12-2011

Correlation Between the TCAP Test and ThinkLink Learnings Predictive Assessment Series Test in Reading Math and Science in a Tennessee School System.

Jared Edwin Day
East Tennessee State University

Correlation Between the TCAP Test and ThinkLink Learning's Predictive Assessment Series Test in Reading, Math, and Science in a Tennessee School System

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 in Educational Leadership

by

Jared E. Day

December 2011

Dr. Virginia Foley, Chair

Dr. Eric Glover

Dr. James Lampley

Dr. Elizabeth Ralston

Keywords: Benchark Testing, TCAP, Assessment, Standardized Testing, NCLB

ABSTRACT

Correlation Between the TCAP Test and ThinkLink Learning's Predictive Assessment Series Test in Reading, Math, and Science in a Tennessee School System

by

Jared Edwin Day

The purpose of the study was to determine if a correlation existed between the Predictive Assessment Series (PAS) Test, marketed by Discovery Education, and the Tennessee Comprehensive Assessment Program (TCAP) Achievement Test in reading, math, and science for grade 4, grade 6, and grade 8. The study included 4th-grade, 6th-grade, and 8th-grade students during the 2008-2009 school year who had taken the ThinkLink Predictive Assessment Series for reading, math, and science in February 2009 and had taken the TCAP reading, math, and science test in April 2009.

The approach of the study was quantitative in nature. Data were collected from one school system in East Tennessee. The school system had 5 elementary schools and 1 middle school. Data collection tools used in the study included results from the TCAP test using the paper and pencil format and a computer test, the ThinkLink PAS. Student scaled scores were used for determining the degree of correlation between the TCAP and PAS tests. The data were analyzed using the Statistical Program for the Social Sciences.

Based on the analysis and findings of this study, using the ThinkLink PAS test appears to have been successful in predicting how well students will perform on the state

assessment. Overall, the correlations between the PAS and TCAP were consistent across grades, across gender within grade levels, and with Title I and Non-Title I students. The findings also show that it was possible to calculate a predicted TCAP score in reading, mathematics, and science. This was an important finding because the ability of the PAS assessment to predict TCAP scores could be another tool to provide educators the opportunity to target students who are potentially at risk of not meeting state benchmark proficiency levels. Based on the findings, there appears to be a strong relationship between the ThinkLink PAS benchmark assessment and the TCAP assessment in reading, math, and science for grade 4, grade 6, and grade 8. The relationships between PAS and TCAP tests in reading, math, and science were consistent across gender within grade levels. According to the results of the test of homogeneity of slopes, the relationships between PAS and TCAP tests in reading, math, and science were also consistent across Title I and Non-Title I schools. The test of homogeneity of slopes showed the slopes regression lines for the scores of Title I and Non-Title I students were the same (parallel) for grade 4, grade 6, and grade 8. Overall, the correlations between PAS and TCAP scores for Title I and Non-Title I students were moderately strong to very strong. The predictive validity of the PAS provides educators valuable time to reteach grade level skills to students who are at risk of scoring nonproficient on the TCAP.

DEDICATION

I dedicated this study to my parents who gave me constant support and encouragement which helped me stay on track in order to fulfill a personal goal. They both instilled in me at an early age the value of education and having a strong work ethic. Their strong belief in hard work has made me a stronger individual. An anonymous quote that my grandfather often recited comes to mind, “Once a task has begun, never leave it until it’s done. Though the task be great or small, do it well or not at all.”

ACKNOWLEDGEMENTS

I would like to extend my thanks to my family and friends for providing me with encouragement throughout this experience.

I wish to extend my gratitude to my chairperson Dr. Virginia Foley. Her willingness to help, dedication, and caring spirit helped me throughout this endeavor.

My sincere appreciation goes to my outstanding committee: Dr. Elizabeth Ralston, Dr. Eric Glover, and Dr. James Lampley. Each of them provided me with valuable input and support that made this experience possible.

Finally, I would like to thank my colleague at Anderson Elementary, Principal Andrew Brown, for his flexibility while I finished my coursework and dissertation.

CONTENTS

Page

ABSTRACT 2

DEDICATION 4

ACKNOWLEDGEMENTS 5

LIST OF FIGURES 9

Chapter

1. INTRODUCTION 11

Statement of the Problem 13

Research Questions 14

Significance of the Study 15

Definitions of Terms 16

Delimitations and Limitations 20

Overview of the Study 20

2. REVIEW OF LITERATURE 21

No Child Left Behind Act 21

Computer-Based Testing vs. Paper-and-Pencil Tests 23

Background of ThinkLink Learning's Predictive Assessment Series (PAS) 29

Background of Tennessee Comprehensive Assessment Program (TCAP) 32

Background of the Tennessee Value-Added Assessment System (TVAAS) 34

Accountability 36

Benchmark Assessments 39

3. METHODOLOGY 43

Research Design 43

Population 44

Instrumentation 45
Tennessee Comprehensive Assessment Program (TCAP) 46
ThinkLink Learning's Predictive Assessment Series (PAS) 46
Data Collection 47
Data Analysis 48
Research Questions and Hypotheses 48
Summary 54
4. DATA ANALYSIS 55
Introduction 55
Analysis of Research Questions 56
Research Question #1 56
Research Question #2 58
Fourth Grade Reading 58
Sixth Grade Reading 60
Eighth Grade Reading 62
Fourth Grade Math 64
Sixth Grade Math 66
Eighth Grade Math 67
Fourth Grade Science 69
Sixth Grade Science 70
Eighth Grade Science 72
Research Question #3 73
Fourth Grade Reading 74
Sixth Grade Reading 75
Fourth Grade Math 77
Sixth Grade Math 79
Fourth Grade Science 81

Sixth Grade Science 83
5. SUMMARY OF FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS 86
Introduction 86
Summary of Findings 86
Demographics for Grade 4 87
Demographics for Grade 6 87
Demographics for Grade 8 87
Research Question #1 88
Research Question #2 89
Research Question #3 90
Conclusions 91
Conclusion #1 91
Conclusion #2 92
Conclusion #3 92
Recommendations for Practice 92
Recommendations for Further Research 93
REFERENCES 95
APPENDIX: Director's Letter 103
VITA 104

LIST OF FIGURES

Figure Page

1. Scatter Plot of the Regression Lines of Fourth Grade PAS and TCAP Reading Scores for Males and Females 60
2. Scatter Plot of the Regression Lines of Sixth Grade PAS and TCAP Reading Scores for Males and Females 62
3. Scatter Plot of the Regression Lines of Eighth Grade PAS and TCAP Reading Scores for Males and Females 64
4. Scatter Plot of the Regression Lines of Fourth Grade PAS and TCAP Math Scores for Males and Females 65
5. Scatter Plot of the Regression Lines of Sixth Grade PAS and TCAP Math Scores for Males and Females 67
6. Scatter Plot of the Regression Lines of Eighth Grade PAS and TCAP Math Scores for Males and Females 68
7. Scatter Plot of the Regression Lines of Fourth Grade PAS and TCAP Science Scores for Males and Females 70
8. Scatter Plot of the Regression Lines of Sixth Grade PAS and TCAP Science Scores for Males and Females 71
9. Scatter Plot of the Regression Lines of Eighth Grade PAS and TCAP Science Scores for Males and Females 73
10. Scatter Plot of the Regression Lines of Fourth Grade PAS and TCAP Reading Scores for Title I and Non-Title I Students 75
11. Scatter Plot of the Regression Lines of Sixth Grade PAS and TCAP Reading Scores for Title I and Non-Title I Students 77
12. Scatter Plot of the Regression Lines of Fourth Grade PAS and TCAP

Math Scores for Title I and Non-Title I Students 79
13. Scatter Plot of the Regression Lines of Sixth Grade PAS and TCAP
Math Scores for Title I and Non-Title I Students 81
14. Scatter Plot of the Regression Lines of Fourth Grade PAS and TCAP
Science Scores for Title I and Non-Title I Students 83
15. Scatter Plot of the Regression Lines of Sixth Grade PAS and TCAP
Science Scores for Title I and Non-Title I Students 85

CHAPTER 1

INTRODUCTION

Regular use of benchmark assessments, particularly when aligned with state content standards, is seen as having potential to improve student performance. While annual state testing provides summative measures of achievement, the results are available only after students have moved to the next grade. In contrast, benchmarks are scored immediately, providing valuable information that can alert teachers and administrators to learning gaps before students move on. In a 2005 survey approximately 70 % of school superintendents reported their districts used benchmark assessments (Henderson, 2008).

With the signing of the *No Child Left Behind* legislation by President George W. Bush on January 8, 2002, accountability took on new meaning and has certainly thrust the topic of testing into the mainstream (U.S. Department of Education, 2004). The law stipulates that tests in reading, math, and science are to be given annually in grades 3-8 and once in high school. Because schools face serious consequences for failing to show Adequate Yearly Progress (AYP) proficiency for their students, the incentive to use benchmark assessments has escalated. These short tests offer instant feedback on how well students are achieving success. Many educators view the periodic use of benchmark assessments as a way to assess student achievement and to identify the specific needs of each student. Benchmark testing takes individual test scores and breaks them down by using the identical student categories that the NCLB act uses as well as supplying reports that indicate the progress of individual students (Henderson, Petrosino, Guckenburg, & Hamilton, 2007).

Those in favor of benchmark assessments argue that when used as directed, these tests supply the data necessary to give instructors immediate feedback for individual student's academic needs. Proponents also report that when benchmarks are aligned with state standards, they can assist teachers in determining their students' test outcomes against those standards of the district. On the other hand, critics of standardized benchmark assessments report that these tests promote "teaching to the test." Some also have concern that as demand has risen, quality has not kept up. Olson (2005a) noted vendors have produced benchmark assessments that include a large quantity of test questions but in terms of quality, much is left to be desired. Furthermore, Olson (2005a) pointed out that some critics of benchmark assessments feel that these forms of testing could be better described as being summative tests rather than formative tests. There are even those who fear the money, time, and energy expended in benchmark assessments could divert the focus from those critical elements such as reshaping how teachers interact with their students each day. Furthermore, proponents argue that commercially produced benchmark tests such as Discovery Education's ThinkLink PAS are far from ideal but better than nothing at all. Likewise, the critical point made by many is that educators need to ensure that they are making the best use of the data.

The Discovery Education ThinkLink Predictive Benchmark assessment has incorporated a unique scientific practice that matches diagnostic assessments to mirror a state's curriculum and standardized test. It relies on a research-based program that addresses and meets the requirements of Stage 5 of the NCLB research guidelines.

Discovery Education claims that its predictive assessments predict student proficiency, mastery, and AYP performance with 80% to 90% accuracy. The goal of the Discovery Education ThinkLink PAS is to provide teachers with timely and reliable data from predictive tests so that educators have the ability to target areas of concern and plan instruction throughout the school year (California Learning Resource Network, 2008).

Statement of the Problem

For many years Discovery Education's *ThinkLink* Predictive Assessment Series (*PAS)* has been the sole benchmark assessment used in grades two through eight in the school system being studied. With higher expectations of students' performance, the system's school leaders opted to administer the *ThinkLink PAS*® tests three times a year for the purpose of maximizing students' success on standardized tests. The purpose of this study is to investigate the degree of correlation between the PAS, a computerized predictive assessment marketed by ThinkLink Learning, a business unit of Discovery Education, and the Tennessee Comprehensive Assessment Program (TCAP) test in reading, math, and science. Because the TCAP is a high-stakes test that is used to measure the academic success or failure of schools, it is, therefore, imperative that educators employ any tool available to ensure that children are well equipped (Teachers' Guide, 1999). The study includes in gender and socioeconomic status as determined by enrollment in Title I schools. The researcher was unable to expand the study to include the impact of free and reduced lunch status because it is federally protected. The PAS tests are administered in the fall, winter, and spring. These tests are created so that they mirror and match the state test. ThinkLink claims that the PAS test is highly accurate at predicting student proficiency, mastery, and AYP performance. ThinkLink PAS cites their research as showing that the PAS test predicts student performance with 80% to 90% accuracy.

Benchmark tests have been available for about 10 years. They are designed to evaluate the level of student mastery of skills so that educators can monitor student progress toward state mandated goals. Within the school system chosen for this study, teachers rely heavily on the PAS test results to guide their planning of instruction. This study was designed to substantiate the accuracy cited by ThinkLink PAS and to determine whether or not the PAS test is helping educators maximize student success on the TCAP tests in the spring each year (ATP, 2002).

The information obtained from this study will be interesting for both teachers and administrators. It could reveal new knowledge to the field of K-12 student assessments and assist teachers and administrators to make educated decisions when it is time to make the next system-wide predictive benchmark assessment adoption. Likewise, this study might be useful for other school systems contemplating the best predictive benchmark assessment tool for their students.

Research Questions

The following research questions guided this study:

1. Are there relationships between the scaled scores of the PAS and the scaled scores of the TCAP in reading, math, and science for students in Grade 4, Grade 6, and Grade 8?
2. Are the relationships between the PAS and TCAP tests in reading, math, and science the same for both male and female students?

3. Are the relationships between the PAS tests and TCAP tests in reading, math, and science the same for students who are attending Title I and Non-Title I schools?

Significance of the Study

The fact that many school systems use benchmark tests such as ThinkLink PAS underscores the need to gather and analyze the available data about the assessments used as a predictor indicator. This research is valued because it looks at an assessment tool using technology that provides timely and accurate information. The data obtained can then be used to gauge student progress and more importantly predict student achievement on high stakes tests.

The National Center for Educational Accountability and others have determined that one common characteristic of high-achieving districts is the use of periodic benchmark assessments (Olson, 2005). A 2005 survey of superintendents indicated that an estimated 70% of school districts used some form of benchmark testing and, as many as 80% projected their use for the upcoming school year (Olsen, 2005a). School systems across the country continue to move toward the use of benchmark testing that provide more readily useable student achievement data at regular intervals. Computer-based benchmark assessment tools can provide the timeliness needed to meet the demands of today's schools. These data provide teachers the opportunity to adjust their instruction accordingly.

Administrators and classroom teachers need to know the potential for increasing proficiency levels that benchmark testing may hold. By examining these issues, this study might help the school district to redirect resources in a manner that would most likely have the biggest payoff in proficiency gains. Additionally, an examination of this study should add to the discussion surrounding benchmark testing programs. The researcher hopes that such discussion leads to both answers and questions for further research. The information obtained through this study could also help other districts as they struggle with the most effective way to realize the best use of benchmark testing such as Discovery Education's ThinkLink PAS.

Definition of Terms

1. Benchmark Assessment: A benchmark assessment is a formative assessment, usually with two or more equivalent forms so that the assessment can be administered to the same children at multiple times over a school year without evidence of practice effects. In addition to formative functions, benchmark assessments allow educators to monitor the progress of students against state standards and to predict performance on state exams (Brown & Coughlin, 2007).
2. Correlation: The nature, or extent, of the relationship between two variables (Hinkle et al., p. 617).
3. Criteria-referenced Test: A measurement that focuses on performance of an individual as measured against a standard or a set of prespecified criteria rather than against performance of others who take the same test (Harvey, 2004-2011).
4. Criterion Validity: The ability of a measure to predict performance on a second measure of the same construct computed as a correlation. If the second measure is taken after the first, the ability is described as predictive validity (Brown & Coughlin, 2007).

5. Formative Assessment: An assessment designed to provide information to guide instruction (Brown & Coughlin, 2007).
6. Non-Title I Schools: Schools that do not qualify for federal funds (U.S. Department of Education, 2002, p. 13).
7. Norm-referenced Test: A measurement of achievement that is standardized on a group of test takers whose performance is evaluated in relation to the performance of others. It gives a comparison of student performance in five content areas against a national norm group of students taking a similar test. The expectation is that the average score for a school or school system will be at the national average (Tennessee Report Card, p. 1).
8. Pearson Product-Moment Correlation Coefficient: The index of the linear relationship between two variables, called the Pearson *r* (Hinkle et al., p. 620).
9. Preassessment Tool: Can help determine what needs to be reviewed, emphasized, or introduced for the first time. These tools may include oral or written feedback, formal or informal methods, a broad or narrow focus (Teaching Today, 2009).
10. Predictive Assessment System (PAS): A predictive assessment designed to assess student progress to meeting state standards. This assessment is used by many schools as a preassessment tool (ThinkLink Learning, 2005).
11. Predictive Validity: The ability of one assessment tool to predict future performance either in some activity or on another assessment of the same construct (Bredekamp & Shepard, 1989, Kurdek & Sinclair, 2001).

12. *R*. : *R*. relates to multiple correlations and is the square root of R-squared (Salkind, 2005).
13. Reliability: The degree to which test scores for a group of test takers are consistent over repeated applications of a measurement procedure and hence are inferred to be dependable and repeatable for an individual test taker. Low reliability means that scores should not be trusted for decision-making (Herman, Osmundson, & Dietel, 2010).
14. Scatter Plot: A scatter plot is a plot of each set of scores on separate axes. A positive trend is shown if as one set of values increases the other set tends to increase. A negative trend is indicated if as one set of values increases the other set tends to decrease. The general shape of the collection of data points indicates whether the correlation is direct (positive) or indirect (negative) (Salkind, 2005).
15. Standardized Test: A measurement that is given to a specific population and then the means, standard deviations, standardized scores, and percentiles are calculated. The scores are then compared by taking an individual score and comparing it with the established norm group score (Gay et al., 2006).
16. State Content Standards: The knowledge and skills that all students should know and be able to do for each grade level and academic subject area. This includes the minimum standards for school systems to follow and to communicate to the public (Brown & Coughlin, 2007).
17. Statistical Package for the Social Sciences (SPSS): One of many Windows-based statistical software packages used to analyze a large data set is call Statistical Package for the Social Science (Salkind, 2005).

