19	56.11	2.24	47	.03
Free/Reduced	30	43.27			
6th grade Math					
Paid	19	57.58	2.53	47	.02
Free/Reduced	30	44.23			
7th grade Math					
Paid	19	60.37	1.26	47	.21
Free/Reduced	30	52.83			
8th grade Math					
Paid	19	62.42	1.67	47	.10
Free/Reduced	30	54.10			

*critical value of t =1.68 α= .05

As shown in Table 5, the results of the test were significant for 3rd, 4th, 5th, and 6th grade scores. Third grade students paying for meals had a higher mean NCE than students on free or reduced lunch. The results, $t(47) = 2.24$, p= .03, indicated that a statistical significance existed and I rejected the null hypothesis ($Ho3_1$). The η^2 (.05) indicated a low/medium effect size.

Fourth grade students paying for meals had a higher mean NCE than students on free or reduced lunch. The results, $t(47) = 2.85$, p= .01, indicated that a statistical significance existed and I rejected the null hypothesis ($Ho3_2$). The η^2 (.08) indicated a medium effect size.

Fifth grade students paying for meals had a higher mean NCE than students on free or reduced lunch. The results, $t(47) = 2.24$, $p = .03$, indicated that a statistical significance existed and I rejected the null hypothesis ($Ho3_3$). The η^2 (.05) indicated a low/medium effect size.

Sixth grade students paying for meals had a higher mean NCE than students on free or reduced lunch. The results, $t(47) = 2.53$, $p = .02$, indicated that a statistical significance existed and I rejected the null hypothesis ($Ho3_4$). The η^2 (.07) indicated a medium effect size.

The test results for seventh and eighth grade students failed to reveal a statistically significant difference. Seventh grade students paying for meals had a higher mean NCE than students on free or reduced lunch, however, the results, $t(47) = 1.26$, $p = .21$, indicated that a statistical significance did not exist and I failed to reject the null hypothesis ($Ho3_5$).

Eighth grade students paying for meals also had a higher mean NCE than students on free or reduced lunch, but the results of the analysis, $t(47) = 1.67$, $p = .10$, indicated that a statistically significant difference did not exist. I failed to reject the null hypothesis ($Ho3_6$).

Research Question 4

To what extent, if any, do relationships exist between a student's 3rd grade Reading / Language Arts TCAP score and the same student's 4th, 5th, 6th, 7th, and 8th grade TCAP scores on the Math portion of the *TerraNova*?

From Research Question 4, the following hypotheses were developed and tested:

$Ho4_1$: There is no relationship between the 3rd grade NCE Reading / Language Arts scores of students and the same student's 3rd grade NCE scores on the Math portion of the TCAP.

$Ho4_2$: There is no relationship between the 3rd grade NCE Reading / Language Arts scores of students and the same student's 4th grade NCE scores on the Math portion of the TCAP.

$Ho4_3$: There is no relationship between the 3rd grade NCE Reading / Language Arts scores of students and the same student's 5th grade NCE scores on the Math portion of the TCAP.

$Ho4_4$: There is no relationship between the 3rd grade NCE Reading / Language Arts scores of students and the same student's 6th grade NCE scores on the Math portion of the TCAP.

$Ho4_5$: There is no relationship between the 3rd grade NCE Reading / Language Arts scores of students and the same student's 7th grade NCE scores on the Math portion of the TCAP.

$Ho4_6$: There is no relationship between the 3rd grade NCE Reading / Language Arts scores of students and the same student's 8th grade NCE scores on the Math portion of the TCAP.

In order to test $Ho4_1$, $Ho4_2$ $Ho4_3$, $Ho4_4$, $Ho4_5$, and $Ho4_6$, a statistical analysis was performed to determine a correlation coefficient. The interval scale correlation coefficient used in this case was the Pearson product-moment coefficient (r). The predetermined alpha (α) is .05. Table 6 shows the results of the data analysis.

Table 6

Correlation of 3rd Grade Reading/Language Arts Scores and Math Scores

		3rd grade Reading/Language	
	N	*r*	*p*
3rd grade Math	49	.73	<.01
4th grade Math	49	.69	<.01
5th grade Math	49	.61	<.01
6th grade Math	49	.67	<.01
7th grade Math	49	.74	<.01
8th grade Math	49	.70	<.01

$\alpha = .05$

As shown in Table 6, a significant relationship exists between students' 3rd grade Reading/Language Arts NCE scores and the same students'3rd, 4th, 5th, 6th, 7th, and 8th grade Math NCE scores. Because the *p*-value in the tested hypotheses, $Ho4_1$, $Ho4_2$, $Ho4_3$, $Ho4_4$, $Ho4_5$, and $Ho4_6$ (<.01), is less than the predetermined alpha (.05), I rejected the null hypotheses ($Ho4_1$, $Ho4_2$, $Ho4_3$, $Ho4_4$, $Ho4_5$, and $Ho4_6$). The Pearson Correlation Coefficient (*r*) of the tested hypotheses resulted in an *r* equal to or greater than .61(*r*=> .61*),* this indicated a medium-strength positive relationship. The strongest relationship was exhibited between 3rd grade Reading/Language Arts and 7th grade Math (*r*= .74).

Research Question 5

To what extent, if any, do differences exist between boys and girls on 3rd, 4th, 5th, 6th, 7th, and 8th grade TCAP NCE scores on the Reading / Language Arts portion of the *TerraNova*?

From Research Question 5, the following hypotheses were developed and tested:

$Ho5_1$: There is no difference between boys and girls and NCE scores in Reading / Language Arts on the 3rd grade TCAP.

$Ho5_2$: There is no difference between boys and girls and NCE scores in Reading / Language Arts on the 4th grade TCAP.

$Ho5_3$: There is no difference between boys and girls and NCE scores in Reading / Language Arts on the 5th grade TCAP.

$Ho5_4$: There is no difference between boys and girls and NCE scores in Reading / Language Arts on the 6th grade TCAP.

$Ho5_5$: There is no difference between boys and girls and NCE scores in Reading / Language Arts on the 7th grade TCAP.

$Ho5_6$: There is no difference between boys and girls and NCE scores in Reading / Language Arts on the 8th grade TCAP.

An independent-samples t test was conducted to evaluate whether the mean TCAP NCE score on the Reading/Language Arts portion of the *TerraNova* would differ depending upon the gender of a student. The NCE score was the test variable and the grouping variable was whether a student was male or female. Table 7 shows the results of the data analysis.

