RELATIONSHIP BETWEEN COLLEGE READINESS,

OKLAHOMA STATE TESTING PROGRAM, AND EXPLORE

By

Rick Martin

Bachelor of Science in Education
East Central University
Ada, Oklahoma
1984

Master of Education
University of Central Oklahoma
Edmond, Oklahoma
1988

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
Doctor of Education
July, 2010

RELATIONSHIP BETWEEN COLLEGE READINESS,

OKLAHOMA STATE TESTING PROGRAM, AND EXPLORE

Dissertation Approved:

Dr. Ken Stern
Dissertation Advisor

Dr. Ed Harris

Dr. Janice Miller

Dr. Mike Woods

Dr. A. Gordon Emslie
Dean of the Graduate College

ACKNOWLEDGEMENTS

I would like to extend my appreciation to all who assisted me with time, support, encouragement, and shared knowledge while I worked on my degree. I greatly appreciate the staff at the schools that provided the data to make the study possible.

I particularly want to thank my committee members who provided feedback and direction when needed. Thanks to Dr. Harris and Dr. Miller for filling in on my committee after previous committee members left the University. A special thanks to Dr. Stern for continuing to serve as my advisor following his retirement from the University. I greatly appreciate the time and expertise Dr. Miller shared on statistics in assisting me in data analyses. Thank you Dr. Woods for giving your time to serve as an outside member and the contributions you made to improve my work.

Special thanks must go to my family, especially my wife Debbie, as she took care of our children while I was away many hours working on my degree. Thanks to my children, Luke and Sara for understanding while I was away from home for so many hours. I also want to thank my oldest daughter, DeAndra, for helping her Mom with Luke and Sara and providing emotional support to both of us.

I appreciate the relationship I developed with my cohort while working on the program. I really appreciate Paul for being able to work together completing many classes along the way and sharing stories of our coaching backgrounds.

Finally, I would like to thank the faculty of Oklahoma State University's graduate school who were of great assistance during my years of study.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

Figure

CHAPTER I

INTRODUCTION

This study examined a part of Oklahoma's system of academic accountability for the requirements of the No Child Left Behind Act (NCLB) of 2001, using third-seventh grade math and reading scores from the Oklahoma State Testing Program (OSTP) and their relationship to the EXPLORE (grade 8) test. According to ACT (2008), EXPLORE is a part of ACT's Educational Planning and Assessment System (EPAS) designed to help students increase their academic readiness for college. EXPLORE measures student achievement in English, mathematics, reading, and science.

NCLB, a reauthorization of the Elementary and Secondary Education Act (ESEA), first enacted in 1965 and last reauthorized in 1994, required annual testing beginning in the 2005-06 school year in third-eighth grade in reading and mathematics and by 2007-08 in science at least once in elementary, middle, and high school. These tests had to be aligned with each state's academic standards. A sample of fourth and eighth graders in each state had to participate in the National Assessment of Educational Progress (NAEP) testing program in reading and math every other year to provide a comparison for state results. All students must reach a "proficient" level on state tests by the 2013-14 school year. Every school is required to meet state "adequate yearly progress" (AYP) targets toward this goal (based on a legislative formula spelled out in the law) for both their student population as a whole and for certain demographic subgroups.

NCLB prescribed penalties for schools receiving Title I funding failing to meet the target for two consecutive years. Such schools must be provided technical assistance and students must be offered a choice of other public schools to attend. Schools failing to achieve AYP for three consecutive years must also offer students supplemental educational services, including private tutoring. Schools failing to meet AYP beyond this point are subject to outside corrective measures, which could include governance changes (Education Week, 2004).

Carey (2006) contends NCLB gives states wide discretion to define what students must learn, how that knowledge should be tested, and what test scores constitute "proficiency." He accuses many states of taking advantage of the autonomy to make their test scores look much better than they really are and to minimize the number of schools facing scrutiny under NCLB.

Cronin, Dahlin, Adkins, & Kingsbury (2007) state that NCLB offered many promises, one of the most important being a pledge to "Mr. and Mrs. Smith" that they would get an annual snapshot of how their "Susie" is doing in school. Their fourth-grader lives in suburban Detroit, and her parents get word that she has met the proficiency standards on Michigan's state test in reading and math. The Smiths understandably take this as good news; their daughter must be "on grade level" and on track to do well in school, maybe even go to college.