18. Tennessee Comprehensive Assessment Program (TCAP): The Tennessee Comprehensive Assessment Program (TCAP) assesses content areas in reading, mathematics, science, and social studies. The TCAP is a criterion-referenced test based on the Tennessee standards. In the state of Tennessee students in grades 3-8 are administered the Tennessee Comprehensive Assessment Program Achievement Test each spring. This is a timed, multiple choice assessment that measures skills in reading, language arts, mathematics, science, and social studies. The results are provided to parents, teachers, and administrators (Tennessee Report Card, p. 1).
19. Tennessee Value-Added Assessment System (TVAAS): A tool that gives feedback to school leaders and teachers on student progress. It allows districts to follow student achievement over time and provides schools with a longitudinal view of student performance. TVAAS provides valuable information for teachers to make informed instructional decisions (Tennessee Report Card, p. 1).
20. Title I Schools: Refers to schools that receive funds under Title I of the Elementary and Secondary Education Act (ESEA). Title I supports programs to improve the academic performance of students from low-income families. This category is the method used to analyze economically disadvantaged (United States Department of Education, 2003).
21. Validity: The extent to which an assessment measures what it is supposed to measure and the extent to which inferences and actions made on the basis of test scores are appropriate and accurate determines test validity (Messick, 1980).

Delimitations and Limitations

Delimitations of this study included:

1. The population of the study was delimited to students in fourth, sixth, and eighth grade having taken the ThinkLink PAS computerized, benchmark test in February and the TCAP test in April during the 2008-2009 school year.
2. The population was delimited to a school system that used a benchmark testing program for the past 7 years.
3. This study was delimited to students enrolled in fourth, sixth, and eighth grade in six public schools in a northeastern Tennessee school system during the 2008-2009 school year.

Limitations of this study included:

1. This study was limited to those fourth, sixth, and eighth grade students who were administered both the ThinkLink PAS test and the TCAP test during the 2008-2009 school year.

The main limitation of this study is one of limited generalizability.

Overview of the Study

This study was arranged into five chapters. Chapter 1 contains an introduction to the study, statement of the problem, applicable research questions, significance of the study, definitions of terms, and delimitations and limitations. Chapter 2 provides a review of literature related to the issues addressed in the study. Chapter 3 includes research methodology and design. Chapter 4 presents the results of the study. Lastly, the summary, conclusions, and recommendations are the focus of Chapter 5.

CHAPTER 2

REVIEW OF THE LITERATURE

No Child Left Behind Act

The NCLB Act of 2001 brought mixed reactions, positive and negative, from a large number of stakeholders. The primary focus of the law is to guarantee that all students – regardless of economic status, race, ethnicity, language spoken at home, or disability – be able to obtain proficiency in reading, math, and science by 2014 (Center for Public Education, 2006). NCLB passed with bipartisan support by Congress in 2001 and was signed into law by President George W. Bush in 2002. The new law was a representation of the education reform plan of the President and contained the most changes to the Elementary and Secondary Education Act since it was enacted in 1965 (Tennessee Department of Education, 2003).

The accountability requirements of the NCLB Act put responsibility for student achievement on schools (Casbarro, 2004). In order to fulfill a part of the requirements schools in each state must assess students annually in reading and math in grades three through eight and again before they graduate from high school (Neil, 2003).

> This requirement must be met by the 2005-2006 school year and science assessments in key grades will follow in the 2007-2008 school year. Due to these requirements, at least 36 states will have to develop more than 200 new tests within the next few years to be in compliance with the federal law. (Gandal & McGiffert, 2003, p.39)

The focus under NCLB is to close the achievement gap, especially in reading, math, and science. The achievement gap is a demonstration of the difference between how well economically disadvantaged and minority students perform on standardized tests compared to their peers (Tennessee Department of Education, 2005, p. 15).

With the reauthorization of ESEA in 1994, states were required to plan and adopt standards. Under NCLB students are required to be tested more often, and the tests developed are based on rigorous state standards that define specifically what students should know and be able to do at a certain age and grade level (Resnick, 2003). States and districts must in addition demonstrate progress in closing the achievement gap between traditionally low-performing groups of students and their peers. According to NCLB students should be performing at the proficient level on state achievement tests by the 2012 school year. In order to meet the criteria established states need to shift an additional 4% to 6% of their students into the proficient category every year. Using the National Assessment of Educational Progress (NAEP) tests as a measure, only 3 of 33 states made even 1% gains in reading per year from 1992 to 1998 (Neil, 2003). Schools that have not met their annual AYP target must expend time and resources for additional services to help bring up student performance. By 2014 NCLB mandates that all students will show 100% proficiency in reading, math, and science. Schools could face sanctions being applied if they fail to meet the standards (Center for Public Education, 2006). Schools that fail to achieve AYP goals face daunting corrective actions such as replacement of school faculty, implementing a new curriculum, extension of the school day or year, parental choice options, and complete reorganization (Guilfoyle, 2006).

Teacher qualification was addressed for the first time in NCLB act of 2001. States must ensure that all students are being taught by highly qualified teachers (Feller, 2006). In 2006, 33 states reported that at least 90% of their students were being taught by highly qualified teachers (Henderson, 2008). This requirement has put more impetus on states and districts to focus on teacher recruitment and retention.

In recent decades there has been a rising interest in standardized testing and the use of the scores from testing to determine the accountability of schools (United States Department of Education, 2005). As reported by the American Educational Research Association, spending on K-12 tests for the 50 states has almost doubled from $165 million in 1996 to $330 million in 2000 (McAdams, 2002). Test scores have become the main source of data examined when determining the effectiveness of a school for its students (Amrein & Berliner, 2002).

Computer-Based Testing vs. Paper-and-Pencil Tests

The passage of the No Child Left Behind (NCLB) Act has impacted greatly the direction testing has taken. With higher expectations being put on student performance, it has led to education stakeholders pursuing a more effective means of measuring student knowledge than the use of traditional paper-and-pencil tests (Wang, 2008). Many administrators foresee the use of computer-based testing on state assessments because of the advancement of technology. One of the key advantages of computer-based assessment over paper-and-pencil testing is that the computer-based allows instructors and students immediate feedback. Computer-based testing also increases test security, decreases the costs for mailing tests back to the state testing facility, and gives administrators flexibility when scheduling test times. In addition, computer-based testing offers the use of multimedia innovative item responses that are not available with the paper-and-pencil tests (Bennett, 2001, 2002; National Association of State Boards of Education, 2001; National Center for Education Statistics, 2000; National Commission on Excellence in Education, 1983). Using computers for test administration is justified because of the increased usage of computers in schools. Educators have found that computers have become an essential tool to enhance their instruction and assessment. More importantly, computers have received positive acceptance from students and teachers (Wang, 2008). In the future plans have already been started to implement computer-based assessments throughout the educational systems of our country (Bennett, 2001, 2002).

Studies conducted by the Princeton, New Jersey, based Educational Testing Service indicated that students' performance on computer-delivered tests is dependent, in part on their competency with technology. According to Olson (2005b) the studies focused on the results of students who had responded to mathematics and writing items on a test from the National Assessment of Educational Progress using paper-and-pencil vs. computer. The results showed that 8^{th} grade average scores for students using the computerized test were about four points less than those of students who had used the paper-and-pencil version. Also, 5% more fourth grade students answered correctly to test items on paper than on a computer. The statistics from the National Assessment of Educational Progress illustrated that it was essential for tested grade levels to have computer instruction to increase input speed and accuracy. When comparing the computer-based writing test and the paper and pencil test, the results did not show a significant difference. Again, the students with better computer skills were successful getting higher scores (Olson, 2005b).

Clariana and Wallace (2002) noted that there is increased evidence to verify that identical paper and computer-based tests will not offer the same results. This occurrence has been identified as the *test mode effect*. Bunderson (1989) conducted 3 studies that indicated high performances for computer-based tests, 11 studies indicated no significant differences, and 9 studies indicated a superior rating for paper-based tests. According to these findings the possibility of a particular test giving the same results on paper and computer are just about 50%.

When examining the test mode effect, the need for paper and computer forms of the test to be the same are necessary. Mourant, Lakshmanan, and Chantadisai (1981) have reported that students become more tired when reading words on a computer screen rather than reading the same words on paper. Wilson (2001) has shown that fonts have also been responsible for computer versus paper differences. Perhaps the two greatest differences between the two methods of testing are perceived interactivity and physical size of the computer display. A computer screen can only display about one third of the information printed on a standard sheet of paper. Haas and Hayes (1986) noted that when a test question required more than one page, computer scores showed to be lower than paper-and-pencil ones. This could be attributed to the difficulty of reading the text on the computer screen (Bugbee & Bernt, 1990). On the paper-based test numerous test items were arranged on a page. Students could easily turn pages backward or forward to view other questions. This example of interactivity proved to yield higher scores for paper-and-pencil administration.

When using computer testing scoring becomes an instant, easy task. The chance of making errors while checking tests is diminished (Bahr & Bahr, 1997). In 1990 Bugbee and Bernt studied 265,000 tests that had been taken by using computers and stated that students were more in favor of taking a computerized test due to the immediate scoring of the test. Wise and Plake (1990) noted a saving on resources such as paper and personnel when using computers. The time crunch for testing is aided when using computer-based testing because they can be taken anywhere or at anytime the proper hardware and software are available.

Because computer technology has grown tremendously, computer-based testing may soon incorporate audio, video, and animation. Parshall (1999) explained that audio tests may greatly change the way measurement in certain areas is done. Zenisky and Sireci (2002) predicted that many new innovations in computer-based testing will alter the test taking experience for many examinees.

As testing becomes more computer-based, test takers who can type computer keys fairly fast will be at an advantage over those who cannot. Furthermore, the gap in performance on multiple choice tests when comparing men and women, ethnic groups, or people of varied socioeconomic backgrounds could become greater as a result of the computerized testing (National Center for Fair and Open Testing, 1998).

Due to the growth of computer-based testing, studies were launched to examine the availability of the Internet to students in schools (Davis, 1998). In 1999 the National Center for Education Statistics (NCES) reported that 63% of all instructional classrooms had Internet access. This was 20 times more than 5 years earlier. NCES reported in 1999 that 95% of all schools had Internet connection (NCES, 2000). These figures suggest that schools were becoming connected to the Internet at a fast rate. Some students were not receiving computer access at the speed of others. In high-poverty schools the number of students to Internet computers was 16:1. Low-poverty schools were 7:1 (NCES, 2000). According to Clariana and Wallace (2002) computer familiarity is the most important factor to consider in the test mode effect. Their concerns are mainly for students identified as having reduced computer access such as females and minorities. In comparison, higher-attaining students will excel with any new assessment tool and will quickly adjust to test taking strategies (Watson, 2001). In the investigation conducted by Clariana and Wallace (2002) higher-attaining students likely made accommodations rapidly and, therefore, were more successful with computer-based assessment. As familiarity with computers rises, then computer familiarity should not be a hindrance to some.

Some researchers are asking which assessment mode more accurately shows the students' actual knowledge. According to Bugbee (1996) test developers should show that computer-based and paper-based test versions are equivalent, and/or must give information to identify the scaling process used to equate the two tests. Clariana and Wallace (2002) stated that additional time and effort must be used in order to improve test items. Their findings indicate that even by using identical items on computer-based and paper-based tests it is not necessarily going to provide equivalent measures of student learning (Bugbee, 1996).

A study by Ward, Hooper, and Hannafin (1989) indicated no difference in test performance between paper-and-pencil and computer-based testing, but it revealed a considerable difference in anxiety level. Those being tested by using a computer showed a higher anxiety level. Shermis and Lombard (1998) in a study on test anxiety found there was no statistically significant relationship between self-reported measures of computer anxiety and test anxiety. Consequently, the measures were then combined in a prediction model. The results showed that anxiety and age became significant predictors. Low anxiety and older age were associated with high math scores. A similar study explored the effect of prior computer use to a students' willingness to test by using a computer (Bugbee & Bernt, 1990). As the researchers expected, the study revealed a less negative response for the use of computers that was significantly linked to more regular use of computers. Results also showed that more computer experience did not necessarily increase the election by the student to test by computer when a choice of both types of testing was available. Moreover, the study found that a person's feelings about using a computer were also linked to the type of task being asked to accomplish.

There is a large body of research that explores the comparability of scores from paper-and-pencil tests and computer-based tests. According to Bunderson, Inouye, and Olsen (1989) and Wise, Barnes, Harvey, and Plake (1989) computer-based and paper-based test version results are very similar. From the test taker's viewpoint computer assessment was easier (Park, 2003). More recent research showed that some students, when doing the writing version, felt more confident and comfortable by using the computer (Russell, 1999; Russell & Haney, 1997; Russell & Plati, 2001a, 2001b). The results suggested that computer-based testing may be the better choice over paper-and-pencil testing to measure students' writing abilities.

Even though there is a growing amount of interest among testing companies to prepare online testing for state assessments, Trotter (2002) states that state education officials are for the most part too conservative to branch out and purchase a new method of assessment. Therefore, Trotter reported that this hesitancy might have an effect on the growth of the market. Bennett (1999) pointed out that radical improvements in assessment will springboard from the three areas of technology, measurement, and cognitive science. Of the three areas new technology will prove to be the most powerful force for change. Although it is difficult to foresee the long-term direction that large-scale assessment will take, it is a certainty that technological improvements will enhance the practice of educational assessment (Bennett, 1999).

Background of ThinkLink Learning's Predictive Assessment Series (PAS)

Discovery Education and ThinkLink Learning's PAS is a preassessment tool designed to measure the knowledge and skills tested by state standardized tests. For the 2004-2005 school year ThinkLink Learning offered its formative assessment program to almost 1,000 schools and approximately 300,000 students (ThinkLink Learning, 2005). More recently ThinkLink reported that it administered 3 million assessments to students during the 2006-2007 school year (ThinkLink Learning, 2005). ThinkLink claimed that the PAS predicts the proficiency of students, mastery of subject matter, and AYP performance with 80% - 90% accuracy (CLRN, 2008). Each state determines what will be the requirements needed to reach proficiency on its achievement tests. In Tennessee the proficiency levels are listed as: not proficient, proficient, and advanced (ThinkLink Learning, 2005). ThinkLink provides three tests to be given during the academic year. Test 1 is taken at the beginning of the school year to measure content from the previous year. The first test is used as a preassessment tool. Teachers use the tests results to determine which skills the students are weak in from the previous grade levels. By using the data given, teachers and administrators plan strategies and select resources to reach all students (CLRN, 2008). Test 2 in early winter is on content for the current school year's summative test. Then, Test 3 is given in the spring just before the state testing dates. Test 3 is for predictions about whether students are likely to reach proficiency on the state tests. Tests 2 and 3 results can be looked at to view growth of individual students as the year has progressed. ThinkLink Learning (Sausner, 2005) describes its periodic predictive testing to the painting technique pointillism. Hardin Daniel, Vice president of sales and marketing of ThinkLink Learning (Sausner, 2005), said, "If you get real close to the painting you can see the individual brushstrokes. Every once in a while the teacher needs to back up and get that overall view of, 'How are we doing according to what the state is measuring?"

ThinkLink Learning's Predictive Assessment Series assesses student progress toward meeting state standards for reading/language arts, math, and science. The tests are re-evaluated and studied by ThinkLink experts to maintain a high correlation with state standards. As established by the No Child Left Behind (NCLB) Act states are required to provide their own annual tests for grades three through eight to measure students' learning as required by the standards (Fleishman & Safer, 2005). This has led to a heightened emphasis on the use of data. The accountability requirements of the federal (NCLB) legislation helped increase ThinkLink's usage (Wayman, 2004). As Earl and Katz (2002) noted data use is now not a choice for school leaders but a must. In terms of improving student performance on the end of the year state assessments, ThinkLink claimed "there were 399 Tennessee schools, representing over 100,000 students in grades 3 to 8 who used ThinkLink's Predictive Assessment Series. Of the 205 schools that met AYP in 2002-2003, ThinkLink helped 200 or 98% maintain AYP in 2003-2004. Of the 194 schools that did not meet AYP in 2002-2003, ThinkLink helped 137 of 194 schools or 71% improve and meet AYP benchmarks in 2003-2004" (ThinkLink Learning, 2005).

Research has indicated that practice tests do not improve student learning and fail to cause test scores to escalate (Daniel & Wheeler, 2006). It is the data gathered from formative assessments that result in higher test scores (ThinkLink, 2005). Generally, practice tests do not provide the instructional feedback that can be gained by the use of formative assessments. To assist in making the data understandable for teachers, each ThinkLink report is color-coded and, thereby, very easy to comprehend. Reports can be generated that identify mastery of each student for a specific objective (Daniel & Wheeler, 2006). In addition, a growth score is provided to aid in watching student Adequate Yearly Progress (AYP). The teacher receives immediate feedback on what students have mastered, what they partially know, and what they have not mastered. These reports are available online to teachers and administrators by using a password protected account (ThinkLink Learning, 2005). Administrators also have access to reports that give a snapshot of data that shows the percent of students achieving mastery by grade and subject. There is in addition summary data that compares scores by grade and by school across the district (ThinkLink Learning, 2005).

Herman and Baker (2005) noted, "A test has diagnostic value to the extent that it provides useful feedback for instructional planning for individuals and groups. A test with high diagnostic value will tell us not only whether students are performing well, but also why students are performing at certain levels and what to do about it." ThinkLink benchmark tests give complete feedback on the performance of students in a format that is user-friendly. The benchmark tests are available for teacher and student use after testing. The test questions can be read and discussed by teachers and students. The diagnostic value is greater as students and teachers are able to talk about correct and incorrect responses (CLRN, 2008). Timely reports about potential learning problems permit the school to implement corrective measures sooner rather than later.

A fair benchmark test should also give an accurate assessment of diverse subgroups. To eliminate bias ThinkLink test items are reviewed for fairness regarding gender, race, and other categories (Daniel & Wheeler, 2006). Accommodations are also provided for students needing Braille tests, large print, or audio recordings.

Background of Tennessee Comprehensive Assessment Program (TCAP)

In the state of Tennessee students in grades three to eight take achievement tests as part of the Tennessee Comprehensive Assessment Program (TCAP). "The TCAP Achievement Test has fresh, non-redundant test items and is customized yearly to measure academic basic skills in reading, language arts, mathematics, science, and social studies" (Tennessee Department of Education, 2004, p.1). TCAP uses the Tennessee Criterion Referenced Test (CRT) to measure student learning.