Table 7

Comparison of Reading/Language Arts Scores by Gender

Subtest	*N*	*M*	*t*	*df*	*p*
3rd grade					
Female	19	58.35	2.33	47	.02
Male	30	44.57			
4th grade					
Female	19	55.46	1.76	47	.09
Male	30	46.57			
5th grade					
Female	19	54.46	2.64	47	.01
Male	30	44.57			
6th grade					
Female	19	61.15	2.13	47	.04
Male	30	44.57			
7th grade					
Female	19	57.92	2.72	47	.01
Male	30	48.87			
8th grade					
Female	19	58.65	3.09	47	<.01
Male	30	44.57			

*critical value of $t = 1.68$ $\alpha = .05$

As shown in Table 7, it was concluded that a significant difference existed in 3rd, 5th, 6th, 7th, and 8th grade Reading / Language Arts scores between female and male students. Third grade female students had a higher mean NCE than male students. The results, $t(47) = 2.33$, $p = .02$, indicated that a statistical significance existed and I rejected the null hypothesis ($Ho5_1$). The η^2 (.06) indicated a medium effect size.

Fifth grade female students had a higher mean NCE than male students. The results, $t(47) = 2.64$, $p = .01$, indicated that a statistical significance existed and I rejected the null hypothesis ($Ho5_3$). The η^2 (.07) indicated a medium effect size.

Sixth grade female students had a higher mean NCE than male students. The results, $t(47) = 2.13$, $p = .04$, indicated that a statistical significance existed and I rejected the null hypothesis ($Ho5_4$). The η^2 (.05) indicated a low/medium effect size.

Seventh grade female students had a higher mean NCE than male students. The results, $t(47) = 2.72$, $p = .01$, indicated that a statistical significance existed and I rejected the null hypothesis ($Ho5_5$). The η^2 (.07) indicated a medium effect size.

Eighth grade female students had a higher mean NCE than male students. The results, $t(47) = 3.09$, $p = <.01$, indicated that a statistical significance existed and I rejected the null hypothesis ($Ho5_6$). The η^2 (.09) indicated a medium effect size.

I failed to reject the null hypotheses for 3rd grade Reading /Language Arts.

Research Question 6

To what extent, if any, do differences exist between boys and girls on 3rd, 4th, 5th, 6th, 7th, and 8th grade TCAP NCE scores on the Math portion of the *TerraNova*?

From Research Question 6, the following hypotheses were developed and tested:

$Ho6_1$: There is no difference between boys and girls and NCE scores in Math on the 3rd grade TCAP.

$Ho6_2$: There is no difference between boys and girls and NCE scores in Math on the 4th grade TCAP.

$Ho6_3$: There is no difference between boys and girls and NCE scores in Math on the 5th grade TCAP.

$Ho6_4$: There is no difference between boys and girls and NCE scores in Math on the 6th grade TCAP.

$Ho6_5$: There is no difference between boys and girls and NCE scores in Math on the 7th grade TCAP.

$Ho6_6$: There is no difference between boys and girls and NCE scores in Math on the 8th grade TCAP.

An independent-samples t test was conducted to evaluate whether the mean TCAP NCE score on the Math portion of the *TerraNova* would differ depending upon the gender of a student. The NCE score was the test variable and the grouping variable was whether a student was male or female. Table 8 shows the results of the data analysis.

Table 8

Comparison of Math Scores by Gender

Subtest	*N*	*M*	*t*	*df*	*p*
3rd grade					
Female	19	57.85	1.48	47	.15
Male	30	48.87			
4th grade					
Female	19	51.42	.78	47	.44
Male	30	47.13			
5th grade					
Female	19	54.00	2.12	47	.03
Male	30	41.74			
6th grade					
Female	19	54.73	2.16	47	.04
Male	30	43.39			
7th grade					
Female	19	60.12	1.61	47	.11
Male	30	50.83			
8th grade					
Female	19	60.50	1.38	47	.18
Male	30	53.74			

*critical value of $t = 1.68$ $\alpha = .05$

As shown in Table 7, the analysis demonstrates that a significant difference existed in 5th and 6th grade Math scores between female and male students. Fifth grade female students had a higher mean NCE than male students. The results, $t(47) = 2.12$, $p = .03$, indicated that a statistical significance existed and I rejected the null hypothesis ($Ho6_3$). The η^2 (.05) indicated a low/medium effect size.

Sixth grade female students had a higher mean NCE than male students. The results, $t(47) = 2.16$, $p = .04$, indicated that a statistical significance existed and I rejected the null hypothesis ($Ho6_4$). The η^2 (.05) indicated a low/medium effect size.

The mean NCE Math scores of female students were consistently higher than the scores of male students; however, the results were not statistically significant. The results for 3rd grade, $t(47) = 1.45$, $p = .15$, indicated that a statistical significance did not exist and I failed to reject the null hypothesis ($Ho6_1$).

The results for 4th grade, $t(47) = .78$, $p = .44$, indicated that a statistical significance did not exist and I failed to reject the null hypothesis ($Ho6_1$).

The results for 7th grade, $t(47) = 1.61$, $p = .11$, indicated that a statistical significance did not exist and I failed to reject the null hypothesis ($Ho6_1$).

The results for 8th grade, $t(47) = 1.38$, $p = .18$, indicated that a statistical significance did not exist and I failed to reject the null hypothesis ($Ho6_1$).

Research Question 7

To what extent, if any, do differences exist in TCAP NCE gain scores between students on free/reduced lunch and paid students for 2000, 2001, 2002, 2003, and 2004 on the Reading / Language Arts portion of the *TerraNova*?

From Research Question 7, the following hypotheses were developed and tested:

$Ho7_1$: There is no difference in the 2000 TCAP NCE gain scores between students on free/reduced lunch and paid students on the Reading / Language Arts portion of the *TerraNova*.

$Ho7_2$: There is no difference in the 2001 TCAP NCE gain scores between students on free/reduced lunch and paid students on the Reading / Language Arts portion of the *TerraNova*.

$Ho7_3$: There is no difference in the 2002 TCAP NCE gain scores between students on free/reduced lunch and paid students on the Reading / Language Arts portion of the *TerraNova*.

$Ho7_4$: There is no difference in the 2003 TCAP NCE gain scores between students on free/reduced lunch and paid students on the Reading / Language Arts portion of the *TerraNova*.

$Ho7_5$: There is no difference in the 2004 TCAP NCE gain scores between students on free/reduced lunch and paid students on the Reading / Language Arts portion of the *TerraNova*.