> Unfortunately, there's a lot that Mr. and Mrs. Smith don't know. They don't know that Michigan sets its "proficiency passing score"—the score a student must attain in order to pass the test—among the lowest in the land. So Susie must be "proficient" in math in the eyes of Michigan education bureaucrats but still could have scored worse than five-sixths of the other fourth-graders in the country. Susie's parents and teaches also don't know that Michigan has set the bar particularly low for younger students, such that Susie is likely to fail the state test by the time she gets to sixth grade—and certainly when she reaches eighth grade—even if she makes regular progress every year. And they also don't know that "proficiency" on Michigan's state tests has little meaning outside the Wolverine State's borders; if Susie lived in California or Massachusetts or South Carolina, she would have missed the "proficiency" cut-off by a mile. Mr. and Mrs. Smith know that little Susie is "proficient." What they don't know is that "proficient" doesn't mean much. (Cronin et al., 2007, p. 2)

Carey (2006) created a state ranking system called "The Pangloss Index," after the character in Voltaire's Candide:

> Dr. Pangloss was an inveterate optimist, a man who insisted, in the face of all evidence to the contrary, that we live in the best of all possible worlds. Far too many states are using their discretion under NCLB to follow Pangloss' lead. (pp. 4-5)

Carey (2006) further contends that this has created large state-to-state variation in the percentage of students who are deemed "proficient;" Mississippi reports the highest level of fourth-graders in the nation proficient in reading (89%) compared to South Carolina, the lowest level of fourth graders proficient in reading, with 35% of their fourth-graders reaching the proficiency level. He questions if Mississippi is the best in the nation at teaching elementary students to read since they rank next to last on the NAEP results in fourth grade reading with only 18% of their students reaching the proficient level. By contrast, Massachusetts has the highest fourth grade NAEP reading scores in the nation, yet ranks fifth from the bottom based on the March, 2006 report. "State and NAEP assessments don't cover exactly the same content, so comparisons between the two aren't totally precise. But these kinds of through-the-looking-glass results leave little doubt that states like Mississippi have set academic standards exceptionally low" (Carey, 2006, p. 7). Carey (2006) indicated that in 2004-05, Oklahoma reported 83% (ninth highest in the nation) of their fourth grade students as proficient or advanced in reading on their state test, while the level reported on the 2005 NAEP as reaching proficient or advanced was 25%.

The concern over differences in test scores by fourth and eighth-grade students in reading and mathematics on the Oklahoma State Testing Program (OSTP), compared with that seen on a representative sample of students on reading and math assessments on the NAEP, led the Oklahoma legislature to create the Achieving Classroom Excellence (ACE) II Task Force to study five issues:

1. Comparison of the Priority Academic Student Skills with other states' curricular standards, primarily states that score highest on the NAEP;

2. Alignment of the Priority Academic Student Skills with the NAEP standards;

3. Feasibility of realigning the state performance level standards to NAEP performance level standards;

4. Differences in achievement levels among states based on exclusion rates on the NAEP; and

5. Feasibility of aligning the cut scores on state-mandated tests to NAEP cut scores. (Oklahoma, 2007)

(More detailed results of the findings are presented in Chapter II.)

Cronin et al. (2007) contend that one cannot answer questions about states' cut scores being too high or low and if they are internally consistent by examining academic standards alone. A state may have strong standards yet its test is easy to pass. It could have poor standards, yet expect high performance of its test-takers. It might have standards that are carefully aligned from one grade to the next, yet do poorly in setting cut scores.

> To examine cut scores carefully, you need a yardstick external to the state itself, something solid and reliable that state-specific results can be compared with. The most commonly used measuring stick is the NAEP, yet NAEP is a less-than- perfect benchmarking tool. (p. 3)

NAEP is designed to report scores of large groups of American students such as by state, and subgroups of gender, race, and ethnicity. Prior to NCLB, NAEP was voluntary for schools and states, but NCLB requires states to participate in biennial fourth-and eighth-grade reading and math main NAEP assessment to receive Title I funds. Since scores are not reported, critics of NAEP question students' motivation to perform their best during NAEP testing. On state tests, students receive their individual scores, and encouragement and support from parents, teachers, and administrators (Yeager, 2007).

NAEP exclusion criteria and rates vary greatly across states potentially affecting scores. For example, each state defines who its English Language Learner (ELL) students are, resulting in some states including certain students in this category while others exclude them, thus distorting the comparison (Yeager, 2007).

Yeager (2007) lists another limitation of NAEP; that of providing a snapshot data of achievement at a particular point in time for fourth and eighth grade students. The same students are not tracked over time, making it difficult to determine the outcomes of education policies and reforms or to measure individual school performance.