The state of Tennessee has used the Tennessee Comprehensive Assessment program to measure students' achievement since 1989. Currently the TCAP test uses pertinent information to evaluate students, teachers, and schools using the criteria established by NCLB accountability standards. The TCAP achievement test, which is a timed, multiple-choice, criterion-referenced test, measures basic skills that are directly linked to state standards (Tennessee Department of Education, 2006). In the spring Tennessee students in grades three through eight complete the TCAP Achievement test. Under the NCLB law all students in all subgroups have to be included on district and state assessment programs (Asp, 2000). Students who are English-language learners and have attended school in the United States for 3 consecutive years must complete reading assessments that are written in English. To meet the needs of English language learners and students with disabilities various accommodations have been permitted. For grades three through eight the TCAP achievement test gives criterion-referenced performance information (Paige, 2006). Test results are reported to parents, teachers, and administrators and these outcomes are reviewed by the school staff to improve the instructional needs of students in Tennessee (Tennessee Department of Education, 2006).

Every year the Tennessee Department of Education issues a report card for the state and for each public school system and school in the state. This report card uses letter grades to indicate performance on academic and nonacademic measures (Pruett, 2002). For grades three through eight, academic information is based upon cumulative 3-year averages in two areas for each of the five subject areas. The first area, academic achievement, is derived from the normal curve equivalent (NCE) average for schools and districts. The second area is drawn from the average value-added growth for each subject area for schools and districts. Value-added assessment does comparison by using students' scores on the previous years' tests to establish if they are improving academically (Hellend, 2001). A database is maintained that contains achievement test results for all students taking the test over the past 3 years (Baker, Xu, & Detch, 1995).

By following the progress of individual students the problems of socioeconomic factors becomes less of a hindrance (Sanders, 1998).

Background of Tennessee Value-Added Assessment System (TVAAS)

The Tennessee Value-Added Assessment System (TVAAS) was begun in 1992 as an important part of comprehensive education reform method that measured teaching and learning (Center for Greater Philadelphia, 2004). A former University of Tennessee professor, William Sanders, developed this statistical measurement tool. Its purpose was to hold educators and schools accountable for student learning. This measurement system has enabled researchers to make predictions using test data to determine student growth in a school year (Hershberg, 2004a). By tracking individual students over time and using value-added, the impact of teacher instruction on students' learning and growth can be measured. Using scale score data, TVAAS developed a profile of academic growth for each student (Holloway, 2000).

Value-added assessment can be used in a number of valuable ways. Administrators might find it helpful when making personnel assignments, student placement, resource allocation, and staff development training. The value-added model could help other states and districts to formulate comprehensive accountability systems that could be used to evaluate curriculum, professional development, and teaching methods to determine their effect on academic achievement (Hershberg, 2004b). Evidence has been shown from the value-added model that differences in classroom teachers' effectiveness was the main determiner in improving student academic growth (Holloway, 2000). Sanders and Rivers (1996) conducted a study using students in Tennessee. The results indicated that students having an effective teacher in math for 3 consecutive years scored 50 percentage points higher than those students who had ineffective teachers. From this study the profound effect of teachers on student achievement is exhibited (Hershberg, 2004a).

The main function of TVAAS has been to meet the accountability requirements of the Tennessee Education Improvement Act by giving information about the learning gains of students as predicted by the previous 3-year period (Tucker & Stronge, 2005). The basic information presented by the Tennessee Department of Education (2005) using TVAAS is as follows:

Student Level:

1. gains for each subject for the 3 most recent years,
2. 3-year average gains, and
3. comparison of gains to be averaged for the school, school district, state, and nation.

Teacher Level:

1. average gains of students in each subject and grade level taught by the teacher in the 3 most recent years,
2. average gains of students in the school district in each subject and grade level during the current year, and
3. comparison of average gains to those for the school district, state, and nation (TDOE, 2005).

Hershberg (2004b) reported that value-added assessment has provided two important benefits since the inception of NCLB. It has offered educators an avenue to improve their instruction as well as determining a way to measure school performance. According to Brandt (2000) the value-added approach might be the fairest method to use when comparing the effectiveness of teachers and schools on students' academic achievement.

Accountability

The enactment of the federal No Child Left Behind Act in 2002 made performance-based education accountability a federal mandate. Performance-based accountability's attraction is the promise that all students, even the disadvantaged, will master basic knowledge and skills (Tennessee Department of Education, 2005). Experts have reported that No Child Left Behind's mandates have compelled teachers to focus mainly on high-stakes testing rather than on improving learning and planning interesting educational experiences that will enable students to enter society as prepared citizens (Noddings, 2005). As Casbarro (2004) has stated accountability is increased when higher and more rigorous standards are implemented. Furthermore, greater accountability leads to more testing and, as a result of the testing, comes heightened stress and anxiety. "By raising the bar, we have created one of the most stress-filled learning environments in history" (Casbarro, 2004, p. 37).

The current importance put on testing as a tool of education reform goes back many years to a time when tests were used to change pedagogical ideas and practices. In the United States this use of testing extends back to 1845 in Boston when Horace Mann replaced a traditional oral exam with a standardized written essay test. According to history, in Italy during the 15th Century teacher salaries were linked to student exam performance (Madaus & O'Dwyer, 1999).

A considerable amount of past data and recent research verifies that as the stakes increase the curriculum becomes more limited in order to concentrate on the content being tested (Madaus & O'Dwyer, 1999, p.33). Pressure to raise test scores to comply with NCLB accountability encourages schools to increase time on tested areas and decrease time on nontested content (Neil, 2003). According to one school principal, "The art, music, and everything else are basically out the window . . . something has to go" (Herszenhorn, 2003).

A national survey found that teachers in high-stakes states were four times more likely than those in low-stakes states to spend more than 30 hours a year on test preparation, such as reviewing topics, working similar test items, and using commercial materials to enhance test performance (Pedulla et al., 2003). Teachers also consider the form (multiple choice, essay, short answers, etc.) that questions on high-stakes tests are using. Research has been conducted that indicates that test format does influence instruction both in a positive and negative way (Nichols, Glass, & Berliner, 2007). Tests in states that require written response from students to test questions show an increase in higher-order thinking skills and writing being taught (Taylor, Shepard, Kinner, & Rosenthal, 2003). Likewise, there are studies that show a decrease in the use of more time-consuming instructional strategies and expanded enrichment activities (Pedulla et al., 2003; Taylor et al., 2003). In addition, a recent study showed that the format of the state test may cause adverse use of technology for instruction (Russell & Abrams, in press).

As reported by Abrams and Madaus (2003) there is a need to improve state testing programs by using more than one measure of student achievement. The assessments would not allow students several opportunities to take the same test, but would allow other forms of measurement to be used. Most people recognize the importance of accountability, but the emphasis put on one test a year is stressful for students and teachers (Nichols, Glass, & Berliner, 2005). State testing programs might also look at other indicators when determining what students know before imposing high-stakes consequences on students and schools (Abrams & Madaus, Nov. 2003, p. 34).

There are two sides to the debate regarding high-stakes testing. Those who are in favor of standardized testing see it as the only fair method of determining how schools and students perform (Neill, 2006). Those in opposition express their dissatisfaction with using a single test to adequately assess the performance of an individual student or school (Owens, 2002). Supporters of high-stakes testing affirm that teachers need to be held accountable, and the test scores can be used to enhance educational instruction and offer better professional development for teachers (Amrein & Berliner, 2002). Those who are against high-stakes testing are quick to argue that it encourages schools to "teach to the test;" therefore, the results might show improvement, but in reality little improvement in learning has been accomplished (Green, Winters, & Forster, 2003).

The state of Tennessee has developed a Tennessee Accountability Plan to hold kindergarten through eighth grade schools accountable. Ninety-five percent of students must be tested and reach 83% proficiency in reading, language, and writing and 79% proficiency in mathematics. Schools must maintain a 93% attendance rate or show improvement. A 95% confidence interval has been applied to determine if targets are met (Winstead, 2006).

Benchmark Assessments

Benchmark assessments are being used in many school districts and systems. Throughout the country to increase achievement test scores and to meet mandates stipulated in the NCLB act of 2001, benchmark assessments are viewed as the way for schools to meet state standards (Henderson, 2008). Usually the benchmark assessments are administered 3 to 5 times a year and give teachers and administrators immediate data to measure students' progress as well as helping teachers adjust instruction (Herman & Baker, 2005; Olson, 2005a). Because school districts have been worried about student performance on end-of-the-year state tests, benchmark testing has become a high growing area in the assessment industry (Olson, 2005a). In terms of feasibility Herman and Baker stated,

> Benchmark testing should be worth the time and money that schools invest in it. Well-designed benchmark tests can contribute to as well as measure student learning. But if such tests are not well designed, they can waste students' and teachers' valuable time and energy, ultimately detracting from good teaching and meaningful learning (2005, p. 54).

Most benchmark assessments take approximately one 1 each for reading and mathematics, but may include other subjects (Pasquier & Gomz-Zwiep, 2006). Test results are broken down by the same student categories required under the federal NCLB Act such as by race, income, disability, and English proficiency (Coffey, 2009). A 2005

Education Week survey of superintendents found that approximately 70% reported using benchmark assessments in their districts (Olson, 2005b). There are only a few studies of benchmark assessments' effects on student performance on state tests. The large amount of information gathered on the effects of formative assessments indicates consistently the positive effects of formative assessment on student learning (Black & Wiliam, 1998a, 1998b). Black and Wiliam (1998a) reported that positive gains are even more pronounced for low-achieving students than the general student population (Henderson, et al., 2007). Whether these trends will be true for benchmark assessments has yet to be determined.

Critics of high-stakes, standardized benchmark assessments argue that this type of testing leads to educators "teaching to the tests" (Zehr, 2006). Some critics argue that increased furor for benchmark testing has led to a decrease in quality. An Eduventures report noted that many vendors have placed an emphasis on the quantity of test questions as opposed to the quality. Although the test companies may have tens of thousands of exam items, many of the items have not been extensively field-tested or undergone a rigorous review (Olson, 2005). In addition, critics warn that even the best benchmark tests are not true formative assessments that are meant to give immediate help to adjust teaching and learning as it is happening (Herman & Baker, 2005). With benchmark assessments, the teacher has already moved on and the results are likely used for remediation purposes (Olson, 2005).

Those who advocate the use of benchmark tests suggest that if used correctly the data from ongoing assessments can improve classroom practices that will heighten learning (Coffey, 2009). Proponents also claim that if benchmarks are in alignment with state standards, teachers are enabled to use pertinent data to make better instructional decisions (Henderson, 2008).

Because teachers are now encouraged to use benchmark data in a variety of ways to relate effectively with their students, more attention should be paid to teacher development on how to use data to improve learning. Some teachers become frustrated and show resistance when forced to analyze student data (Olson, 2005b). Teacher acceptance of data usage could be accomplished by school districts supplying assistance with the use and management of the data acquired by testing. Teachers will begin to see the value of more frequent assessments in their classrooms (Pasquier & Gomz-Zwiep, 2006). Instructors will see the benefits received from immediate feedback about the quality of their instruction (Herman & Baker, 2005)

In addition to the use of data teachers should have access to supplementary materials that will help to support identified learning gaps. This area has been addressed by several school districts in that they have established support teams made up of content and curriculum experts (Popham, 2006). They meet regularly with classroom teachers to address strengths and weaknesses in student learning and determine the next steps to be undertaken by the teacher to meet the needs of various learners (Olson, 2005a).

Assessment experts warn that benchmark testing should be worth the time and money that schools invest in them (Henderson, 2007). Well-designed benchmark tests can enhance learning as well as measure student achievement (Pasquier & Gomz-Zwiep, 2006). If they are not well-designed, they can be a deterrent to learning because the tests are not only a waste of valuable learning time but, more importantly, detract from purposeful teaching and student learning (Steinberg & Henrique, 2001). In order to determine if the benchmark tests are worthwhile, administrators ultimately need to study the results. Like state tests benchmark tests will accomplish their purpose only if we watch their consequences and continue to improve their quality. Herman and Baker (2005, p.53) conclude that: “If the benchmark tests are doing their job, there should be a strong predictive relationship between students’ performance on the benchmark tests and students’ performance on the state assessments.”

CHAPTER 3

METHODOLOGY

This chapter discusses the methodology and procedures used in this study to evaluate the relationships between students' performance in reading, math, and science on the Predictive Assessment Series (PAS) and the Tennessee Comprehensive Assessment Program (TCAP) in grade 4, grade 6, and grade 8. This chapter focuses on the research design, population and data collection, instrumentation, data analysis, hypotheses, and a summary.

Research Design

The purpose of this study was to explore the relationship of Predictive Assessment Series reading, math, and science scores and TCAP reading, math, and science scores of fourth, sixth, and eighth grade students. The goal was to identify the predictive validity of the Predictive Assessment Series benchmark instrument.

A nonexperimental, exploratory, quantitative, correlational research design was used for the study. This was determined because the independent variables were not manipulated and no treatment or intervention was provided for the study participants. Normal testing data were used for the study. The data collection tools consisted of a criterion-referenced test that is completed by fourth, sixth, and eighth-grade students using the paper and pencil format, and a computer based benchmark test, the Predictive Assessment Series marketed by ThinkLink Learning and Discovery Education.

Implementing a quantitative design, this study included the fourth, sixth, and eighth grade PAS and TCAP scores in reading, math, and science from the 2008-2009 school year, with the researcher obtaining permission from a school district in northeastern Tennessee. It should be noted that this type of study is not subject to the same types of threats to internal and external validity that are typically found in experimental studies.

Population

This study was conducted in one school system in East Tennessee. The school system has five elementary schools and one middle school. Two of the elementary schools and the middle school qualify as Title I. The district had adopted the use of computerized testing for its students in 2002. The school system administered the Predictive Assessment Series test to all students in Grade two through Grade eight, in September, November, and February of each year. This particular school system educates more than 3,800 students in five elementary schools, one middle school, and one high school. District-wide, 51% of the students are male and 49% female, with an ethnic make-up of 92% White, 4.9% African American, 1.8% Hispanic, 0.9% Asian, and 0.3% Native American/Alaskan. With respect to socioeconomic status, 47% of the district's students are economically disadvantaged as defined by participation in the free-or-reduced priced meals program.

The population for the study included fourth, sixth, and eight-grade students during the 2008-2009 school year who had taken the ThinkLink Predictive Assessment Series for reading, math, and science in February of 2009 and taken the TCAP reading, math, and science test in April of 2009. Students who were within these grade levels that did not complete both the PAS and TCAP test were eliminated from the study. Of the 902 fourth, sixth, and eighth-grade students tested 879 had taken both tests.

The criterion for inclusion was that students have to have participated both in PAS and TCAP testing during the academic year 2008-2009. Data were gathered with permission from the school system. Data collection for this dissertation did not require student participation beyond normal testing. Criteria included in the study consisted of the following:

- Male or female
- Students in fourth, sixth, and eighth grade in the Bristol Tennessee City School system
- Must have taken the ThinkLink PAS test in February of 2009 and taken the TCAP in April of 2009, irrespective of gender, ethnicity, socioeconomic status, and English-language proficiency.

Participants were excluded from the study if a student did not take either the PAS test in February of 2009 or the TCAP test in April of 2009. Testing data must be available from each of the before mentioned assessments in order for a student's results to be included in the study.

Instrumentation

I used the Tennessee Comprehensive Assessment Program (TCAP) Achievement Test to gather fourth, sixth, and eighth grade student's academic performance data. The state of Tennessee mandates that students in grades three through eight take the TCAP each spring. The Achievement Test is a timed, multiple choice assessment that measures skills in reading, language arts, mathematics, science, and social studies. The TCAP achievement tests were published by Pearson Education, Inc. The TCAP test was required to be administered in six consecutive days between March 30 and April 24, 2009. The TCAP test for fourth, sixth, and eighth graders included reading, mathematics, science, and social studies.

Tennessee Comprehensive Assessment Program (TCAP)

The Tennessee Comprehensive Assessment Program (TCAP) Achievement Test is a paper-and-pencil assessment. The test is categorized as a criterion-referenced test. This indicates that the results are reported with student performance against a standard. The TCAP test was used to obtain fourth-grade, sixth-grade, and eighth grade student's academic performance in reading, math, and science. The TCAP evaluates student mastery in reading, language arts, math, social studies, and science. For the purpose of this study the reading, math, and science results were used. The TCAP measures academic achievement and whether or not it is improving over a period of time, and it helps to determine if instructional programming is giving the results that are desired.

ThinkLink Learning's Predictive Assessment Series Test

The Predictive Assessment Series (PAS) is an online-delivered test. It is a standards-based assessment tool that is based on the Tennessee benchmarks and content standards. The goal of the PAS test is to give educators immediate diagnostic data especially about those skills where mastery has not been reached. With PAS school districts are able to have consistent, reliable feedback that allows teachers to focus on state standards throughout the school year. As a result educators can make more timely decisions about educational programs and strategies for the needs of individual students.

Data Collection

An exempt status was obtained from the East Tennessee State University Review Board. Likewise, permission to conduct the study using data from the aforementioned district was received (see Appendix). The results of the study were also shared with the said district's director of testing.

The researcher collaborated with the director of testing to retrieve standardized test reports through Pearson Access. This cross-platform system provides equal functionality and performance while generating reports for tests given in each of the six schools. The reports included gender and socioeconomic status of each student in addition to test scores. The names of students were not released by the school system to the researcher. To ensure that each child's identity was protected, each student's name was omitted from the reports by the school system's director of testing.

The fourth-grade, sixth-grade, and eighth-grade reading, math, and science TCAP were given to all students. The teachers followed Tennessee's TCAP Teacher Guide for test administration guidelines.

The computerized reading, math, and science tests were given to all fourth-grade, sixth-grade, and eighth-grade students in each participating school's computer lab. All test administrators and teachers followed Discovery Education's published guidelines for test-taking procedures. In addition, all participants were given a standard test-taking environment. Again, the test administrators and teachers were to ensure that a standard environment was provided. Classroom teachers served as proctors for each of the assessments. However, the teachers did not give any added assistance to students with any test item. The TCAP results became available 2 months after the test was administered. The results of Discovery Education's Predictive Assessment Series were available immediately online.

Data Analysis

Descriptive and inferential statistics were used in this study. Data collected for the study were entered into a data file for analysis using the Statistical Program for the Social Sciences (SPSS). The criterion for establishing the statistical significance was set at an alpha level of .01.