An independent-samples t test was conducted to evaluate whether the mean TCAP NCE gain score on the Reading/Language Arts portion of the *TerraNova* would differ depending upon the free/reduced or paid lunch status of a student. The NCE gain score was the test variable and the grouping variable was whether a student was classified as receiving free or reduced lunch or whether the student paid full lunch price. Table 9 shows the results of the data analysis.

Table 9

Comparison of NCE Gain Scores in Reading/Language Arts

Subtest	*N*	*M*	*t*	*df*	*p*
2000 NCE gain scores					
Free/Reduced	30	5.57	1.27	47	.21
Paid	19	3.47			
2001 NCE gain scores					
Free/Reduced	30	2.27	.93	47	.36
Paid	19	4.47			
2002 NCE gain scores					
Free/Reduced	30	2.87	3.67	47	<.01
Paid	19	14.26			
2003 NCE gain scores					
Free/Reduced	30	1.13	2.19	47	.03
Paid	19	8.00			
2004 NCE gain scores					
Free/Reduced	30	4.07	1.34	47	.19
Paid	19	0.00			

*critical value of t =1.68 α= .05

As shown in Table 9, the results of the analysis varied across grade levels. The mean gain scores for 2001, 2002, and 2003 were greater for students on paid lunch status than for students on free/reduced lunch. Students classified as free or reduced lunch exhibited greater gains in 2000 and 2004. However, only the gain scores for 2002 and 2003 were determined to have been statistically significant.

The results for 2002 gain scores, $t(47) = 3.67$, p= <.01, indicated that a statistical significance existed and I rejected the null hypothesis ($Ho7_3$). The η^2 (.13) indicated a medium/large effect size.

The results for 2003 gain scores, $t(47) = 2.19$, $p = .03$, indicated that a statistical significance existed and I rejected the null hypothesis ($Ho7_4$). The η^2 (.05) indicated a low/medium effect size.

The results for 2000 gain scores, $t(47) = 1.27$ $p = .21$, indicated that a statistical significance did not exist and I failed to reject the null hypothesis ($Ho7_1$).

The results for 2001 gain scores, $t(47) = .93$ $p = .36$, indicated that a statistical significance did not exist and I failed to reject the null hypothesis ($Ho7_2$).

The results for 2001 gain scores, $t(47) = 1.34$ $p = .19$, indicated that a statistical significance did not exist and I failed to reject the null hypothesis ($Ho7_5$).

Research Question 8

To what extent, if any, do differences exist in TCAP NCE gain scores between students on free or reduced lunch and paid students for 2000, 2001, 2002, 2003, and 2004 on the Math portion of the *TerraNova*?

From Research Question 8, the following hypotheses were developed and tested:

$Ho8_1$: There is no difference in 2000 TCAP NCE gain scores between students on free or reduced lunch and paid students on the Math portion of the *TerraNova*.

$Ho8_2$: There is no difference in 2001 TCAP NCE gain scores between students on free or reduced lunch and paid students on the Math portion of the *TerraNova*.

$Ho8_3$: There is no difference in 2002 TCAP NCE gain scores between students on free or reduced lunch and paid students on the Math portion of the *TerraNova*.

$Ho8_4$: There is no difference in 2003 TCAP NCE gain scores between students on free or reduced lunch and paid students on the Math portion of the *TerraNova*.

$Ho8_5$: There is no difference in 2004 TCAP NCE gain scores between students on free or reduced lunch and paid students on the Math portion of the *TerraNova.*

An independent-samples t test was conducted to evaluate whether the mean TCAP NCE gain score on the Reading/Language Arts portion of the *TerraNova* would differ depending upon the free or reduced or paid lunch status of a student. The NCE gain score was the test variable and the grouping variable was whether a student was classified as receiving free or reduced lunch or whether the student paid full lunch price. Table 10 shows the results of the data analysis.

Table 10

Comparison of NCE Gain Scores in Math

Subtest	*N*	*M*	*t*	*df*	*p*
2001 NCE gain scores					
Free/Reduced	30	.37	.36	47	.72
Paid	19	1.89			
2002 NCE gain scores					
Free/Reduced	30	1.03	.11	47	.92
Paid	19	.68			
2003 NCE gain scores					
Free/Reduced	30	8.27	1.94	47	.06
Paid	19	2.63			
2004 NCE gain scores					
Free/Reduced	30	1.43	.36	47	.72
Paid	19	2.42			

*critical value of t =1.68 α= .05

As shown in Table 10, the mean gain scores of students on free/reduced lunch status were greater than students who paid for lunch in 2002, and 2003. In 2001, students classified as paid exhibited greater NCE gain scores than students who were on free or reduced lunch.

There was no statistically significant difference determined to have existed between students who paid for their lunch and students who received free or reduced lunch. I failed to reject hypotheses $Ho8_1$, $Ho8_2$, $Ho8_3$, $Ho8_4$, and $Ho8_5$.

Summary

This chapter included a listing of 8 research questions along with 50 hypotheses formulated to test the questions. Analyses of data, along with tables presenting the findings of the analysis, are included. In chapter 5, the findings are summarized and interpreted and from the analysis, conclusions are made. In addition, limitations and recommendations for practice and further considerations are given.

CHAPTER 5

FINDINGS, CONCLUSIONS, LIMITATIONS, AND RECOMMENDATIONS

The purpose of this study was to examine the relationships between and among variables encompassing student performance on the Reading / Language Arts and Math sections of the *TerraNova* portion of the Tennessee Comprehensive Assessment Program over a six-year period within a rural public school setting in East Tennessee. Variables such as initial student scores on the Reading / Language Arts and Math portions of the *TerraNova* portion of the TCAP, subsequent scores of the same students in following years, participation in the school's free or reduced lunch program, and student gender were examined. A summary of the findings, conclusions, and recommendations for further research and for practice follow.

Summary of the Study

In recent years, high stakes testing has risen to the forefront as a means to measure student progress. The interest in the progress of individual students as they weave their way through the American education system has been on the upswing, due in large part to the reaction of elected officials to the public outcry for the increased accountability of public schools, as well as the administrators and educators who staff them, for the quality of education received by students. The resulting scores of these typically annual tests are used by local, state, and national officials to determine how well equipped students are in comparison to other students educated in neighboring districts, states, and even other global settings.