Loveless (2007) found that benchmarks used in scoring the NAEP are set too high, causing inordinately large numbers of students to be classified as less than proficient in math and reading. He based this on a study by (Phillips, 2007) at the American Institutes for research that correlated The Trends in International Mathematics and Science Survey Test (TIMSS), which assesses eighth graders' knowledge of mathematics, to the NAEP tests. In this correlation, students from 46 countries were scored on the NAEP test, 19 nations scored below basic, 22 scored basic, and 5 scored proficient. The United States scored basic on this analysis. Singapore, the highest performing nation in the world in mathematics, would have 73% of their students meeting the NAEP proficiency standard followed by Hong Kong, with 66% proficient, South Korea with 65%, Chinese Taipel with 61%, Japan with 57%, and Belgium (Flemish) with 40%. "If 25% to 50% of the students in the top-scoring countries in the world fail to meet an American standard of proficiency, one wonders how realistic that standard is as a universal expectation”(Loveless, p. 13).

Several other prestigious organizations reviewed the NAEP achievement levels and questioned their validity, including the Government Accountability Office, the National Academy of Sciences and the National Academy of Education. The latter report concluded that NAEP cut scores are too high and linked the flaw to the test’s weak content:

> The possibility that the cut scores are systematically too high is consistent with the findings from the panel’s content-expert studies in reading and mathematics, which showed that because there were no advanced items to measure the content of the descriptions, the experts moved higher and higher on the score scale in search of such an item. (Loveless, 2007, p. 13)

In 1996, CTB-McGraw Hill, an educational publishing and test development company, developed the Bookmarking cut score setting method. This method is now used by 28 states, including Oklahoma, and by NAEP. In this method, judges review the test and discuss performance categories. The test publisher ranks test questions from easiest to most difficult based on students’ past performance on the item. The judge’s note or “bookmark” where they think the performance categories should lie along the continuum of test items. This process continues for three rounds with cut scores based on the median value of the judges’ decisions. During the third round, judges are given data showing the affect of potential cut scores for students taking the test (Rotherham, 2006).

Rotherham (2006) advises that academic standards and assessments begin with the development of the former followed by the development of student assessments and then by setting cut scores. Cut scores set the standard of passing or meeting of the proficiency level:

> Just as academic standards are ultimately the result of professional judgment rather than absolute truth, there is no "right" way to set cut scores, and different methods have various strengths and weaknesses. The problem is that, though passionate feelings abound, there is no source of agreement about what, for instance, a fifth-grader should know and be able to do in mathematics or what sort of text they (sic) should be able to comprehend. (pp. 3-4)

Statement of the Problem

NCLB requires each state to establish challenging academic standards for all students in reading and math and to test students annually to see if they are reaching these standards. Districts and schools within the districts not achieving the state's required level are required to provide reforms that get tougher the longer there is failure to reach the goal. The bar is raised every year until 2014 when all students are required to reach the "proficiency" level as defined by each state (Carey, 2007). Critics of NCLB (Kahlenberg, 2008; Rothstein, Jacobsen & Wilder, 2006; and Linn, 2006) contend that no standard can be challenging to typical and advanced students and achievable by below-average students.

NCLB gives states broad discretion to determine what students must learn, how that knowledge should be tested, and what tests and scores constitute “proficiency.” Critics (Carey, 2006; Carey, 2007; Cronin et al., 2007) contend many states have taken advantage of this autonomy to make their educational performance look much better than it really is. In March 2006, states submitted the latest in a series of annual reports to the U.S. Department of Education describing their progress under NCLB. Several states used their standard-setting flexibility to inflate the progress achieved by their schools, thus limiting the number of schools facing penalties under the law. Some states claimed that 80% to 90% of their students were proficient in reading and math, even though NAEP put the figure at 30% or below (Carey, 2006).

External validity is necessary for purposes of accountability. Such validity is the extent test performance is related to another valued, independent, and direct measure of which the test is designed to assess (Lally, 2000). One form of external validity is predictive validity, a measure of whether achievement on one assessment can predict achievement on some future assessment of the same content and skills (Sattler, 1992). NAEP has been the predominant instrument to determine the external validity of state tests. Since one of the limitations of NAEP is that it does not track individual students, and looks only at scores during grades 4 and 8, it is difficult to use in tracking the outcomes of education reforms and policies (Yeager, 2007). Loveless (2007) also offered evidence that the high cut score standards set on NAEP made it an unrealistic instrument to measure progress and gauge state tests. For the purposes of this study, predictive validity will be used to determine the relationship between the OSTP and the EXPLORE test. The latter, a part of ACT’s Educational and Planning Assessment system, is typically given in the eighth or ninth grade and provides information for college readiness benchmark scores (ACT, 2008). The EXPLORE also gives the percent of students scoring at or below your score, describes the skills and knowledge students already have, and gives ideas for improving even more in different subject areas.