Research Questions and Hypotheses

The following research questions guided this study:

Research Question 1: Are there relationships between the scaled scores of the PAS and the scaled scores of the TCAP in reading, math, and science for students in grade four, grade six, and grade eight? This research question was answered using the Pearson product-moment correlation coefficients of the TCAP and PAS scaled scores for students in grades four, six, and eight for each of the subject areas. The following null hypotheses were tested:

$H1_1$: There is no relationship between the PAS test in reading and the TCAP in reading among 4^{th} graders.

$H1_2$: There is no relationship between the PAS test in reading and the TCAP in reading among $6th^{th}$ graders.

$H1_3$: There is no relationship between the PAS test in reading and the TCAP in reading among 8^{th} graders.

$H1_4$: There is no relationship between the PAS test in math and the TCAP in math among 4th graders.

$H1_5$: There is no relationship between the PAS test in math and the TCAP in math among 6thth graders.

$H1_6$: There is no relationship between the PAS test in math and the TCAP in math among 8th graders.

$H1_7$: There is no relationship between the PAS test in science and the TCAP in science among 4th graders.

$H1_8$: There is no relationship between the PAS test in science and the TCAP in science among 6thth graders.

$H1_9$: There is no relationship between the PAS test in science and the TCAP in science among 8th graders.

Research Question 2: Are the relationships between the February PAS and TCAP tests in reading, math, and science the same for male and female students? To answer this research question the GLM procedure in SPSS was used to test the homogeneity of (regression) slopes. The following tests of the homogeneity of slopes for males and females were tested:

$Ho2_1$: The slopes of the regression lines for fourth grade PAS and TCAP reading scores for males and females are homogeneous (parallel)

$Ho2_2$: There is no difference in the intercepts of the regression lines for fourth grade PAS and TCAP reading scores between males and females.

$Ho2_3$: There is no difference in the fourth grade PAS and TCAP reading correlations for male and female students.

$Ho2_4$: The slopes of the regression lines for sixth grade PAS and TCAP reading scores for males and females are homogeneous (parallel).

$Ho2_5$: There is no difference in the intercepts of the regression lines for sixth grade PAS and TCAP reading scores between males and females.

$Ho2_6$: There is no difference in the sixth grade PAS and TCAP reading correlations for male and female students.

$Ho2_7$: The slopes of the regression lines for eighth grade PAS and TCAP reading scores for males and females are homogeneous (parallel).

$Ho2_8$: There is no difference in the intercepts of the regression lines for eighth grade PAS and TCAP reading scores between males and females.

$Ho2_9$: There is no difference in the eighth grade PAS and TCAP reading correlations for male and female students.

$Ho2_{10}$: The slopes of the regression lines for fourth grade PAS and TCAP math scores for males and females are homogeneous (parallel).

$Ho2_{11}$: There is no difference in the intercepts of the regression lines for fourth grade PAS and TCAP math scores between males and females.

$Ho2_{12}$: There is no difference in the fourth grade PAS and TCAP math correlations for male and female students.

$Ho2_{13}$: The slopes of the regression lines for sixth grade PAS and TCAP math scores for males and females are homogeneous (parallel).

$Ho2_{14}$: There is no difference in the intercepts of the regression lines for sixth grade PAS and TCAP math scores between males and females.

$Ho2_{15}$: There is no difference in the sixth grade PAS and TCAP math correlations for male and female students.

$Ho2_{16}$: The slopes of the regression lines for eighth grade PAS and TCAP math scores for males and females are homogeneous (parallel).

$Ho2_{17}$: There is no difference in the intercepts of the regression lines for eighth grade PAS and TCAP math scores between males and females.

$Ho2_{18}$: There is no difference in the eighth grade PAS and TCAP math correlations for male and female students.

$Ho2_{19}$: The slopes of the regression lines for fourth grade PAS and TCAP science scores for males and females are homogeneous (parallel).

$Ho2_{20}$: There is no difference in the intercepts of the regression lines for fourth grade PAS and TCAP science scores between males and females.

$Ho2_{21}$: There is no difference in the fourth grade PAS and TCAP science correlations for male and females students.

$Ho2_{22}$: The slopes of the regression lines for sixth grade PAS and TCAP science scores for males and females are homogeneous (parallel).

$Ho2_{23}$: There is no difference in the intercepts of the regression lines for sixth grade PAS and TCAP science scores between males and females.

$Ho2_{24}$: There is no difference in the sixth grade PAS and TCAP science correlations for male and female students.

$Ho2_{25}$: The slopes of the regression lines for eighth grade PAS and TCAP science scores for males and females are homogeneous (parallel).

$Ho2_{26}$: There is no difference in the intercepts of the regression lines for eighth grade PAS and TCAP science scores between males and females.

$Ho2_{27}$: There is no difference in the eighth grade PAS and TCAP science correlations for male and female students.

Research Question 3: Are the relationships between the PAS tests and TCAP tests in reading, math, and science the same for students attending Title I and Non-Title I schools? As in Research Question 2, this question was answered using the GLM procedure in SPSS to test the homogeneity of (regression) slopes. The following null hypotheses were tested:

$Ho3_1$: The slopes of the regression lines for fourth grade PAS and TCAP reading scores for Title I and Non-Title students are homogeneous (parallel).

$Ho3_2$: There is no difference in the intercepts of the regression lines for fourth grade PAS and TCAP reading scores between Title I and Non-Title I students.

$Ho3_3$: There is no difference in the fourth grade PAS and TCAP reading correlations for Title I and Non-Title I students.

$Ho3_4$: The slopes of the regression lines for sixth grade PAS and TCAP reading scores for Title I and Non-Title I students are homogeneous (parallel).

$Ho3_5$: There is no difference in the intercepts of the regression lines for sixth grade PAS and TCAP reading scores between Title I and Non-Title I students.

$Ho3_6$: There is no difference in the sixth grade PAS and TCAP reading correlations for Title I and Non-Title I students.

$Ho3_7$: The slopes of the regression lines for fourth grade PAS and TCAP math scores for Title I and Non-Title I students are homogeneous (parallel).

$Ho3_8$: There is no difference in the intercepts of the regression lines for fourth grade PAS and TCAP Math scores between Title I and Non-Title I students.

$Ho3_9$: There is no difference in the fourth grade PAS and TCAP Math correlations for Title I and Non-Title I students.

$Ho3_{10}$: The slopes of the regression lines for sixth grade PAS and TCAP Math scores for Title I and Non-Title I students are homogeneous (parallel).

$Ho3_{11}$: There is no difference in the intercepts of the regression lines for sixth grade PAS and TCAP math scores between Title I and Non-Title I students.

$Ho3_{12}$: There is no difference in the sixth grade PAS and TCAP math correlations for Title I and Non-Title I students.

$Ho3_{13}$: The slopes of the regression lines for fourth grade PAS and TCAP Science scores for Title I and Non-Title I students are homogeneous (parallel).

$Ho3_{14}$: There is no difference in the intercepts of the regression lines for fourth grade PAS and TCAP science scores between Title I and Non-Title I students.

$Ho3_{15}$: There is no difference in the fourth grade PAS and TCAP Science correlations for Title I and Non-Title students.

$Ho3_{16}$: The slopes of the regression lines for sixth grade PAS and TCAP science scores for Title I and Non-Title I students are homogeneous (parallel).

$Ho3_{17}$: There is no difference in the intercepts of the regression lines for sixth grade PAS and TCAP science scores between Title I and Non-Title I students.

$Ho3_{18}$: There is no difference in the sixth grade PAS and TCAP science correlations for Title I and Non-Title I students.

Summary

Chapter 3 consisted of the presentation of the research design, population, instrumentation, data collection, data analysis, and research questions and null hypotheses used in this study. The study's results were derived from quantitative data obtained from the *Predictive Assessment Series* benchmark scores and TCAP scores of fourth, sixth, and eighth-grade students in an Eastern Tennessee school district. In addition, the testing instruments, *Predictive Assessment Series* and TCAP, were described and explained. Null hypothesis based on research questions were listed and statistical tests were identified for each. Chapter 4 contains the results from the analysis.

CHAPTER 4

DATA ANALYSIS

Introduction

The research questions presented in Chapter 1 and the hypotheses introduced in Chapter 3 are addressed in this chapter. The purpose of this study was to determine if a correlation existed between the Predictive Assessment Series (PAS) Test, and the Tennessee Comprehensive Assessment Program (TCAP) Achievement Test in reading, math, and science for grade four, grade six, and grade eight. Test scores of students taking the February PAS Test and the TCAP in the spring of 2009 were compared. Test scores were collected from five elementary schools and one middle school from one school system in East Tennessee. This study was guided by three research questions and the corresponding null hypotheses.

Demographic information of the population encompassed Title I and Non-Title status and gender. Data from 308 (34.1%) fourth-grade students, 286 (31.7%) sixth-grade students, and 308 (34.1%) eighth-grade students in an urban school district in northeastern Tennessee were used in this study. The study included all students in the district in grades four, six, and eight who had taken both the PAS and TCAP tests during the academic year 2008-2009. The PAS test was administered by the school district in February of 2009 and the TCAP was given in April of 2009. The PAS is given three times a year and this study looks at the third test. Because of absences and students transferring to other schools, some students did not have both a PAS score and a TCAP score. The population consisted of 449 (49.8%) males and 453 (50.2%) females. In fourth grade, 108 (35.1%) students attended schools that qualify for Title I funds and 200 (64.9%) students attended Non-Title I school. In sixth grade, 114 (39.9%) students attended schools that qualify for Title I funds and 172 (60.1%) attended Non-Title I schools. Altogether, the study included 372 (62.6%) students from a Non-Title I school and 222 (37.4%) attending a Title I school. All of the eighth graders in the school district attend the same middle school, therefore, Title I and Non-Title I data could not be included in the study for the 8^{th} grade students.

Analysis of Research Questions

Data for this study were compiled from the results of the 2009 PAS and TCAP tests. Various statistical methods were used to analyze the data. The organization of this chapter follows the order of the research questions as listed in Chapters 1 and 3.

Research Question #1

Are there relationships between the scaled scores of the PAS and the scaled scores of the TCAP in reading, math, and science for students in grade four, grade six, and grade 8?

Table 1 shows the results for the correlations for the ThinkLink PAS and TCAP scores for reading, math, and science in grade four, grade six, and grade eight.

Table 1

Correlations for PAS and TCAP Scores for Reading, Math, and Science by Grade Level

	N	*R*	R^2	*P*
PAS and TCAP Reading				
4th grade	293	.698	.487	< .001
6th grade	273	.735	.540	< .001
8th grade	281	.783	.613	< .001
PAS and TCAP Math				
4th grade	291	.708	.501	< .001
6th grade	267	.762	.581	< .001
8th grade	284	.801	.642	< .001
PAS and TCAP Science				
4th grade	285	.726	.527	< .001
6th grade	270	.710	.504	< .001
8th grade	285	.737	.543	< .001

All correlations were significant at the .001 level and all the null hypotheses were rejected. All nine correlations showed a strong positive relationship between the PAS and TCAP tests. The relationships ranged from a low of .698 in fourth grade reading and a high of .801 in eighth grade math. Overall, for the population of this study the strongest correlations were found in 8th grade (r = .801 for math, r = .783 for reading, and r = .737 for science) and in the area of math (r =.801 for eighth grade, r =.762 for sixth grade, and r = .708 for fourth grade).

A Pearson product-moment correlation was used to calculate the r value between the TCAP observed scores and the PAS scores. The results of the correlation indicated a strong positive correlation between the TCAP scores and the PAS scores. Therefore, the PAS scores were useful in predicting the TCAP scores in reading, math, and science during the 2008-2009 school year for this population.

Research Question # 2

Are the relationships between PAS and TCAP tests in reading, math, and science the same for both male and female students?

The population consisted of 449 (49.8%) males and 453 (50.2%) females. The male and female students were tested in the same testing environments and at the same time of day.

Fourth Grade Reading

To evaluate the differences, if any, in the relationships between fourth grade PAS and TCAP reading scores for males and females, three hypotheses were tested:

$Ho2_1$: The slopes of the regression lines for fourth grade PAS and TCAP reading scores for males and females are homogeneous (parallel).

$Ho2_2$: There is no difference in the intercepts of the regression lines for fourth grade PAS and TCAP reading scores between males and females.

$Ho2_3$: There is no difference in the fourth grade PAS and TCAP reading correlations for male and female students.

The test of the homogeneity of slopes for fourth grade TCAP reading scores regression on fourth grade PAS scores showed the slopes regression lines for males and females were the same (parallel), F (1, 289) = 1.536, p = .216. Also, there was no difference between the intercepts of the regression lines for fourth grade males and females, F (1, 289) = 1.313, p = .253. While the correlation between fourth grade PAS and TCAP reading scores for females (r = .736) was stronger than the correlation for males (r = .652), there was no significant difference between the two correlations, Fisher's z = -1.38, p = .150 (.168). All three null hypotheses were retained. Figure 1 shows the two regression lines for fourth grade males and females are very similar.

To demonstrate the strength and direction of the relationships between TCAP and PAS assessments, a scatter plot of each correlation was created. Figures 1 through 15 graphically display the relationships between the assessments and their corresponding coefficient.

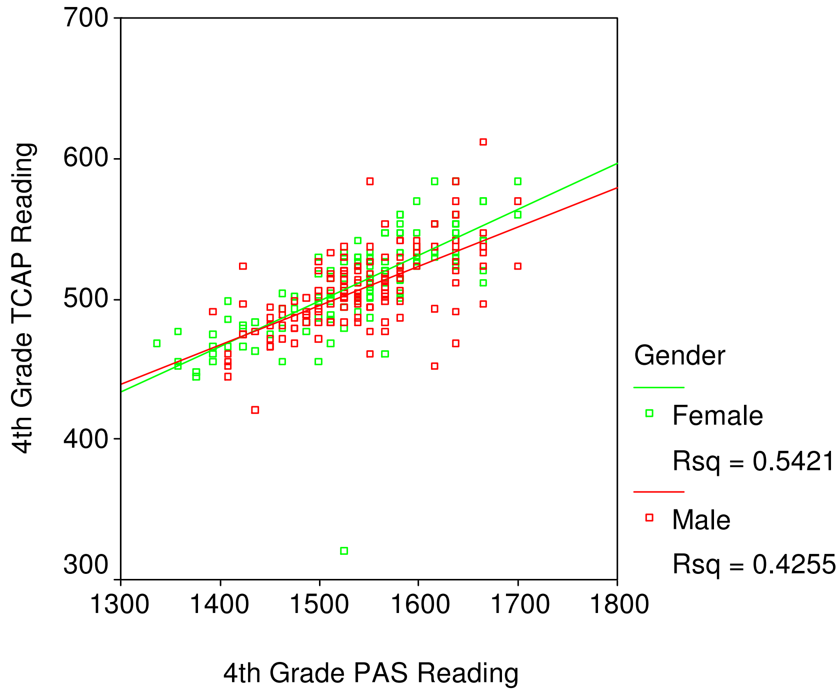

Figure 1. Scatter Plot of the Regression Lines of Fourth Grade PAS and TCAP Reading Scores for Males and Females

Notes: $\hat{y}_{males} = 75.298 + .280x$; $\hat{y}_{females} = 9.854 + .326x$

Sixth Grade Reading

To evaluate the differences, if any, in the relationships between sixth grade PAS and TCAP reading scores for males and females, three hypotheses were tested:

$Ho2_4$: The slopes of the regression lines for sixth grade PAS and TCAP reading scores for males and females are homogeneous (parallel).

$Ho2_5$: There is no difference in the intercepts of the regression lines for sixth grade PAS and TCAP reading scores between males and females.

$Ho2_6$: There is no difference in the sixth grade PAS and TCAP reading correlations for male and female students.

The test of the homogeneity of slopes was used to analyze whether or not the slopes of the regression lines for sixth grade PAS and TCAP reading scores for males and females were parallel. Figure 2 shows the two regression lines for sixth grade males and females are very similar. The summary of the findings for sixth grade TCAP reading scores regression on sixth grade PAS scores showed the slopes regression lines for males and females were the same (parallel), $F(1, 269) = .008$, $p = .929$. In addition, the test for the difference in intercepts of the regression lines for sixth grade males and females was not significant, $F(1, 269) < .001$, $p = .998$. Although the correlation between sixth grade PAS and TCAP reading scores for males ($r = .752$) was stronger than the correlation for females ($r = .723$), there was no significant difference between the two correlations, Fisher's z = .52, $p = .603$. All three null hypotheses were retained.

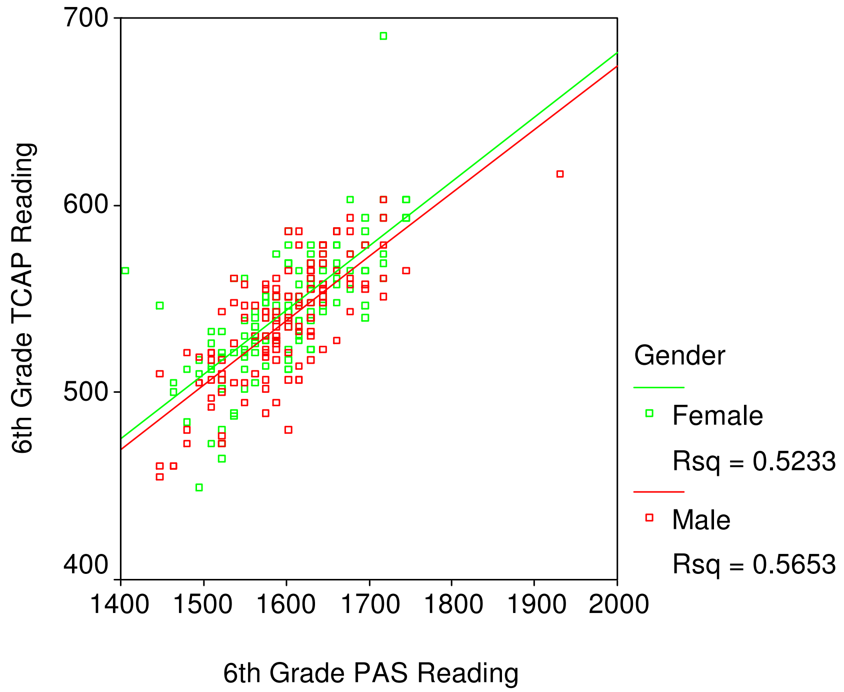

Figure 2. Scatter Plot of the Regression Lines of Sixth Grade PAS and TCAP Reading Scores for Males and Females

Notes: $\hat{y}_{\text{males}} = -7.240 + .341x$; $\hat{y}_{\text{females}} = -7.106 + .344x$

Eighth Grade Reading

To evaluate the differences, if any, in the relationships between eighth grade PAS and TCAP reading scores for males and females, three hypotheses were tested:

$Ho2_7$: The slopes of the regression lines for eighth grade PAS and TCAP reading scores for males and females are homogeneous (parallel).