The review of literature documented the historical perspectives surrounding continuing educational reform regarding assessment and accountability linked to political actions and public outcry, such as: the *Elementary and Secondary Education Act*, the release of *A Nation at Risk*, and the adoption of the *No Child Left Behind Act*. The introduction of high stakes testing (using both norm-referenced and criterion-referenced exams) as a means to measure student growth in

academic performance in an effort to ensure greater accountability was reviewed. It was noted that this increased call for accountability encompassed those within the educational system that the public deemed responsible, including policy makers, local administrators, and educators. It was noted that Tennessee has led the way nationally in efforts to ensure the quality of student education within their public school systems with the adoption of the Tennessee Value Added Assessment Program, conceived and designed by William Sanders. This program purports to accurately measure yearly student growth by using a mixed methodology of statistical analysis designed to remove the effects of variables such as previous level of ability, socioeconomic status, and mobility. Finally, issues such as the use of high stakes testing, TVAAS and the use of value-added assessment models, some barriers to student achievement, the importance of prior achievement as a predictive variable, and student accountability were examined.

This quasi-experimental design of this study focused on the relationship between and among variables unique to individual students, such as initial student scores on the Reading / Language Arts and Math sections of the *TerraNova* portion of the TCAP, subsequent scores of the same students in following years, participation in the school's free or reduced lunch program, student gender, and a comparison of gain scores over a six-year period. The sample consisted of students in the third-, fourth-, fifth-, sixth-, seventh-, and eighth-grade who were enrolled in Grainger County Schools during the 1999-2004 school years and who took the *TerraNova* for six consecutive years beginning in 1999 and including 2000, 2001, 2002, 2003, and concluding with 2004. Of the 58 students who were third-, fourth-, fifth-, sixth-, seventh-, and eighth-graders during 1999-2004, 9 were excluded because they did not take the *TerraNova* all six years of the study. The resulting population was 49 students and it involved 14 teachers and a single elementary school.

Summary of Findings

The analysis focused on eight research questions. The independent variables for the study were gender, socioeconomic status, and grade level (test score reporting year). The dependent variables consisted of *TerraNova* value-added scores (NRT) and proficiency scores (CRT) translated to Normal Curve Equivalent (NCE) scores on the Reading / Language Arts and Mathematics portion of the TCAP. A combination of *t* test for independent samples, examination of effect size using eta square (η^2), and an analysis of data to determine correlation coefficient using Pearson's product moment coefficient (r) were used in 50 hypotheses. The findings of the study provide answers to the 8 research questions. The following restates each research question and provides a summary of the findings related to it.

Research Question 1

To what extent, if any, are there relationships between students' 3rd grade TCAP scores in Reading / Language Arts and Math and the same students' scores as the students progressed from the 3rd through the 8th grades?

An analysis of data to determine correlation coefficient using Pearson's correlation coefficient was conducted to test 10 hypotheses that were developed in an effort to answer this research question. Using a predetermined level of alpha ($\alpha = .05$) to determine the possibility that the findings were due to chance and associating a correlation coefficient (r) of 0 to indicate that the two variables being tested have no relationship with one another, the analysis of data produced consistent results. Of the 10 null hypotheses tested, all were rejected due to findings indicating that a statistically significant relationship existed between the scores of the same subjects across grade levels. Strong positive relationships were found to exist between the TCAP scores of the students in following subjects and across all grade levels, with the weakest r (640)

demonstrated between 3rd grade Reading/Language Arts scores and 8th grade Reading/Language Arts scores.

Research Question 2

To what extent, if any, are there differences between the 3rd, 4th, 5th, 6th, 7th, and 8th grade TCAP NCE scores of students on free or reduced lunch and students on paid lunch on the Reading / Language Arts portion of the *TerraNova*.

A t-test for independent samples was used to test 5 hypotheses that were developed in an effort to answer this research question. Using a predetermined level of alpha ($\alpha = .05$) to determine the possibility that the findings were due to chance and associating the critical value of *t* (1.68) with the degrees of freedom *(df* = 47), the analysis of data produced consistent results for 4th, 6th, 7th, and 8th grades. Of the 6 null hypotheses tested, 5 were rejected due to findings that indicated a statistically significant difference existed in Reading/language Arts NCE scores. It was determined that students on paid lunch scored higher across certain grade levels than students on free and reduced lunch. The eta square for grades 4th ($\eta^2 = .07$), 5th ($\eta^2 = .05$), and 8th ($\eta^2 = .07$) indicated a medium effect size. The eta square for grades 6th ($\eta^2 = .14$) and 7th ($\eta^2 = .15$) indicated a large effect size.

Despite the fact that the mean NCE of students on free/reduced lunch was significantly less than the mean NCE of students on paid lunch status, the calculated value of *t* (1.49) was less than the determined critical value of *t* (1.68) and I failed to reject the null hypotheses for 3rd grade Reading /Language Arts.

Research Question 3

To what extent, if any, are there differences between the 3rd, 4th, 5th, 6th, 7th, and 8th grade TCAP NCE scores of students on free or reduced lunch and students on paid lunch on the Math portion of the *TerraNova*.

A t-test for independent samples was used to test a total of 5 hypotheses which were developed in an effort to answer this research question. Using a predetermined level of alpha (α = .05) to determine the possibility that the findings were due to chance and associating the critical value of *t* (1.68) with the degrees of freedom *(df* = 47), the analysis of data produced significant results for 3rd, 4th, 5th, and 6th grades. Of the 6 null hypotheses tested, 4 were rejected due to findings that indicated a statistically significant difference existed in Math NCE scores. It was determined that students on paid lunch scored higher across certain grade levels than students on free and reduced. The eta square for grades 3rd (η^2 = .05), 4th (η^2 = .08), 5th (η^2 = .05), and 6th (η^2 = .07) indicated a medium effect size.

Despite the fact that the mean NCE score of 7th and 8th grade students on free/reduced lunch was less than the mean NCE of students on paid lunch status, the calculated values of *t*, 7th =1.26 (*p*=.21) and 8th = 1.67 (*p*=.10) were less than the determined critical value of *t* (1.68) and I failed to reject the null hypotheses for 7th and 8th grade Math.

Research Question # 4

To what extent, if any, do relationships exist between a student's 3rd grade Reading / Language Arts TCAP score and the same student's 4th, 5th, 6th, 7th, and 8th grade TCAP scores on the Math portion of the *TerraNova*?