$Ho2_8$: There is no difference in the intercepts of the regression lines for eighth grade PAS and TCAP reading scores between males and females.

$Ho2_9$: There is no difference in the eighth grade PAS and TCAP reading correlations for male and female students.

The test of the homogeneity of slopes for eighth grade TCAP reading scores regressed on eighth grade PAS scores showed the slopes regression lines for males and females were the same (parallel), F (1, 277) = .215, p = .643. Furthermore, there was no significant difference between the intercepts of the regression lines for eighth grade males and females, F (1, 277) = .307, p = .580. The results of the analysis demonstrated the correlation between eighth grade PAS and TCAP reading scores for males (r = .781) was stronger than the correlation for females (r = .779). However, the difference between the two correlations was not significant, Fisher's z = .04, p = .968. It should be noted that all three null hypotheses were retained. For the most part the two regression lines for eighth grade males and females in Figure 3 show the two regression lines for eighth grade males and females were very similar.

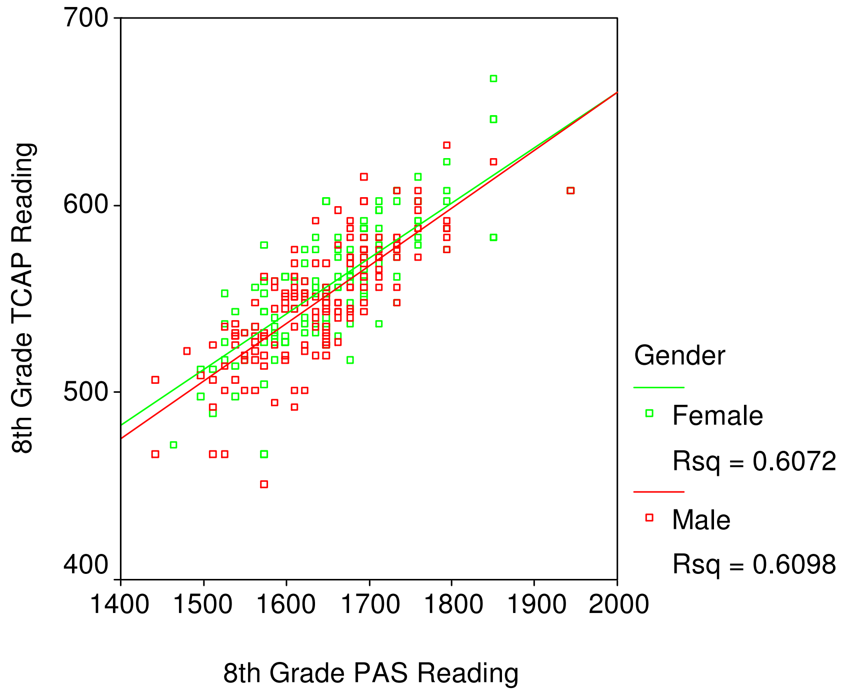

Figure 3 Scatter Plot of the Regression Lines of Eighth Grade PAS and TCAP Reading Scores for Males and Females

Notes: $\hat{y}_{\text{males}} = 42.004 + .309x$; $\hat{y}_{\text{females}} = 68.720 + .296x$

Fourth Grade Math

To evaluate the differences, if any, in the relationships between fourth grade PAS and TCAP math scores for males and females, three hypotheses were tested:

$Ho2_{10}$: The slopes of the regression lines for fourth grade PAS and TCAP math scores for males and females are homogeneous (parallel).

$Ho2_{11}$: There is no difference in the intercepts of the regression lines for fourth grade PAS and TCAP math scores between males and females.

$Ho2_{12}$: There is no difference in the fourth grade PAS and TCAP math correlations for male and female students.

The test of the homogeneity of slopes for fourth grade TCAP math scores regressed on fourth grade PAS scores showed the slopes regression lines for males and females were the same (parallel), F (1, 287) = 3.583, p = .059. Also, there was no difference between the intercepts of the regression lines for fourth grade males and females, F (1, 287) = 3.587, p = .059. While the correlation between fourth grade PAS and TCAP math scores for females (r = .744) was stronger than the correlation for males (r = .678), the difference between the two correlations was not significant, Fisher's z = -1.13, p = .259. All three null hypotheses were retained. Figure 4 shows the two regression lines for fourth grade males and females are very similar.

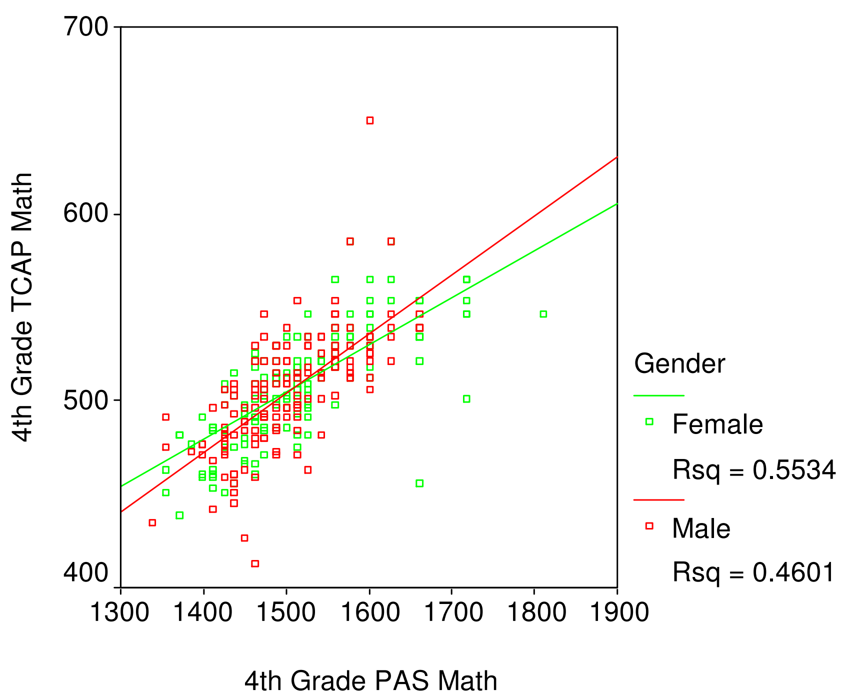

Figure 4. Scatter Plot of the Regression Lines of Fourth Grade PAS and TCAP Math Scores for Males and Females

Notes: $\hat{y}_{males} = 28.998 + .317x$; $\hat{y}_{females} = 125.444 + .253x$

Sixth Grade Math

To evaluate the differences, if any, in the relationships between sixth grade PAS and TCAP math scores for males and females, three hypotheses were tested:

$Ho2_{13}$: The slopes of the regression lines for sixth grade PAS and TCAP math scores for males and females are homogeneous (parallel).

$Ho2_{14}$: There is no difference in the intercepts of the regression lines for sixth grade PAS and TCAP math scores between males and females.

$Ho2_{15}$: There is no difference in the sixth grade PAS and TCAP math correlations for male and female students.

The test of the homogeneity of slopes for sixth grade TCAP math scores regressed on sixth grade PAS scores showed the slopes regression lines for males and females were the same (parallel), F (1, 263) = .017, p = .897. Similarly, there was no difference between the intercepts of the regression lines for sixth grade males and females, F (1, 263) = .006, p = .938. While the correlation between sixth grade PAS and TCAP math scores for females (r = .769) was stronger than the correlation for males (r = .757), the difference between the two correlations was not significant, Fisher's z = - .23, p = .818. All three null hypotheses were retained. Figure 5 shows the two regression lines for sixth grade males and females are very similar.

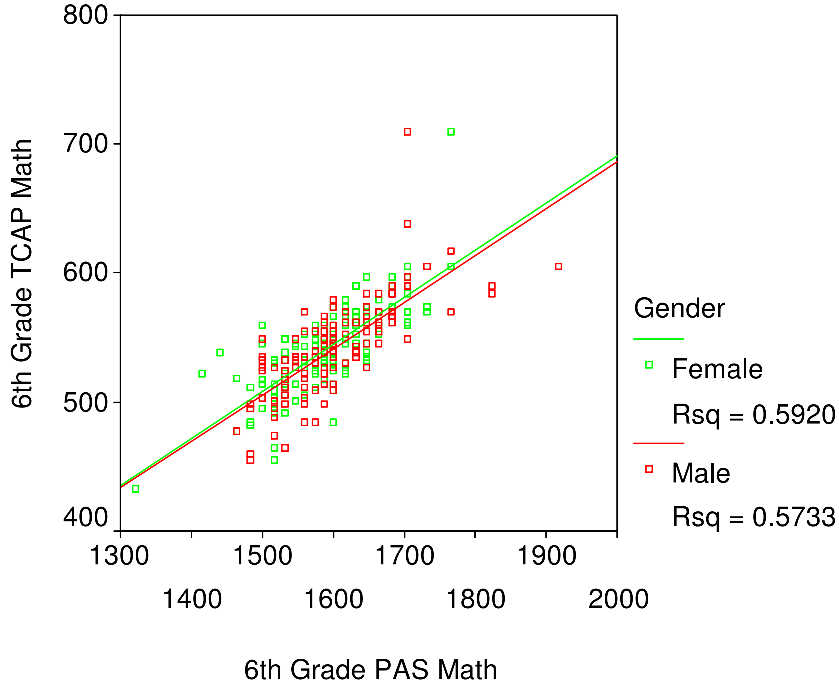

Figure 5. Scatter Plot of the Regression Lines of Sixth Grade PAS and TCAP Math Scores for Males and Females

Notes: $\hat{y}_{males} = -33.540 + .360x$; $\hat{y}_{females} = -38.234 + .365x$

Eighth Grade Math

To evaluate the differences, if any, in the relationships between eighth grade PAS and TCAP math scores for males and females, three hypotheses were tested:

$Ho2_{16}$: The slopes of the regression lines for eighth grade PAS and TCAP Math scores for males and females are homogeneous (parallel).

$Ho2_{17}$: There is no difference in the intercepts of the regression lines for eighth grade PAS and TCAP Math scores between males and females.

$Ho2_{18}$: There is no difference in the eighth grade PAS and TCAP Math correlations for male and female students.

The test of the homogeneity of slopes for eighth grade TCAP math scores regressed on eighth grade PAS scores showed the slopes regression lines for males and females were the same (parallel), F (1, 280) = .255, p = .614. In addition, there was no difference between the intercepts of the regression lines for eighth grade males and females, F (1, 280) = .337, p = .562. Although the correlation between eighth grade PAS and TCAP math scores for males (r = .815) was stronger than the correlation for females (r = .789), the difference between the two correlations was not significant, Fisher's z = .271, p = .542. All three null hypotheses were retained. Figure 6 shows the two regression lines for eighth grade males and females are very similar.

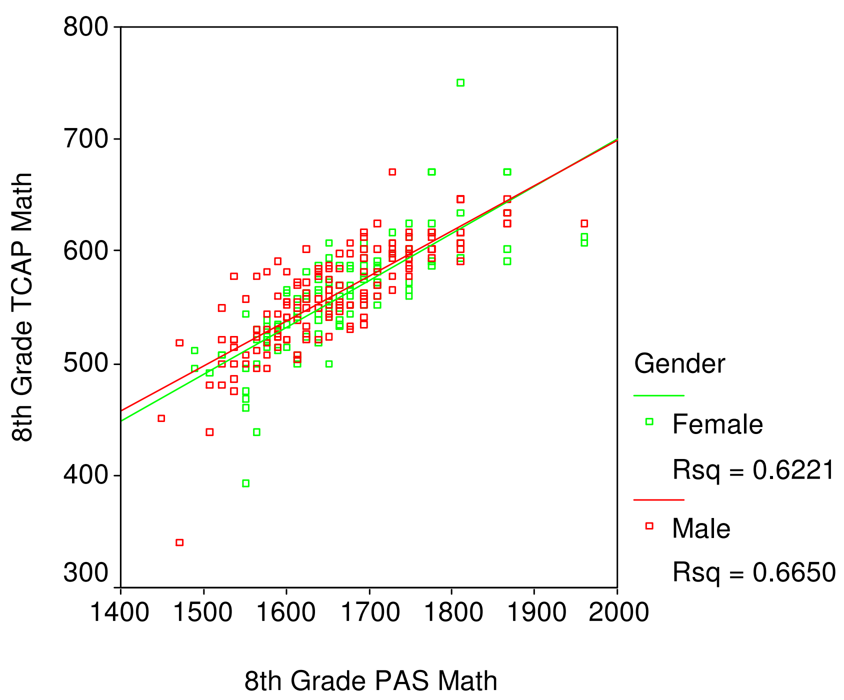

Figure 6. Scatter Plot of the Regression Lines of Eighth Grade PAS and TCAP Math Scores for Males and Females

Notes: $\hat{y}_{males} = -103.413 + .401x$; $\hat{y}_{females} = -138.872 + .419x$

Fourth Grade Science

To evaluate the differences, if any, in the relationships between fourth grade PAS and TCAP science scores for males and females, three hypotheses were tested:

$Ho2_{19}$: The slopes of the regression lines for fourth grade PAS and TCAP science scores for males and females are homogeneous (parallel).

$Ho2_{20}$: There is no difference in the intercepts of the regression lines for fourth grade PAS and TCAP science scores between males and females.

$Ho2_{21}$: There is no difference in the fourth grade PAS and TCAP science correlations for male and female students.

The test of the homogeneity of slopes for fourth grade TCAP science scores regressed on fourth grade PAS scores showed the slopes regression lines for males and females were the same (parallel), F (1, 281) = 2.393, p = .123. There was no difference between the intercepts of the regression lines for fourth grade males and females, F (1, 281) = 2.429, p = .120. While the correlation between fourth grade PAS and TCAP science scores for females (r = .753) was stronger than the correlation for males (r = .688), the difference between the two correlations was not significant, Fisher's z = - 1.13, p = .259. All three null hypotheses were retained. Figure 7 shows the two regression lines for fourth grade males and females are very similar.

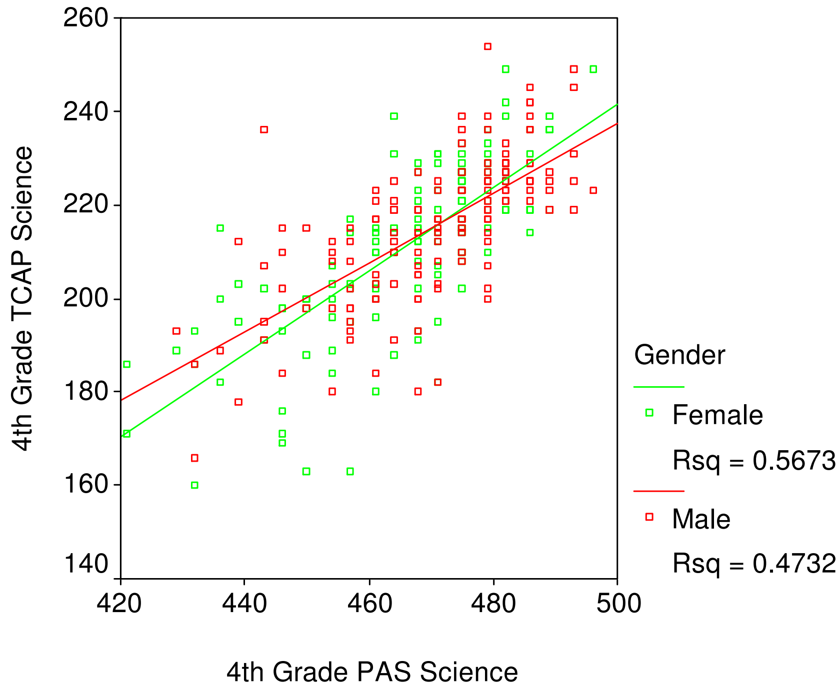

Figure 7. Scatter Plot of the Regression Lines of Fourth Grade PAS and TCAP Science Scores for Males and Females

Notes: $\hat{y}_{\text{males}} = -134.268 + .744x$; $\hat{y}_{\text{females}} = -202.295 + .888x$

Sixth Grade Science

To evaluate the differences, if any, in the relationships between sixth grade PAS and TCAP science scores for males and females, three hypotheses were tested:

$Ho2_{22}$: The slopes of the regression lines for sixth grade PAS and TCAP Science scores for males and females are homogeneous (parallel).

$Ho2_{23}$: There is no difference in the intercepts of the regression lines for sixth grade PAS and TCAP science scores between males and females.

$Ho2_{24}$: There is no difference in the sixth grade PAS and TCAP science correlations for male and female students.

The test of the homogeneity of slopes for sixth grade TCAP Science scores regressed on sixth grade PAS scores showed the slopes regression lines for males and females were the same (parallel), *F* (1, 266) = .075. Subsequently, there was no difference between the intercepts of the regression lines for sixth grade males and females, *F* (1, 266) = .074. *p* = .786. While the correlation between sixth grade PAS and TCAP science scores for females (*r* = .713) was stronger than the correlation for males (*r* = .695), there was no significant difference between the two correlations, Fisher's z = -.29, *p* = .772. All three null hypotheses were retained. Figure 8 shows the two regression lines for sixth grade males and females are very similar.

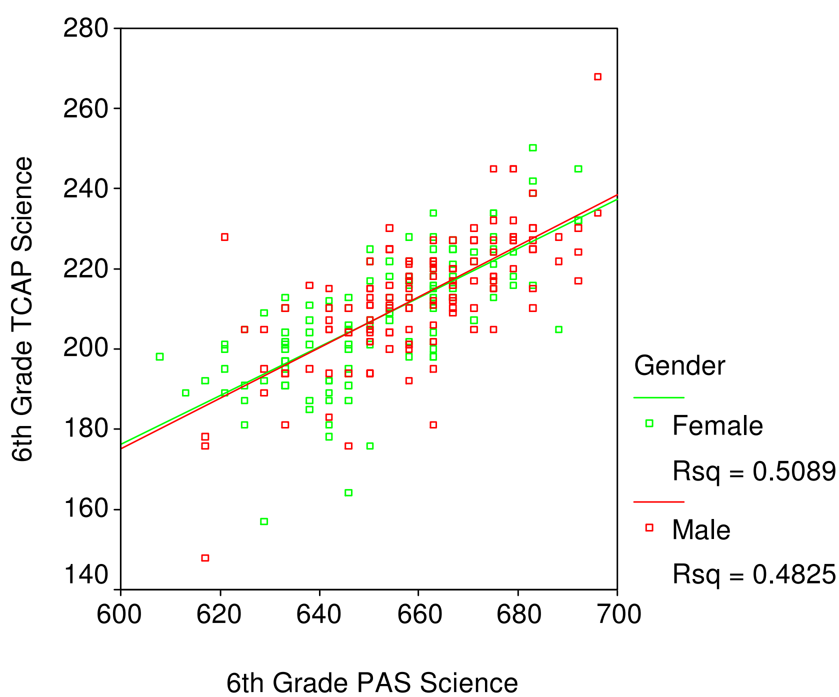

Figure 8. Scatter Plot of the Regression Lines of Sixth Grade PAS and TCAP Science Scores for Males and Females

Notes: $\hat{y}_{males} = -205.292 + .634x$; $\hat{y}_{females} = -191.531 + .613x$

Eighth Grade Science

To evaluate the differences, if any, in the relationships between eighth grade PAS and TCAP science scores for males and females, three hypotheses were tested:

$Ho2_{25}$: The slopes of the regression lines for eighth grade PAS and TCAP science scores for males and females are homogeneous (parallel).

$Ho2_{26}$: There is no difference in the intercepts of the regression lines for eighth grade PAS and TCAP science scores between males and females.

$Ho2_{27}$: There is no difference in the eighth grade PAS and TCAP science correlations for male and female students.

The test of the homogeneity of slopes for eighth grade TCAP science scores regressed on eighth grade PAS scores showed the slopes regression lines for males and females were the same (parallel), F (1, 281) = .013, p = .909. Also, there was no difference between the intercepts of the regression lines for eighth grade males and females, F (1, 281) = .013, p = .909. While the correlation between eighth grade PAS and TCAP science scores for males (r = .765) was stronger than the correlation for females (r =.707), the difference between the two correlations was not significant, Fisher's z = 1.06, p = .289. All three null hypotheses were retained. Figure 9 shows the two regression lines for eighth grade males and females are very similar.

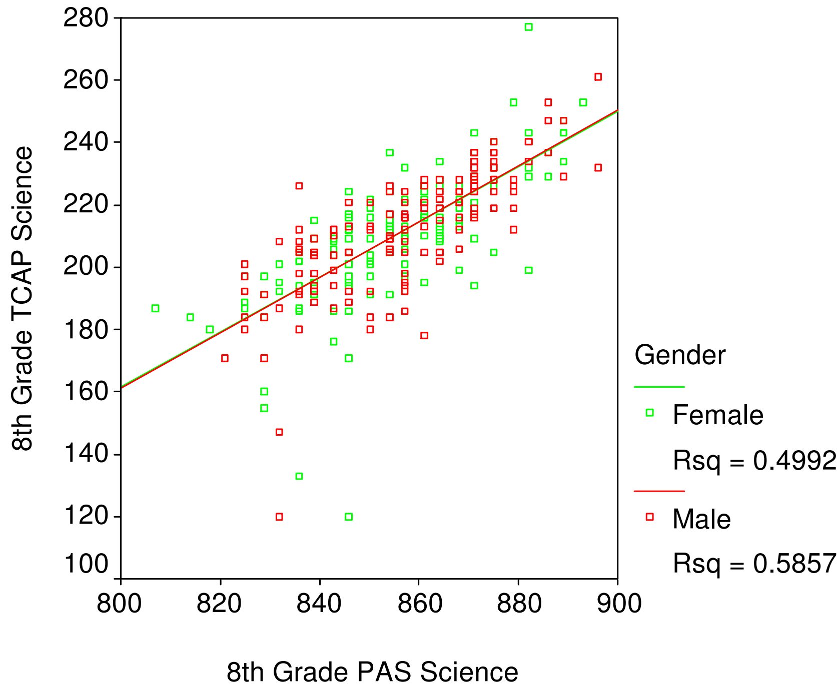

Figure 9. Scatter Plot of the Regression Lines of Eighth Grade PAS and TCAP Science Scores for Males and Females.

Notes: $\hat{y}_{\text{males}} = -551.070 + .890x$; $\hat{y}_{\text{females}} = -541.594 + .879x$

Research Question # 3

Are the relationships between PAS tests and TCAP tests in reading, math, and science the same for students attending Title I and Non-Title I schools?

The population included 372 (62.6%) students who attended schools that qualify for Title I funds and 222 (37.4%) students who attended Non-Title I schools. In fourth grade 200 (64.9%) students attended schools that qualify for Title I funds and 108 (35.1%) students attended Non-Title schools. In sixth grade 172 (60.1%) students attended Title I schools and 114 (39.9%) students attended Non-Title I schools.

Fourth Grade Reading

To evaluate the differences, if any, in the relationships between fourth grade PAS and TCAP reading scores for Title I and Non-Title I students, three hypotheses were tested:

$Ho3_1$: The slopes of the regression lines for fourth grade PAS and TCAP reading scores for Title I and Non-Title I students are homogeneous (parallel).

$Ho3_2$: There is no difference in the intercepts of the regression lines for fourth grade PAS and TCAP reading scores between Title I and Non-Title I students.

$Ho3_3$: There is no difference in the fourth grade PAS and TCAP reading correlations for Title I and Non-Title I students.

The test of the homogeneity of slopes for fourth grade TCAP reading scores regressed on fourth grade PAS scores showed the slopes regression lines for Title I and Non-Title I students were the same (parallel), $F\ (1, 289) = 1.357$, $p = .245$. Likewise, there was no difference between the intercepts of the regression lines for fourth grade Title I and Non-Title I students, $F\ (1, 289) = .828$, $p = .364$. The correlation between fourth grade PAS and TCAP reading scores for Non-Title I students ($r = .747$) was statistically stronger than the correlation for Title I students ($r = .548$). The difference between the two correlations was statistically significant, Fisher's z = 2.84, $p = .005$. The correlation for Non-Title I students ($r = .747$) was strong, whereas, the correlation for Title I students ($r = .548$) was moderate. The null hypotheses for the parallel slopes of the regression lines and the difference between the intercepts were retained, while the null hypothesis for the difference between the correlations for Non-Title I students and Title I students was rejected. Figure 10 shows the two regression lines for fourth grade Title I and Non-Title I students are similar.

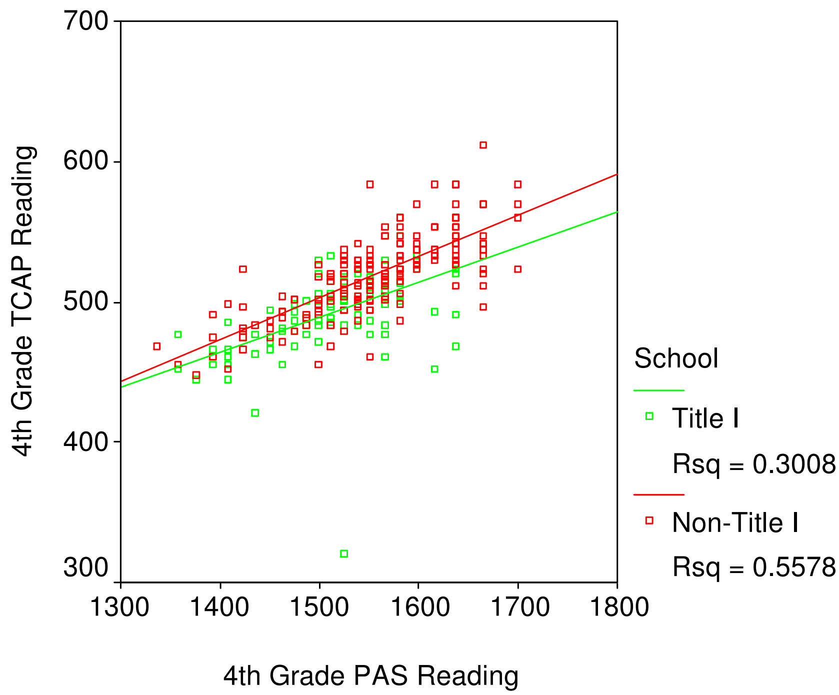

Figure 10. Scatter Plot of the Regression Lines of Fourth Grade PAS and TCAP Reading Scores for Title I and Non-Title I students.

Notes: $\hat{y}_{\text{Non-Title I}} = 58.139 + .296x$; $\hat{y}_{\text{title I}} = 112.143 + .251x$

Sixth Grade Reading

To evaluate the differences, if any, in the relationships between sixth grade PAS and TCAP reading scores for Title I and Non-Title I students, three hypotheses were tested:

$Ho3_4$: The slopes of the regression lines for sixth grade PAS and TCAP reading scores for Title I and Non-Title I students are homogeneous (parallel).

$Ho3_5$: There is no difference in the intercepts of the regression lines for sixth grade PAS and TCAP reading scores between Title I and Non-Title I students.

$Ho3_6$: There is no difference in the sixth grade PAS and TCAP reading correlations for Title I and Non-Title I students.

The test of the homogeneity of slopes for sixth grade TCAP reading scores regressed on sixth grade PAS scores showed there was no significant difference, $F(1, 269) = 1.142$, $p = .286$. Moreover, there was no difference between the intercepts of the regression lines for sixth grade Title I and Non-Title I students, $F(1, 269) = 1.140$, $p = .287$. The correlation between sixth grade PAS and TCAP reading scores for Non-Title I students ($r = .772$) was a little stronger than the correlation for Title I students ($r = .660$). The difference between the two correlations was not significant, Fisher's z = 1.86, $p = .063$. All three null hypotheses were retained. Figure 11 shows the two regression lines for sixth grade Title I and Non-Title I students are similar.

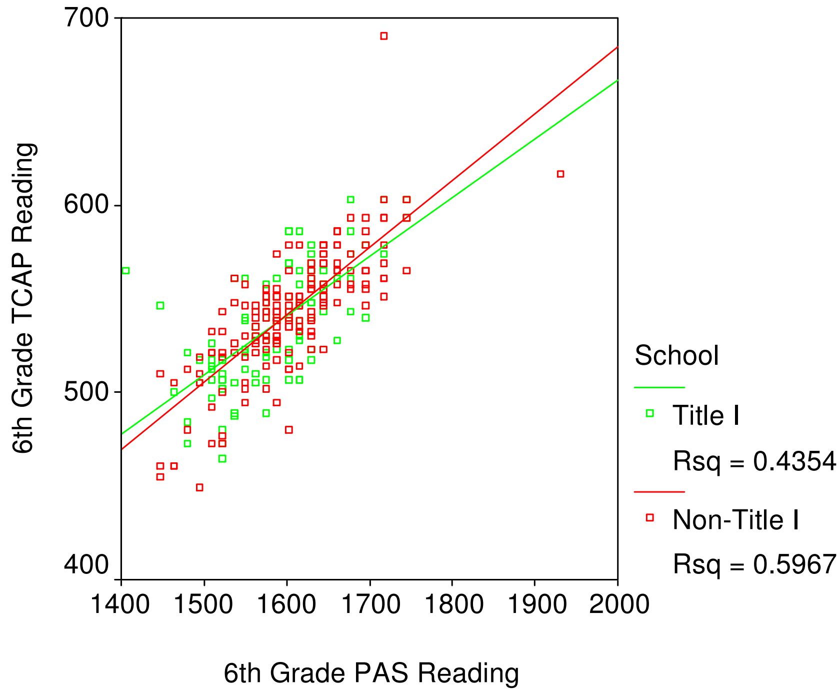

Figure 11. Scatter Plot of the Regression Lines of Sixth Grade PAS and TCAP Reading Scores for Title I and Non-Title I students.

Notes: $\hat{y}_{\text{Non-Title I}} = -33.046 + .359x$; $\hat{y}_{\text{title I}} = 36.539 + .315x$

Fourth Grade Math

To evaluate the differences, if any, in the relationships between fourth grade PAS and TCAP math scores for Title I and Non-Title I students, three hypotheses were tested:

$Ho3_7$: The slopes of the regression lines for fourth grade PAS and TCAP math scores for Title I and Non-Title I students are homogeneous (parallel).

$Ho3_8$: There is no difference in the intercepts of the regression lines for fourth grade PAS and TCAP Math scores between Title I and Non-Title I students.

$Ho3_9$: There is no difference in the fourth grade PAS and TCAP math correlations for Title I and Non-Title I students.

The test of the homogeneity of slopes for fourth grade TCAP Math scores regressed on fourth grade PAS scores showed the slopes regression lines for Title I and Non-Title I students were the same (parallel), F (1, 287) = .097, p = .755. It should also be noted that there was no difference between the intercepts of the regression lines for fourth grade Title I and Non-Title I students, F (1, 289) = .000, p = .990. The correlation between fourth grade PAS and TCAP Math scores for Non-Title I students (r = .715) was stronger than the correlation for Title I students (r = .603). The difference between the two correlations was not statistically significant, Fisher's z = 1.61, p = .107. The null hypotheses for the parallel slopes of the regression lines, the difference between the intercepts, and the null hypothesis for the difference between the correlations for Non-Title I students and Title I students were all retained. Figure 12 shows the two regression lines for fourth grade Title I and Non-Title I students are very similar.

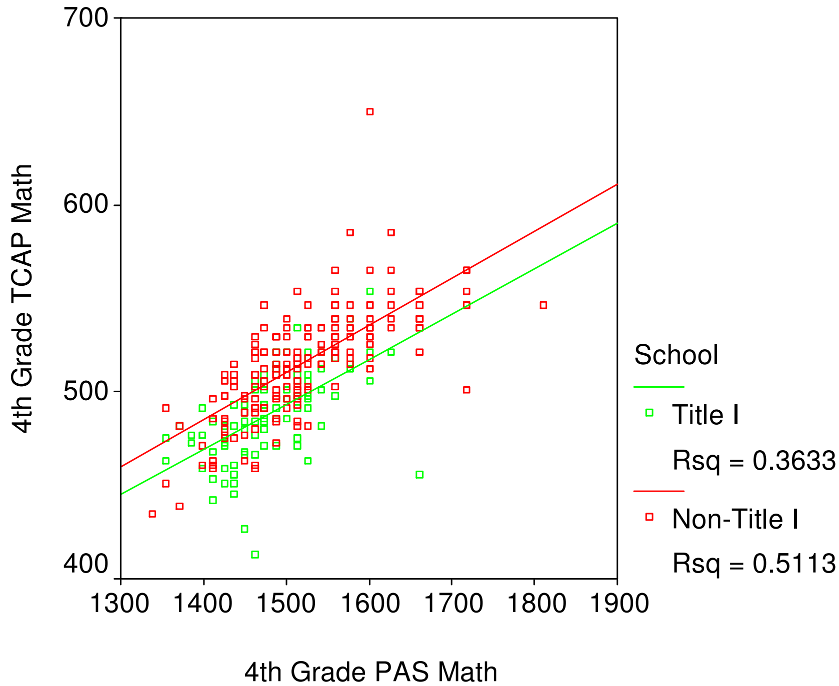

Figure 12. Scatter Plot of the Regression Lines of Fourth Grade PAS and TCAP Math Scores for Title I and Non-Title I students.

Notes: $\hat{y}_{\text{Non-Title I}} = 130.745 + .253x$; $\hat{y}_{\text{title I}} = 131.437 + .241x$

Sixth Grade Math

To evaluate the differences, if any, in the relationships between sixth grade PAS and TCAP math scores for Title I and Non-Title I students, three hypotheses were tested:

$Ho3_{10}$: The slopes of the regression lines for sixth grade PAS and TCAP math scores for Title I and Non-Title I students are homogeneous (parallel).

$Ho3_{11}$: There is no difference in the intercepts of the regression lines for sixth grade PAS and TCAP math scores between Title I and Non-Title I students.

$Ho3_{12}$: There is no difference in the sixth grade PAS and TCAP math correlations for Title I and Non-Title I students.

The test of the homogeneity of slopes for sixth grade TCAP math scores regressed on sixth grade PAS scores showed the slopes regression lines for Title I and Non-Title I students were the same (parallel), F (1, 263) = .812, p = .368. Likewise, there was no difference between the intercepts of the regression lines for sixth grade Title I and Non-Title I students, F (1, 263) = .819, p = .366. The correlation between sixth grade PAS and TCAP math scores for Non-Title I students (r = .780) and the correlation for Title I students (r = .734) were very close. The difference between the two correlations was not significant, Fisher's z = 0.85, p = .395. The null hypotheses for the parallel slopes of the regression lines, the difference between the intercepts, and the null hypothesis for the difference between the correlations for Non-Title I students and Title I students were all retained. Figure 13 shows the two regression lines for sixth grade Title I and Non-Title I students are very similar.