In order to test this question, six hypotheses were developed and a statistical analysis was performed to determine a correlation coefficient. The interval scale correlation coefficient used in this case was the Pearson product-moment coefficient (*r*) with a predetermined alpha (α) of .05. A statistically significant relationship was determined to exist between a student's 3rd grade NCE scores on the Reading / Language Arts portion of the TCAP and the variable examined in

each of the six hypotheses. All of the observed relationships were considered to be medium strength to strong, positive relationships. The highest *r* value was recorded between 3rd grade Reading / Language Arts NCE scores and 7th grade Math NCE scores (r= .74, p= <.01). The lowest *r* value observed occurred between 3rd grade Reading / Language Arts NCE scores and 5th grade Math NCE scores (r= .61, p= <.01). Because this study was not a true experimental study, relationship between these two variables in no way implies causation.

Research Question # 5

To what extent, if any, do differences exist between boys and girls on 3rd, 4th, 5th, 6th, 7th, and 8th grade TCAP NCE scores on the Reading / Language Arts portion of the *TerraNova*?

A t-test for independent samples was used to test a total of 6 hypotheses which were developed in an effort to answer this research question. An independent-samples t test was conducted to evaluate whether the mean TCAP NCE score on the Math portion of the *TerraNova* would differ depending upon the gender of a student. The NCE score was the test variable and the grouping variable was whether a student was male or female. Using a predetermined level of alpha (α = .05) to determine the possibility that the findings were due to chance and associating the critical value of *t* (1.68) with the degrees of freedom *(df* = 47), the analysis of data produced statistically significant results for 3rd, 5th, 6th, 7th, and 8th grades. Of the 6 null hypotheses tested, 5 were rejected due to findings that indicated a statistically significant difference existed in Reading/language Arts NCE scores. It was determined that female students scored higher across certain grade levels than males. The eta square for grades 3rd (η^2 = .06), 5th (η^2 = .07), and 6th (η^2 = .05), 7th (η^2 = .07) and 8th (η^2 = .09) indicated a low/medium effect size.

Despite the fact that the mean NCE of female students was significantly greater than the mean NCE of male students, p=.09 exceeded α (.05) and I failed to reject the null hypotheses for 4th grade Reading /Language Arts.

Research Question # 6

To what extent, if any, do differences exist between boys and girls on 3rd, 4th, 5th, 6th, 7th, and 8th grade TCAP NCE scores on the Math portion of the *TerraNova*?

An independent-samples t test was conducted to evaluate whether the mean TCAP NCE score on the Math portion of the *TerraNova* would differ depending upon the gender of a student. The NCE score was the test variable and the grouping variable was whether a student was male or female. Using a predetermined level of alpha ($\alpha = .05$) to determine the possibility that the findings were due to chance and associating the critical value of *t* (1.68) with the degrees of freedom *(df* = 47), the analysis of data produced statistically significant results for 5th and 6th grades. Despite the fact that female students consistently scored higher than male students, of the 6 null hypotheses tested, only 2 were rejected due to findings that indicated a statistically significant difference existed in Math NCE scores between males and females. The eta square for grades 5th ($\eta^2 = .05$), and 6th ($\eta^2 = .05$) indicated a low/medium effect size.

Despite the fact that the mean NCE of female students was greater than the mean NCE of male students for grades the calculated value of *t* was less than the determined critical value of *t* (1.68) and *p* exceeded α (.05) and I failed to reject the null hypotheses for 3rd, 4th, 7th, and 8thgrade Math.

Research Question # 7

To what extent, if any, do differences exist in TCAP NCE gain scores between students on free/reduced lunch and paid students for 2000, 2001, 2002, 2003, and 2004 on the Reading / Language Arts portion of the *TerraNova*?

An independent-samples t test was conducted to evaluate whether the mean TCAP NCE gain score on the Reading/Language Arts portion of the *TerraNova* would differ depending upon the free/reduced or paid lunch status of a student. The NCE gain score was the test variable and the grouping variable was whether a student was classified as receiving free or reduced lunch or whether the student paid full lunch price.

The results of data analysis testing the hypotheses of this question revealed a statistically significant difference between the TCAP NCE gain scores of students on free/reduced lunch and paid students on the Reading / Language Arts portion of the *TerraNova*. The findings indicate that a statistically significant difference in the "value-added" effect on student test scores was exhibited on gain scores for 2002 and 2003. The NCE gain scores for this year provide evidence that students on free/reduced lunch status scored significantly lower than did the students listed as paid. The results for 2002 gain scores (5^{th} to 6^{th} grades), $t(47) = 3.67$, $p = <.01$, indicated that a statistical significance existed and I rejected the null hypothesis ($Ho7_3$). The η^2 (.13) indicated a medium/large effect size.

The results for 2003 gain scores (6^{th} to 7^{th} grades), $t(47) = 2.19$, $p = .03$, indicated that a statistical significance existed and I rejected the null hypothesis ($Ho7_4$). The η^2 (.05) indicated a low/medium effect size.

The results for 2000, 2001, and 2004 produced no significant differences between the gain scores of students who were listed as free or reduced lunch and students listed as paid.

Research Question # 8

The findings of this analysis resulted in no significant difference between the average TCAP NCE gain of students on free/reduced lunch and paid students on the Math portion of the *TerraNova* for 2000, 2001, 2002, 2003, and 2004.

Conclusions

The study focused on the performance of students over a six-year period. Student NCE scores on the Reading / Language Arts and Math components of the *TerraNova* portion of the TCAP were used as the dependent variable while environmental variables such as gender, grade level, and socioeconomic status were used to compare yearly gains in an effort to discover if and differences or relationships existed. While some of the findings were mixed and failed to outline a clear difference or relationship, instances within the analysis were noted indicating statistically significant results confirming the existence of distinct differences and clear relationships. There were several conclusions drawn from this study.

Conclusion #1

The results of this study echo the findings of earlier research indicating the importance of initial standardized test scores produced in the early grades as a possible predictor of future success on tests as the student progress through the school system. The strong positive relationships exhibited between a student's 3rd grade NCE scores in Reading / Language Arts and Math and the same student's NCE scores in the same subjects in later subsequent years underscore the importance of needed emphasis within these subject areas in the early grades. Primary grade educators whose efforts are directed toward facilitating a student's academic performance toward an individual high level of success on the Reading / Language Arts and

Math subject areas of early grade assessment would undoubtedly be laying the groundwork for the future academic success of their students, excluding the interaction of other variables.

Conclusion # 2

As a variable taken into account during the student's academic journey, a child's socioeconomic status (defined here as free/reduced lunch status) plays a major role in determining individual student academic success. When comparing the mean NCE scores in Reading / Language Arts and Math of students on free/reduced lunch with their classmates on paid lunch status, the results of this study clearly indicate that those on free/reduced lunch are at a disadvantage. All grade levels tested exhibited double-digit differences between the average mean scores of students on paid or free/reduced lunch. The students on free/reduced lunch remained behind their cohorts through the entire six year data set. The only areas of analysis which did not support what the majority of the findings indicated, that being a statistically significant difference in NCE scores, occurred in 3rd grade Reading / Language Arts and 7th/8th grade Math. Even though theses three areas failed to indicate a statistical significance, the average mean scores of students on free/reduced lunch remained far below the scores of their counterparts. The lack of a statistically significant difference may be an indication of effective teaching directed at economically disadvantaged students.

Conclusion # 3

The results indicate that there is no correlation between a student's free/reduced lunch status and their 3rd grade reading scores while at the same time defining a statistically significant positive relationship between socioeconomic status and 3rd grade Math scores. This dichotomy might be explained by intervention programs in place at the system or building level. These programs aimed at providing students an equal start on reading success rather than math skills,

along with excellent facilitation of the programs by educators in the early grades, might be the difference.

Conclusion # 4

The results of this study corroborate the notion that reading is a fundamental skill, necessary and directly linked to a student's present and future academic success. All relationships tested revealed positive correlations between a student's Reading / Language Arts level of ability and the same student's success in Math. The skills needed to perform well on standardized tests designed to measure a student's level of reading mastery are not only utilized by the student in other subject areas across the curriculum, these skills are becoming increasingly important in Math. This is most likely due in part to an increased focus within the curriculum on word problems designed around real life scenarios designed to activate a student's higher-order thinking skills.

Conclusion # 5

The question regarding a difference in NCE scores between males and females turned up mixed results. Analysis of data from the early grades, 3rd and 4th, produced only one example of significant differences between males and females. Female students had a much higher mean NCE scores in Reading / Language Arts. It is worth noting here that even when a statistical difference was not reported, the average NCE scores of females in Reading /Language Arts and Math were consistently higher than the average NCE sores of males. The range of difference never dipped below a 4-point margin, with females scoring higher than males.

Limitations

The study has limitations that need to be mentioned. First, there are varying degrees of teachers' skills. Even though all teachers are highly qualified by NCLB standards, the teachers possess different levels of expertise and years of experience. Another limitation concerns the small size of the population used for the study. The results of this study may not be generalized to other populations.

Recommendations for Practice

Several recommendations for practice can be made as a result of this study. The first would be that every effort should be made within a school system to ensure that students exit the early grades on as equal an academic footing as possible. This might possibly be accomplished by utilizing early intervention programs that target students at risk due to variables including, but not limited to socioeconomic status, parental involvement, and reading readiness. This study did provide support that a direct relationship exists between prior achievement / socioeconomic status and future academic success, viewed as increased achievement on standardized exams. The fact that some of the findings of this study were inconsistent suggests that other factors played a role in student achievement. Another recommendation would be to provide the intervention on a voluntary basis for students. This could possibly come in the form of either a before- or after-school program which could include the entire student population, not just those deemed at risk.

Recommendations for Further Research

Several recommendations for further research were developed as a result of this study. With the demands of NCLB, educators have no choice but to closely monitor students' achievement. A greater emphasis might be placed on student accountability by filtering out the possible effect of external variables such as socioeconomic status and previous achievement by

supplying targeted students with the means to excel through intervention programs. The need for additional research would prompt these recommendations:

1. A longitudinal study of the relationship between the socioeconomic status and students' Reading / Language Arts and Math achievement from grades 3 through 12.
2. Replication of the study, using a larger sample, to evaluate the relationship between the interaction of numerous independent variables and student gains. The variables should include ability level, previous achievement, gender, special education status, and socioeconomic status.
3. Replication of the study using an urban school setting.
4. Replication of the study in a minority school setting.

Summary

The majority of the findings of this study parallel the earlier results cited in the Literature Review. Authors such as Coleman (1966), Willms (1986), and Berliner and Biddle ((1995) argued the fact that factors such as school climate, home environment, and socioeconomic background have a strong relationship to student academic achievement. The results of Research Question 7 clearly echoed these earlier studies and indicated that a relationship existed between students' Reading/language Arts NCE scores and their reported free or reduced lunch status.

One possible reason that the scores of students on free or reduced lunch were not a great deal lower than those on paid status could be a manifestation of the intervention program in place at the school site tested. The Reading intervention program included those 5th grade students deemed to be at-risk due to a) listed as free or reduced lunch and b) listed as non-proficient according to their 4th grade Reading NCE scores. Of the 12 students enrolled in the program, 9 scored at the proficient level on the 5th grade Reading/Language Arts section of the TCAP. This example could be used as a model to help ensure success for at-risk students.

The possibility of using a student's prior achievement as a predictor for future success was presented in the studies of Smith and Meyer (1995) and Brown (1994). The results of this study showed that strong positive relationships existed between students' initial 3rd grade scores in Reading/Language Arts and both Reading and Math scores for subsequent years.

One major area of this study was in disagreement with the widely held view that males generally score higher that females in Math. The results of this study clearly indicated that NCE scores of females were significantly higher that males in all but one grade level.

REFERENCES

Atkins, J. (2005) *The association between the use of accelerated math and students' math achievement.* Doctoral Dissertation. Johnson City, TN: East Tennessee State University.

Baker, A. P., & Xu, D. (1995*). The measure of education: A review of the Tennessee value added assessment system*. Office of Education Accountability. Nashville: TN.

Ballou, D. (2002, Summer). Sizing up test scores. *Education Next,* 10-15.

Berliner, D. C. & Biddle, B. J. (1995). *The Manufactured Crisis: Myths, Fraud, and the Attack on America's Public Schools.* Reading, MA: Addison-Wesley.

Bracey, G. (1997) The seventh bracey report on the condition of public education. *Phi Delta Kappan,* . 79 (2). Retrieved November 15, 2005 from: http://www.pdkintl.org/kappan/kbra9710.htm

Bratton, S. E., Horn, S. P., & Wright, S. P. (1996). *Using and interpreting Tennessee's value-added assessment system: A primer for teachers and principals*. Value Added Research and Assessment Center. Knoxville: TN.

Brown, S. (1994). School effectiveness research and the evaluation of schools. *Evaluation and Research in Education, 8,* 55-68.