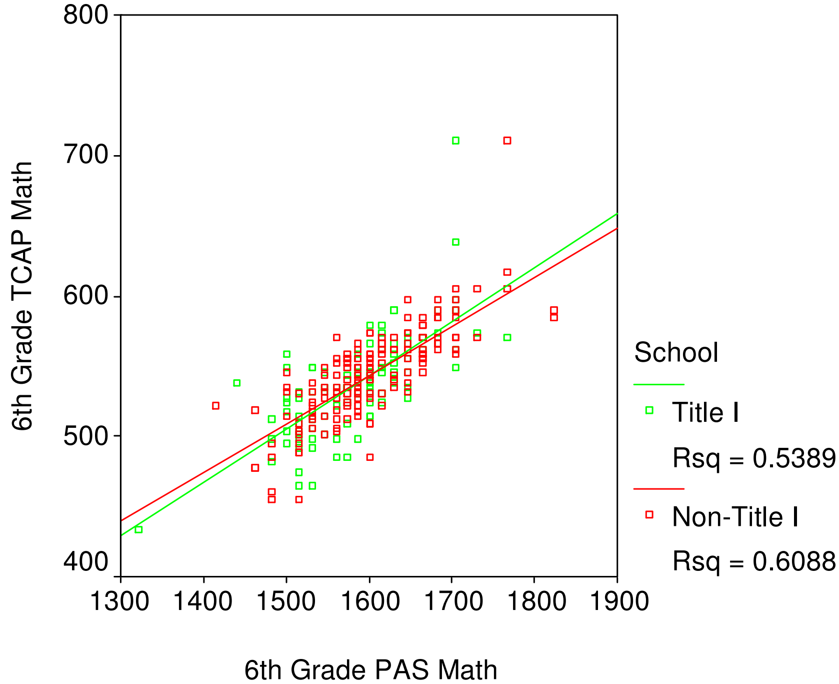

Figure 13. Scatter Plot of the Regression Lines of Sixth Grade PAS and TCAP Math Scores for Title I and Non-Title I students.

Notes: $\hat{y}_{\text{Non-Title I}} = -11.320 + .347x$; $\hat{y}_{\text{title I}} = 68.178 + .382x$

Fourth Grade Science

To evaluate the differences, if any, in the relationships between fourth grade PAS and TCAP science scores for Title I and Non-Title I students, three hypotheses were tested:

$Ho3_{13}$: The slopes of the regression lines for fourth grade PAS and TCAP science scores for Title I and Non-Title I students are homogeneous (parallel).

$Ho3_{14}$: There is no difference in the intercepts of the regression lines for fourth grade PAS and TCAP science scores between Title I and Non-Title I students.

$Ho3_{15}$: There is no difference in the fourth grade PAS and TCAP science correlations for Title I and Non-Title I students.

The test of the homogeneity of slopes for fourth grade TCAP science scores regressed on fourth grade PAS scores showed the slopes regression lines for Title I and Non-Title I students were the same (parallel), F (1, 281) = .018, p = .892. Likewise, there was no difference between the intercepts of the regression lines for fourth grade Title I and Non-Title I students, F (1, 281) = .009, p = .924. The correlation between fourth grade PAS and TCAP science scores for Non-Title I students (r = .726) was almost identical to the correlation for Title I students (r = .715). The difference between the two correlations was not significant, Fisher's z = 0.18, p = .857. The three null hypotheses were all retained. Figure 14 shows the two regression lines for fourth grade Title I and Non-Title I students are very similar.

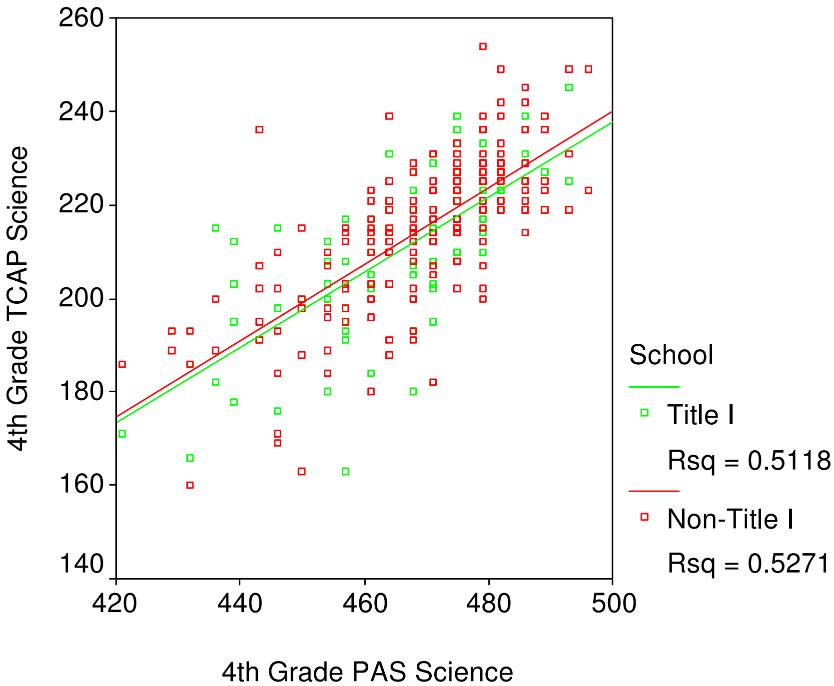

Figure 14. Scatter Plot of the Regression Lines of Fourth Grade PAS and TCAP science Scores for Title I and Non-Title I students.

Notes: $\hat{y}_{\text{Non-Title I}} = -169.605 + .819x$; $\hat{y}_{\text{title I}} = -165.271 + .806x$

Sixth Grade Science

To evaluate the differences, if any, in the relationships between sixth grade PAS and TCAP science scores for Title I and Non-Title I students, three hypotheses were tested:

$Ho3_{16}$: The slopes of the regression lines for sixth grade PAS and TCAP science scores for Title I and Non-Title I students are homogeneous (parallel).

$Ho3_{17}$: There is no difference in the intercepts of the regression lines for sixth grade PAS and TCAP science scores between Title I and Non-Title I students.

$Ho3_{18}$: There is no difference in the sixth grade PAS and TCAP science correlations for Title I and Non-Title I students.

The test of the homogeneity of slopes for sixth grade TCAP science scores regressed on sixth grade PAS scores was not significantly different, F (1, 266) = 1.877, p = .172. Furthermore, there was no difference between the intercepts of the regression lines for sixth grade Title I and Non-Title I students, F (1, 266) = 2.088, p = .150. The correlation between sixth grade PAS and TCAP science scores for Title I students (r = .737) was slightly higher than the correlation for Non-Title I students (r = .717). The difference between the two correlations was not significant, Fisher's z = -0.43, p = .667. All three null hypotheses were retained. Figure 15 shows the two regression lines for sixth grade Title I and Non-Title I students are similar.

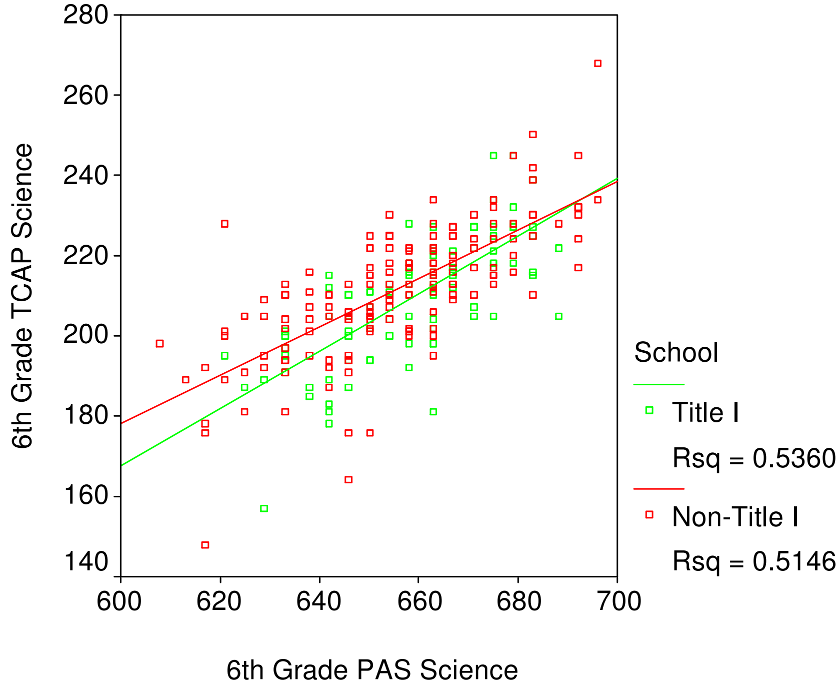

Figure 15. Scatter Plot of the Regression Lines of Sixth Grade PAS and TCAP Science Scores for Title I and Non-Title I students.

Notes: $\hat{y}_{\text{Non-Title I}} = -183.286 + .603x$; $\hat{y}_{\text{title I}} = -261.815 + .716x$

CHAPTER 5

SUMMARY OF FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS

Introduction

The purpose of this study was to determine if a correlation existed between the Predictive Assessment Series (PAS) Test, marketed by Discovery Education, and the Tennessee Comprehensive Assessment Program (TCAP), a criterion-referenced test. The research population was drawn from a Northeast Tennessee school system that administered both the PAS and the TCAP assessments during the 2008-2009 school year. Only students in grade four, grade six, and grade eight who had completed both the TCAP and the PAS assessments were included in the study.

Summary of Findings

The analysis focused on three research questions. The data collection tools included the TCAP a criterion-referenced test that is completed by fourth, sixth, and eighth grade students using paper and pencil to complete, and a computer-based test, the ThinkLink PAS test marketed by Discovery Education. The population consisted of 902 students attending grade four, grade six, and grade eight in the school system participating in the study. Students enrolled in the aforementioned grades during the 2008-2009 school year who completed the PAS test and the TCAP test were included in the study. As a result of school changes and absences some students did not have both a PAS score and a TCAP score. If students from grade four, grade six, and grade eight did not complete both tests, their test scores were excluded from the study due to incomplete information. Data collected for the study were entered into a data file for analysis SPSS for Windows. An alpha level of .001 was used as the criterion for determining statistical significance of the findings for Research Question #1. For Research Question #2 and Research #3, an alpha level of .01 was used to determine the statistical significance of the correlations.

Demographics for Grade 4

Among the grade four students, almost the same number of female students (n = 156, 50.6%) as male students (n=152, 49.4%) participated in the study. With respect to Non-Title I and Title I schools, 200 (64.9%) were from Non-Title I schools and 108 (35.1%) were from Title I schools.

Demographics for Grade 6

Among the grade six students, females (n = 149, 52.1%) outnumbered males (n = 137, 47.9%). In regards to Title status, students attending a Non-Title I school (n = 172, 60.1%) out-numbered those students attending a Title I school (n = 114, 39.9%).

Demographics for Grade 8

Among the grade eight students, males (n = 160, 51.9%) outnumbered females (n = 148, 48.1%).

Research Question #1

Are there relationships between the scaled scores of the PAS and the scaled scores of the TCAP in reading, math, and science for students in grade four, grade six, and grade eight?

This study consisted of 902 students in grade four, grade six and grade eight who were administered the PAS test and the TCAP test in reading, math, and science during the 2008-2009 school year. Pearson Product Moment Correlation statistics were used to analyze the relationship between PAS and TCAP scores in the areas of reading, math, and science. The results indicate a strong to very strong positive relationship between the PAS reading, math, and science scores and corresponding TCAP reading, math, and science scores. These results suggested that a predictive relationship did exist between the PAS and TCAP assessment for the 2008-2009 school year. All correlations were significant at the .001 level and all the null hypotheses were rejected. All nine correlations showed a strong positive relationship between the PAS and TCAP tests. The relationships ranged from a low of .698 in fourth grade reading and a high of .801 in eighth grade math. The strongest relationship was found among eighth graders. Likewise, math showed to have the strongest relationship among the subject areas. For the population as a whole, science had the lowest relationship ($r = .724$) among the three subject areas and fourth grade ($r = .711$) was the lowest among the grade levels.

As stated in the review of literature, there is a large body of research that explores the comparability of scores from paper-and-pencil tests and computer-based tests. According to Bunderson, Inouye, and Olsen (1989) and Wise, Barnes, Harvey, and Plake (1989) computer-based and paper-based test version results are very similar. For this study, the correlations between the PAS computer-based test and the TCAP paper and pencil test were consistently strong across grades, gender, and Title I and Non-Title status.

The stronger correlation between the PAS scores and the TCAP scores for grade eight students could be a result of several factors, including their age, cognitive abilities, and their increased confidence in computer testing. The school district has been using the PAS computer test for 3 years. Shermis and Lombard (1998) found that age and computer anxiety were significant indicators of performance outcomes, which could explain the stronger correlation between the PAS and TCAP among older students.

Research Question #2

Are the relationships between PAS and TCAP tests in reading, math, and science the same for both male and female students?

To evaluate the differences, if any, in the relationships between PAS and TCAP scores the test of homogeneity of slopes was used. The testing of the homogeneity of slopes was chosen to test the difference in the regression slopes (or correlations), as opposed to visually comparing two correlation coefficients. All correlations were significant at the .01 level and all the null hypotheses were rejected.

The results showed that the relationships between PAS and TCAP tests in reading, math, and science were consistent across gender within grade levels. The test of homogeneity of slopes showed the slopes regression lines for males and females were the same (parallel) for grade four, grade six, and grade eight. The highest correlation (r = .815) between PAS and TCAP scores was noted in eighth grade math scores of male students. The lowest correlation (*r* = .652) between PAS and TCAP scores was observed in fourth grade reading scores of male students. Among the grade levels, eighth grade had the strongest correlation (*r* = .815) between the two tests. The strongest correlations among subject areas was found in math.

Research Question # 3

Are the relationships between PAS tests and TCAP tests in reading, math, and science the same for students attending Title I and Non-Title I schools?

To evaluate the differences, if any, in the relationships between PAS and TCAP scores the test of homogeneity of slopes was used. The testing of the homogeneity of slopes was chosen to test the difference in the regression slopes (or correlations), as opposed to visually comparing two correlation coefficients. The results showed that the relationships between PAS and TCAP tests in reading, math, and science were consistent across Title I and Non-Title I schools. The test of homogeneity of slopes showed the slopes regression lines for the scores of Title I and Non-Title I students were the same (parallel) for grade four, grade six, and grade eight. Overall, the correlations between PAS and TCAP scores for Title I and Non-Title I students were moderately strong to very strong. Only one of the null hypotheses was rejected for Research Question #3. The correlation between fourth grade PAS and TCAP reading scores for Non-Title I students (*r* = .747) was statistically stronger than the correlation for Title I students (*r* = .548). As a result, the null hypothesis for the difference between the correlations for Non-Title I students and Title I students was rejected.

Conclusions

Based on the analysis and findings of this study, using the ThinkLink PAS test appears to have been successful in predicting how well students will perform on the state assessment. Overall, the correlations between the PAS and TCAP were consistent across grades, across gender within grade levels, and with Title I and Non-Title I students. The findings also show that it was possible to calculate a predicted TCAP score in reading, mathematics, and science. This was an important finding because the ability of the PAS assessment to predict TCAP scores could be another tool to provide educators the opportunity to target students who are potentially at risk of not meeting state benchmark proficiency levels. With the passage of the *No Child Left Behind* Act, it is necessary for school systems to use the most effective benchmark assessment. Identifying and diagnosing at-risk students early on would provide educators more time for intervention. Research is clear that Discovery Education's ThinkLink PAS test accurately predicted how students would score on the TCAP test during the 2008/2009 school year. The following conclusions emerged from this study:

Conclusion #1

Based on findings from the study, there appears to be a positive relationship between the scaled scores of the PAS and the scaled scores of the TCAP in reading, math, and science for students in grade four, grade six, and grade eight. The 2008-2009 test data that were analyzed showed that the strongest relationships were in eighth grade and in the area of math.

Conclusion #2

Based on the results of the study, there are no differences in the relationships between the PAS and TCAP reading, math, and science scores for males and females in grade four, grade six, and grade eight. The regression lines in reading, math, and science for fourth, sixth, and eighth grade males and females are very similar.

Conclusion #3

The relationships between the PAS tests and TCAP tests in reading, math, and science appear to be similar for fourth and sixth grade students attending a Title I or Non-Title I school. According to the results from the test of the homogeneity of slopes, the correlation between fourth grade PAS and TCAP reading scores for Non-Title I students was slightly stronger than the correlation for Title I students. The only significant finding was in fourth grade reading. The difference between the two correlations was statistically significant in reading for Non-Title I and Title I students in fourth grade.

Recommendations for Practice

The following are recommendations for practice:

1. The continuation of using the ThinkLink Predictive Assessment test in the school system that participated in the study.
2. Other school systems should consider the use of the ThinkLink Predictive Assessment Series (PAS) test or other predictive tests to provide teachers with timely feedback in order to make adjustments to future instruction.
3. All school systems should consider the adoption of ThinkLink PAS or other

preassessment tools that accurately predict the progress of students toward mastery of the state standards.

Recommendations for Further Research

In this age of accountability school systems all over the United States have looked for ways to predict student performance on annual state tests. As a result of the high stakes associated with student performance on the state assessments, many are looking to implementing the use of benchmark assessments to identify students who are potentially at risk of not making state benchmark proficiency levels. Remediation and timely intervention strategies could be provided with early identification and diagnosis. The following are recommendations for further research:

1. A replication of this study should be conducted in another school system within the state of Tennessee that is more reflective of the state's demographic population.
2. A replication of this study should be conducted using an outcome criterion other than the Tennessee Comprehensive Assessment Program standardized assessment.
3. Replication of this study using a larger population size and/or analyzing more than 1 year of data.
4. Use a qualitative research approach to evaluate teachers', parents', administrators', and students' perceptions of the ThinkLink Predictive Assessment Series test or similar assessment.
5. Implementation of a study that evaluates teachers' and schools' differences to

identify strategies that could potentially produce better results.

6. Implementation of a study that evaluates how schools are using data to inform instructional practice including changes to instructional calendars, curriculum mapping, reteaching, and other classroom strategies based on what benchmark assessments reveal.
7. The current study was limited to students in grade four, grade six, and grade eight; future studies should include grade three, grade five, and grade seven in order to increase the population that it may be generalized to.
8. A quantitative research approach to determine if benchmark testing helps change student outcomes.
9. A study to examine the possibilities of replacing current paper and pencil standardized tests with online assessments.
10. A comparison study to evaluate the multiple predictive assessments that are available and determine their strengths and weaknesses.

REFERENCES

Abrams, L., & Madaus, G. (2003). The lessons of high-stakes testing: Research shows that high-stakes tests can affect teaching and learning in predictable and often undesirable ways. *Educational Leadership, 61, 3*, 31-35.

Amrein, A. L., & Berliner, D. C. (2002). *The impact of high-stakes tests on student academic performance: An analysis of NAEP results in states with high-stakes tests and ACT, SAT, and AP test results in state exams with high school graduation exams.* Arizona State University College of Education, Educational Policy Studies Laboratory, Retrieved June 20, 2009, from http://edpolicylab.org/

Asp, E. (2000). Assessment in education: Where have we been? Where are we headed? *Education in a new era: The 2000 ASCD Yearbook.* Alexandria, VA: ASCD.

Bahr, M. W., & Bahr, C. M. (1997). Educational assessment in the next millennium: Contributions of technology. *Preventing School Failure, 41*, 90-94.

Baker, A. P., Xu, D., & Detch, E. (1995). *A review of the Tennessee value-added assessment system.* (Authorization Number 307236), Nashville, TN, Comptroller of the Treasury.

Bennett, R. E. (1999). Using new technology to improve assessment. *Educational Measurement: Issues and Practice, 18*(3), 5-12.

Bennett, R. E. (2001). How the internet will help large-scale assessment reinvent itself. *Education Policy Analysis Archive*, 9(5). Retrieved July 17, 2010 from http://epaa.asu.edu/epaa/v9n5.html

Bennett, R. E. (2002). Inexorable and inevitable: The continuing story of technology and assessment. *Journal of Technology, Learning, and Assessment, 1*(1), 1-24.

Bernt, F. M., Bugbee, A. C., & Arceo, R. D. (1990). Factors influencing student resistance to computer administered testing. *Journal of Research on Computing in Education, 22*(3), 265-275.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 1-65.

Black, P., & Wiliam, D. (1998b). Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan*, 80(2). Retrieved July 8, 2009, from http://www.teachingexpertise.com/articles/black-william-assessment-learning-118

Brandt, R. S. (2000). *Education in a new era: The 2000ASCD yearbook*. Alexandria, VA: ASCD.

Bredekamp, S., & Rosegrant, T. (1992). *Reaching potentials: Appropriate curriculum and assessment for young children* (Vol. 1). Washington, DC: National Association of the Education of Young Children.

Brown, R. S., & Coughlin, E. (2007). *The predictive validity of selected benchmark assessments used in the Mid-Atlantic Region.* Retrieved from: http://www.mhkids.com/media/articles/pdfs/resources/PredictiveValidity.pdf

Bugbee, A. C. (1996). The equivalence of paper-and-pencil and computer-based testing. *Journal of Research on Computing in Education, 28*, 282-299.

Bugbee, A. C., & Bernt, F. M. (1990). Testing by computer: Findings in six years of use 1982-1988. *Journal of Research on Computing in Education, 23*(1), 87-100.

Bunderson, C. V., Inouye, D. K., & Olsen, J. B. (1989). The four generations of computerized educational measurement in R L Linn (ed) *Educational measurement*, Washington, DC: American Council on Education, 367-407.

California Learning Resource Network. (2008). *Electronic learning assessment resources*. Retrieved July 10, 2009, from: http://www.clrn.org/elar/details.cfm?elarid=86

Casbarro, J. (2004). Reducing anxiety in the era of high-stakes testing. *Principal, 83,* 36-38.

Casbarro, J. (2005). The politics of high-stakes testing. *Principal, 84*, 16-20.

Center for Greater Philadelphia (2004). *Value-added assessment.* Retrieved July 23, 2009, from: http://www.cgp.upenn.edu/ope_value.html

Center for Public Education. (2006). *A guide to the No Child Left Behind Act.* Retrieved July 24, 2009, from: http://www.centerforpubliceducation.org/site/c.lvIXIiN0JwE/b.5056891/apps/s/content.asp?ct=6857877

Clariana, R., & Wallace, P. (2002). Paper-based versus computer-based assessment: Key factors associated with the test mode effect. *British Journal of Educational Technology, 33,(5)* 593-602.

Coffey, H. (2009). Benchmark assessments. *Learn NC*. Retrieved July 14, 2009, from http://www.learnnc.org/lp/pages/5317?style=print

Daniel, H., & Wheeler, B. (2006). *The uses of benchmark tests to improve student learning* (ThinkLink Learning/Discovery Education). Retrieved June 6, 2009, from: http://www.leadered.com/06Symposium/pdf/USES%20OF%20BENCHMARK%20TESTS.pdf

Davis, A. (1998). *The limits of educational assessment*. Oxford, England: Blackwell.

Earl, L., & Katz, S. (2002). Leading schools in a data-rich world. In K. Leithwood and P. Hallinger (Eds.), *Second international handbook of leadership and administration*. Dordrecht: Kluwer.

Feller, B. (2006, May 12). No states meet teacher quality goal. *The Washington Post.*

Fleischman S., & Safer, N. (2005). How student progress monitoring improves instruction. *Educational Leadership*, 62, (5) 81-82.

Gandal M., & McGiffert L. (2003). The power of testing. *Educational Leadership, 60*, 5, 39-42.

Green, J. P., Winters, M. A., & Forster, G. (2003). *Testing high stakes tests: Can we believe the results of accountability tests.* (Civic Report 33), Manhattan Institute for Policy Research, Retrieved June 20, 2009 from http://www.manhattan-institute.org/html/cr_33.htm.

Gay, L., Mills, G., & Airasion, P. (2006). *Educational research: Competencies for analysis and application*. Upper Saddle River, NJ: Pearson.

Guilfoyle, C. (2006). NCLB: Is there life beyond testing? *Educational Leadership, 64, 3*, 8-13.

Haas, C., & Hayes, J. R. (1986). What did I just say? Reading problems in writing with the machine. *Research in the Teaching of English*, *20*(1), 22-35.

Harvey, L., (2004 – 2011). *Analytic Quality Glossary,* Quality Research International, http://www.qualityresearchinternational.com/glossary/

Hellend, K. (2001). Value-added assessment. *Evergreen Freedom Foundation, 11*(5), 1-6.

Henderson, S. (2008). Do benchmark assessments make a difference? A first look. WestEd *R&D Alert, 9*, 2, 10.

Henderson, S., Petrosino, A., Guckenburg, S., & Hamilton, S. (2007). *Measuring how benchmark assessments affect student achievement* (Issues & Answers Report, REL 2007-No.039). U. S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands. Retrieved July 21, 2009, from http://ies.ed.gov/ncee/edlabs

Herman, J. L., & Baker, E. L. (2005). Making benchmark testing work. *Educational Leadership*, 63,(3) 48-54.

Herman, J. L., Osmundson, E., & Dietel, R. (2010). *Benchmark assessments for* *improved learning* (AACC Policy Brief). Los Angeles, CA: University of California.

Hershberg, T. (2004a). Measuring what matters. *American School Board Journal, 191,* 27-31.

Hershberg, T. (2004b). *Value-added assessment*. Center for Greater Philadelphia. Retrieved June 13, 2009, from http://www.cgp.upenn.edu/ope_value.html

Herszenhorn, D. (2003, July 23). Basic skills forcing cuts in art classes. *The New York Times*, p. B1.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1998). *Applied statistics for the behavioral sciences* (4th ed.). New York: Houghton Mifflin.

Holloway, J. H. (2000). A Value-added view of pupil performance. *Educational Leadership, 57,* 84-85.

Kurdek, L. A., & Sinclair, R. J. (2001). Predicting reading and mathematic achievement in fourth-grade children from kindergarten readiness scores. *Journal of Educational Psychology, 93*, 451-455.

Madaus, G., & O'Dwyer, L. (1999). A short history of performance assessment: Lessons learned. *Phi Delta Kappan, 80*, 688-695.

McAdams, D. R. (2002). Enemy of the good. *Education Next, 2*(i2), 23.

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist, 35*, 1012-1027.

Mourant, R. R., Lakshmanan R., & Chantadisai, R. (1981). Visual fatigue and cathode ray tube display terminals. *Human Factors, 23*(5), 529-540.

National Association of State Boards of Education. (2001). *Any time, any place, any path, any pace: Taking the lead on e-learning policy.* Retrieved July 20, 2009, from http://www.nasbe.org/Organization_Information/e_learning.pdf

National Center for Education Statistics. (2000, February). *Internet access in U.S. public schools and classrooms: 1994-1999* (NCES 2000-086). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

National Center for Education Statistics. (2000). *Teacher use of computers and the internet in public schools* (NCES 2000-090). Retrieved July 25, 2009, from http://nces.ed.gov/pubs2000/2000090.pdf

National Center for Fair and Open Testing (1998). *Computerized Testing: More*

Questions Than Answers. FairTest Fact Sheet. Retrieved May 16, 2009, from http://www.fairtest.org/facts/computer.htm

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* Retrieved June 26, 2009, from http://www.ed.gov/pubs/NatAtRisk

Neil, M. (2003). The dangers of testing. *Educational Leadership, 60*(5), 43-45.

Neill, M. (2006). The case against high-stakes testing. *Principal*, 85, 4, 28-32.

Nichols, S., Glass, G., & Berliner, D. (2005). *High-stakes testing and student achievement: Problems for the No Child Left Behind Act.* Tempe, AZ: Educational Policy Research Unit.

Nichols, S., Glass, G., & Berliner, D. (2007). *Collateral damage: How high-stakes testing corrupts American's schools*. Cambridge, MA: Harvard Education Press.

Noddings, N. (2005). What does it mean to educate the whole child? *Educational Leadership, 63,* 8-13.

Olson, L. (2005a). Benchmark assessments offer regular checkups on student achievement. *Education Week, 25*(13), 13-14. Retrieved July 19, 2009, from http://www.edweek.com

Olson, L. (2005b). Impact of paper-and-pencil, online testing is compared. *Education Week*, *25*, 1, 14.

Owens, A. M. (2002). Putting schools to the test: How standardized exams are changing education in Canada. *National Post, 4*(i18), B1, B2.

Paige, R. (2006). No Child Left Behind: The ongoing movement for public education reform. *Harvard Educational Review, 76*,(4), 461-473.

Park, J. (2003). A test-taker's perspective. *Education Week, 22*(35), 15.

Parshall, C. G. (1999, April). *Audio computer-based training: Measuring more through the use of speech and non-speech sound.* Paper presented at the American Educational Research Association, Montreal, Quebec, Canada.

Pasquier, M., & Gomz-Zwiep, S. (2006). Developing benchmark assessments: A teacher perspective. Retrieved May 20, 2009, from http://www.caesl.org/conference2006/Benchmark_poster.pdf

Pedulla, J., Abrams, L., Madaus, G., Russell, M., Ramos, M., & Miao, J. (2003).

Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers. *National Board on Educational Testing and Public Policy.*

Popham W. (2006). A Tale of Two Test Types. *Principal*, 85, 4, 12-16.

Pruett, K. (2002). *Making sense of the numbers.* Paper presented at TASL (Tennessee Academy of School Leaders) Data Analysis Workshop, Kingsport, TN.

Resnick, M. (Fall, 2003). NCLB Action alert: Tools & tactics for making the law work. National School Boards Association: Alexandria, Va.

Russell, M. (1999). Testing writing on computers: A follow-up study comparing performance on computer and paper. *Educational Policy Analysis Archives*, 7(20). Available at http://epaa.asu.edu/epaa/v7n20/

Russell, M., & Abrams, L. (2004). Instructional uses of computers for writing: The effect of state testing programs. *Teachers College Record 106(6)*, pp.1332-1357.

Russell, M., & Haney, W. (1997). Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper-and-pencil. *Educational Policy Analysis Archives*, 5(3). Available from http://olam.ed.asu.edu/epaa/v5n3.html

Russell, M., & Plati, T. (2001a), Effects of computer versus paper administration of a state-mandated writing assessment. *Teachers College Record*. Available from http://www.tcrecord.org/PrintContent.asp?ContentID=10709

Russell, M., & Plati, T. (2001b). Mode of administration effects on MCAS composition performance for Grade eight and ten. *Teachers College Record*. Available from http://www.tcrecord.org/Content.asp?ContentID=10709

Salkind, N. (2005). Statistics for people who think they hate statistics (2nd ed.). Thousand Oaks, CA: Sage.

Sanders, W., & Rivers, J. (1996). *Cumulative and residual effects of teachers on future students' academics*. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Sanders, W. L. (1998). Value-added assessment. *The School Administrator Web Edition*, 1-5. Retrieved June 24, 2009, from http://www.aasa.org/publications/sa/1998_12/sanders.htm

Sausner, R. (2005). Making assessments work. *District Administration*. Retrieved on August 2, 2009, from http://districtadministration.ccsct.com//page.cfm?p=1188

Shermis, M. D., & Lombard, D. (1998). Effects of computer-based test administrations on test anxiety and performance. *Computers in Human Behavior, 14* (1), 111-123.

Steinberg, J., & Henrique, D. B. (2001, May 21). When a test fails the schools, careers and reputations suffer. *The New York Times*. Retrieved on July 24, 2009, from http://www.nytimes.com/2001/05/21/business/when-a-test-fails-the-schools-careers-and-reputations-suffer.html?pagewanted=1

Taylor, G., Shepard, L., Kinner, F., & Rosenthal, J. (2003). A survey of teachers' perspectives on high-stakes testing in Colorado: What gets taught, what gets lost. *CSE Technical Report 588.*

Teaching Today (2009). *Creating Standards-Based Lessons: Using pre-assessment.* Retrieved Sept 23, 2011, from *http://teachingtoday.glencoe.com/howtoarticles/creating-standards-based-lessons-using-pre-assessment*

Tennessee Department of Education (2003). No Child Left Behind, A handbook for principals, Office of Federal Programs, 1-72.

Tennessee Department of Education (2004). Understanding TCAP achievement test results. Tennessee Department of Education Evaluation and Assessment Division. *CTB/McGraw Hill 2003*, 1-18.

Tennessee Department of Education. (2005). *A handbook for principals*. Nashville, TN: Author.

Tennessee Department of Education. (2006). *K-12 TCAP testing*. Retrieved May 26, 2009, from http://www.state.tn.us/education/assessment/tsachfaq.shtml

Tennessee Report Card. (2004). *Report card terminology.* Retrieved Sept, 2011, from http://www.k-12.state.tn.us/rptcrd04/rptcrdterms.htm

ThinkLink Learning. (2005). Discovery Education Assessment Website. Viewed on August 2, 2009, from http://www.thinklinklearning.com

Trotter, A. (2002). Testing firms see future market in online assessment. *Education Week on the Web, 20*(4), 6.

Tucker, P. D., & Stronge, J. H. (2005). *Linking teacher evaluation and student achievement*. Alexandria, VA: ASCD.

United States Department of Education. (2002). *No child left behind: A desktop reference.* Washington DC: Office of Elementary and Secondary Education.

United States Department of Education. (2003). *Condition of education 2003.*

Washington, DC: Author.

United States Department of Education. (2004). *No child left behind*. Ed.Gov Website. Retrieved July 21, 2009, from: http://nclb.org/start/facts/teachers/html

United States Department of Education. (2005). *The facts about state standards*. Retrieved on July 24, 2009, from http://www.ed.gov/nclb/accountability/standards/standards.html

Wang, S. (2008). Comparability of computer-based and paper-and-pencil testing in k-12 reading assessments. *Educational and Psychological Measurement, 68, 1*, 5-24.

Ward, T. J., Hooper, S. R., & Hannafin, K. M. (1989). The effect of computerized test on the performance and attitudes of college students. *Journal of Educational Computing Research, 5*, 327-333.

Wayman, J. C., Stringfield, S., & Yakimowski, M. (2004). *Software enabling school improvement through analysis of student data* (CRESPAR Technical Report No. 67). Baltimore: Johns Hopkins University.

Wilson, R. F. (2001). *HTML E-mail: text font readability study.* Results of a survey conducted April, 2001. Available at: http://www.wilsonweb.com/wmt6/html-email-fonts.htm

Winstead, M. (2006). *Presentation notes from principals' academy*. Nashville, TN: Author.

Wise, S. L., Barnes, L. B., Harvey, A. L., & Plake, B. S. (1989). Effects of computer anxiety and computer experience on the computer-based achievement test performance of college students. *Applied Measurement in Education,* 235-241.

Wise, S. L., & Plake, B. S. (1990). Computer-based testing in higher education. *Measurement and evaluation in counseling and development, 23*(1), 3-10.

Zehr, M.A. (2006). Monthly checkups. *Education Week, 25(35).*

Zenisky, A. L., & Sireci, S. G. (2002). Technological innovations in large-scale assessment, *Applied Measurement in Education, 15*, 337-362.

APPENDIX

Director's Letter

May 4, 2009

Dear Director of Schools,

As a student at East Tennessee State University, I am currently involved in my dissertation phase of the Educational Leadership and Policy Analysis doctoral program. My dissertation, *Correlation Between the TCAP Test and ThinkLink Learning's Predictive Assessment Series Test in Reading, Math, and Science in a Tennessee School System*, is to determine if a correlation exists between the Predictive Assessment Series (PAS) Test, marketed by Discovery Education, and the Tennessee Comprehensive Assessment Program (TCAP) Achievement Test in reading, math, and science for grade 4, grade 6, and grade 8.

I am seeking permission to access fourth, sixth, and eighth grade reading, math, and science scale scores from the 2009 TCAP and ThinkLink Learning PAS tests. The scores will be assigned a random number to prevent the identification of any student.

Thank you for your time and response to this request. If you have any questions, please feel free to contact me at xxxxx or by email at xxxxx. The results of this study will be available to you upon your request.

Sincerely,

Jared Day
Program Assistant
Xxxxx Elementary

Permission is granted for Jared Day to utilize fourth, sixth, and eighth grade TCAP and ThinkLink Learning PAS scores of students who were tested in xxxxxxx system.

________________________________ ______________

Signature Date

VITA

JARED E. DAY

Personal Data:	Date of Birth: April 12, 1977
Place of Birth:	Bristol, Tennessee
Marital Status:	Single
Education:	Public Schools, Sullivan County, Tennessee
	East Tennessee State University, Johnson City, Tennessee; History, B.A.; 1999
	East Tennessee State University, Johnson City, Tennessee; Curriculum & Instruction, M.A.T; 2002
	East Tennessee State University, Johnson City, Tennessee; Educational Leadership, Ed.D.; 2011
Professional Experience:	Teacher, Bristol TN City Schools, Bristol, Tennessee; 2003-2006
	Program Assistant, Bristol TN City Schools, Bristol, Tennessee; 2007-Present