Cannon, T. (2000, Fall). Information about norm-referenced scores. *Tennessee Assessment Newsletter,* 1, 2.

CBT McGraw-Hill. (1996). *Terra Nova prepublication technical bulletin.* Monterey, CA: Author.

CBT McGraw-Hill. (1997). *Teacher's guide to the TerraNova.* Monterey, CA: Author.

Chubb, J. E., & Moe, T. M. (1990). Politics, markets, and America's schools. Washington, DC: The Brookings Institution.

Cohen, D. (1995). What standards for national standards? *Phi Delta Kappan, 76*, 751-757.

Coleman, A. L. (2000, May). Fair testing: How schools should protect students' rights in high-stakes testing. *American School Board Journal, 187*(6), 32-35.

Coleman, J. (1966). Equality of educational opportunity. Washington, DC: United States Government Printing Office.

Cooper, J. M., Garrett, S. S., Leighton, M. S., Martorella, P. H., Morine-Dershimer, G. G., Sakder, et al. D., (1990). *Classroom teaching skills* (4th ed.). Lexington, MA: D. C. Heath.

Cremin, L. A. (ED.). (1957). *The republic and the school: Horace Mann on the education of free men.* New York: Thomas Y Crowell.

Cronbach, L. J. (1963). Course improvements through evaluation. *Teachers College Record, 64,* 672-683.

Cunningham, G. K. (1991). Educational testing and educational reform: Some misconceptions. *Contemporary Education, 56,* 238.

Darling-Hammond, L., & Archer, C. (1991). Accountability mechanisms in big city school systems. Retrieved September 15, 2004 from http://www.ericdigests.org/pre-9220/big.htm

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives, 8*(1). Retrieved November 11, 2004 from http://epaa.asu.edu/epaa/v8n1

Dorans, N. (2000). *Distinctions among classes of linkages*. College Board Report # RN-11. New York: The College Board.

Dossett, D., & Munoz, M. A. (2003). *Classroom accountability: A value-added methodology.* Paper presented at the American Educational Research Association in Chicago, IL. Retrieved October 5, 2004 from http://www.jefferson.K12.Ky.US/Departments/AcctResPlan/PDF/ClassroomAccountability.pdf

Du, Y., & Heistad, D. (1999, April). School performance accountability in Minneapolis Public Schools. Paper presented at the Annual American Educational Research Association Conference in Montreal, Canada.

Education Commission of the States. (2001). *No state left behind: The challenges and opportunities of ESEA 2001.* Publication No. GP-02-01. Denver.

FairTest. (2002). National Center for Fair and Open Testing. Retrieved November 27, 2004, from http://www.fairtest.org/facts/mctfcat.html

Fuchs, D. & Fuchs, L. S. 1995. Special education can work. *Issues in educational placement: Students with emotional and behavioral disorders (pp. 363-77)* edited by James M. Kauffman, John W. Lloyd, Daniel P. Hallahan, and Terry A. Astuto. Hillsdale, NJ: Erlbaum.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction* (6th ed.). White Plains, NY: Longman.

Glickman, C. D. (1990). Open accountability for the '90s: Between the pillars. *Educational Leadership, 38-42*.

Good, T. L. et al (1983). Teaching effectiveness: Research findings and policy implications. Paper presented at the National Institute of Education Conference on State and Local Policy Implications of Effective School Research.

Green, S. B., & Salkind, N. J. (2004). *Using SPSS for Windows and Macintosh: Analyzing and understanding data* (4th ed.). Upper Saddle River, NJ: Pearson/Prentiss Hall.

Gutek, G.L. (1988). *Education and schooling in America*. (2nd eds.) Englewood Cliffs, NJ: Prentice Hall.

Hanson, B., Harris, D., Pommerich, M., Sconing, J., & Yi, Q. (2001). *Suggestions for the evaluation and use of concordance results. ACT research report series*. Iowa City, IA: American College Testing Program.

Hill, D. (2000). He's got your number. *Teacher Magazine, 11* (8), 42-47.

Holbein, M. F. D., (1998). Will standards improve academic achievement? *Education, 118*, 559-563.

Houston, P. (2005, February). NCLB: Dreams and nightmares. *Phi Delta Kappan,*79, 469-470.

Kaplan, L. S., & Owings, W. A. (2001). Enhancing teacher and teaching quality: Recommendations for principals. *NASSP Bulletin 85*(628), 64-73.

Kearns, D. T. (1988). An education recovery plan for America. *Phi Delta Kappan,* 565-570.

Knight, G. B. (1989). *Philosophy & education.* (2nd ed.). Berrien Springs, MI: Andrews University Press.

Kupermintz, H. (2003). *School reform proposals: The research evidence.* Retrieved February 7, 2005, from http://www.asu.edu/educ/epsl/EPRU/documents/EPRU%202002-101/Chapter%2011-Kupermintz-Final.htm

Lansa, P. J., & Potter, J. H. (1984). *Building a bridge to span the minority-majority achievement gap.* Paper presented at the National Conference on Desegregation in Postsecondary Education, Durham, NC.

Linn, R. L. (2001, April). The design and evaluation of educational assessment and accountability systems. *CSE Technical Report 539.* CRESST/University of Colorado at Boulder. Retrieved September 21, 2004 from http://www.cse.ucla.edu/CRESST/Reports/TR539.pdf

Long, D., & Cass, M. (2001). Analysist rocks education boat with theory that teacher, not economic status, is more important. *The Tennessean, 42*.

Manno, B. (1998). Five tips for good schools. *Triangle Business Journal, 13* (43), 58.

Meyers, R. H. (1996). Value-added indicators of school performance. In E.A. Hanushek & D. W. Jorgenson (Eds.), *Improving America's schools: The role of incentives* (pp. 197-223). Washington, DC: Brookings.

Meyers, R. H. (2000). Value-added indicators: A powerful tool for evaluating science and mathematics programs and policies. Madison, WI: National Institute for Science Education.

.

Meier, D. (1995, January). How our schools could be. *Phi Delta Kappan*, 77, 369-373.

Montouri, J. (1994, December). Noteworthy. *USA Today, 123* (2595), p. 2.

Munoz, M. A., & Dossett, D. (2001). Equity and excellence: The effect of school and socio-demographic variables on student achievement. *Journal of School Leadership, 11,* 120-139.

National Commission on Excellence in Education. (1983). A nation at risk: the imperative for educational reform. Retrieved November 18, 2005 from: http://www.ed.gov/pubs/NatAtRisk/index.htm

Nelson, K. (2000). Measuring the intangibles. *Classroom leadership: A newsletter for k-12 classroom teachers, 3*(5), p. 1. (Association for Supervision and Curriculum Development (ASCD): A Supplement to Educational Leadership)

Nickerson, R. S. (1989, December). New direction in educational assessment. *Educational Researcher,* 3-6.

Nitko, A. J. (1983). *Educational tests and measurement: An introduction.* New York: Hartcourt Brace Jovanovich.

Odden, A. (1990, June/July). Educational indicators in the United States: The need for analysis. *Educational Researcher*, 24-29.

Orfield, G., & Wald, J. (2000). The high-stakes testing mania hurts poor and minority students the most: Testing, testing. *The Nation, 270* (22), 38.

Ornstein, A. C. (1988, September). The evolving accountability movement. *Peabody Journal of Education, 65,* 19-23.

Parkay, F.W., & Hass, G. (2000). Curriculum planning: A contemporary approach. (7th ed.) Needham Heights, MA: Allyn and Bacon, A Pearson Education Company.

Park, S. H. (2005, January). *Value added training rollout*. Paper presented at a workshop on value-added by the Tennessee Department of Education, Knoxville, TN.

Pipho, C. (1998, January). The value-added side of standards. *Phi Delta Kappan,* 79, 341-342.

Popham, W. J. (1998) *Your school should not be evaluated by standardized test scores!* AASA Online. Retrieved June 30, 2004 from: http://lchc.ucsd.edu//MCA/Mail/xmcamail.1999_04.dir/0173.html

Porter, A. C. (1988, July/August). Understanding teaching: A model for assessment. *Journal of Teacher Education,* 2-7.

Pulliam, J. D., & Van Patten, J. (2003). *History of education in America.* (8th ed.). NJ: Merrill Prentice Hall.

Rivers, J. C. & Sanders, W. L. (2000). Teacher quality and equity in educational opportunity findings and policy implications. Presented at the Hoover/PRI Teacher Quality Conference, Stanford University, May 12, 2000.

Roeder, P. W. (2000). Education reform and equitable excellence: The Kentucky experiment (Update – June 2000) Race, Poverty, and Urban Schools in Kentucky. Retrieved October 12, 2004, from http://www.uky.edu/%7Eproeder/kerabkupdate2.pdf

Salganik, L. H. (1985). Why testing reforms are so popular and how they are changing education. *Phi Delta Kappan, 66,* 607.

Sanders, W. L., (1993). An overview of the: Tennessee value-added assessment system (TVAAS). The University of Tennessee.

Sanders, W. L. (1998). Value-added assessment. *School Administrator, 55*(11), 1-11.

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

Sanders, W. L. & Horn, S. P. (1995). Educational assessment reassessed: The usefulness of standardized and alternative measures of student achievement as indicators for the assessment of educational outcomes. *Educational Policy Analysis Archives, 3* (6), 1-21.

Sanders, W. L. & Horn, S. P. (1998). Research findings from the Tennessee value-added assessment system (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education, 12,* 247-256.

Sapon-Shevin, M. (1993). Gifted education and the protection of privilege: Breaking the silence, opening the discourse. In *Beyond silenced voices: Class, race, and gender in United States Schools* (pp. 25-44), edited by Lois Weis and Michelle Fine. Albany, NY: State University of New York Press.

Smith, K. B., & Meier, K. J. (1995). The case against school choice. Armonk, NY: M.E. Sharpe.

Steadman, l. C. (1993). The condition of education: Why school reformers are on the right track. *Phi Delta Kappan.* 75, 215-225.

Stevens, J., Estrada, S., & Parkes, J. (2000, April). *Measurement issues in the design of state accountability systems.* Paper presented at the American Educational Research Association, New Orleans, LA.

Stringfield, S, & Herman, R. (1997). Research on effective instruction for at-risk students: Implications for the St. Louis public schools. *Journal of Negro Education, 66*, 258-288.

Summers, A. A. (2002, Summer). Expert measures. *Education Next,* 16-19.

Tennessee Department of Education. (2001). A blueprint for learning. Retrieved: October 16, 2005: http://www.tennessee.gov/education/ci/cistandards2001/cibluept/index.htm

Timar, T. B., & Kirp, D. L. (1989). *Managing educational excellence*. Philadelphia: Falmer Press.

U. S. Department of Education. (2002). *No child left behind: A desktop reference.* Jessup, MD: Education Publication Center.

Walberg, H. J. (1986, September). What works in a nation at risk. *Educational Leadership,* 7-10.

Webb, J. (2005). *An analysis of Tennessee gateway exams and the variables related to student results.* Doctoral Dissertation. Johnson City, TN: East Tennessee State University

Webb, P. (2005). *The associations between the scores on the ACT Test and Tennessee's value-added assessment in 281 Tennessee high schools*. Doctoral Dissertation. Johnson City, TN: East Tennessee State University

Whitehurst, G. J. (2002). *Scientifically based research on teacher quality: Research on teacher preparation and professional development.* Presented to White House Conference on Preparing Tomorrow's Teachers, March 5, 2002. Published as Appendix C of *Improving Teacher Quality State Grants. Title II, Part A. Non-Regulatory Draft Guidance.* Washington, DC: U.S. Department of Education. Retrieved November 12, 2004 from http://tepserver.ucsd.edu/courses/tep129/readings/whithurst.pdf

Willms, (1986). Social class segregation and its relationship to pupils' examination results in Scotland. *American Sociological Review, 51(2),* 224-41.

Wright, D. (1999, July/August). Student mobility: A negligible and confounded influence on student achievement. *The Journal of Educational Research (92)* 6, 347-353.

VITA

DAVID CLOUD

Personal Data	Date of Birth: December 16, 1956 Place of Birth: Somers Point, New Jersey Marital Status: Married
Education	Walters State Community College, Morristown, TN A. S. General Education 1977 Carson Newman College, Jefferson City, TN B. A. Liberal Arts/K-8 Licensure 2000 Carson Newman College, Jefferson City, TN M. Ed. Curriculum and Instruction 2001 East Tennessee State University, Johnson City, TN Ed.D., Educational Leadership and Policy Analysis 2005
Professional Experience	Librarian/Media Specialist Joppa School Grainger County, TN 2000/2001 5th Grade Reading/Language Arts Teacher Joppa School Grainger County, TN 2001/2005