Research questions generated from the problem statement address the two major areas of OSTP and predictive validity. Presented first are two predictive validity questions followed by two OSTP questions.

Predictive Validity Research Questions:

1. Are OPI scores in the OSTP third-seventh grade math and reading statistically significantly related to performance in math and reading on the eighth grade EXPLORE test?
2. Do the OPI scores and performance levels on the OSTP third-seventh grade math and reading assessment predict national percentile ranks and college readiness as defined by the EXPLORE test?

OSTP Research Questions:

1. Does OSTP performance at each grade (3-7) and subject (math and reading) follow a normal distribution?

If the performance follows a normal distribution, it will offer evidence that the individual differences in students' ability make it difficult to create a test that is challenging to the most gifted students, yet attainable by lower performing students.

2. Do student performance levels on OSTP math and reading remain consistent from year-to-year (third-seventh grade)?

A comparison of performance levels was used to evaluate if remediation efforts have improved student achievement. Particular focus was on students who scored limited knowledge and unsatisfactory to determine if Oklahoma schools were moving towards the NCLB requirement of all student attaining proficiency by 2014.

Purpose of the Study

The study had two purposes. The first purpose was to define “proficiency levels” (advanced, satisfactory, limited knowledge, and unsatisfactory) and OPI scores on the third-seventh grade math and reading OSTP in a manner that is meaningful to parents, educators, and policy makers. Defining OPI scores and proficiency levels was accomplished by determining a relationship between student achievement on the OSTP and the EXPLORE test in math and reading. The relationship allowed the researcher to estimate national percentile ranks and college readiness based on the relationship to the EXPLORE test. The second purpose examined if the NCLB requirement of all students attaining “proficiency” by 2014 is realistic. The study used two methods to evaluate this proposition. The study examined the distribution of OPI scores within each grade and subject. A normal distribution of scores would offer evidence that the known variations in student ability make it unlikely a test that was challenging to average and above average students would be attainable by lower performing students. Finally, the study tracked students’ performance levels from third to seventh grade in math and reading. Tracking student performance levels from year to year examined the effectiveness of current remediation practices to determine if student achievement is improving, particularly for students who fell in the unsatisfactory and limited knowledge proficiency levels.

Significance of the Study

The study can contribute to the understanding of the results of the OSTP by estimating students' national percentile rank and college readiness level. Additionally, this research can contribute to the literature on the required testing of NCLB, especially Oklahoma's testing. NCLB requires annual testing in third-eight grades in mathematics and reading. States were allowed to design their tests and set their standards. Carey (2006) contends many states have taken advantage of the autonomy to make their test scores appear much better than they really are and to minimize the number of schools facing scrutiny under NCLB. Therefore, parents may not be getting a true picture of their children's progress and, ultimately, their preparedness for success in college and the workforce.

Concerns that the OSTP exams lacked rigor led to the Oklahoma Business and Education Coalition funding a study on Oklahoma's accountability system. Brewer & Killeen (2009) indicated Oklahoma public school students may be misled to think they are doing better academically than they actually are. They accuse Oklahoma of setting the bar too low on state tests and support their claim based on the gap between the percent of students deemed proficient on state tests and the NAEP, which is particularly large in Oklahoma compared to other states. The Oklahoma State Board of Education issued a statement criticizing the report as "error-filled, based on anonymous sources, and lacking data to prove that proposed changes would be effective" (Study, 2009).

Following this report, legislation (SB 1111) was sponsored to restructure the system to improve oversight of student testing, including strengthening the Office of Accountability, created in the school reform act of 1990, to function independently of the Oklahoma State Department of Education (OSDE). The Office of Accountability would handle both the education data system and the testing-accountability functions (Study, 2009). SB 1111 passed the Oklahoma State House and Senate but was later vetoed by the governor (Hoberock, 2009). In the summer of 2009, the OSDE increased the difficulty of the OSTP third-eighth grade math and reading exams. The purposes were to:

1. Increase rigor by raising standards for Grades 3-8 student achievement on the OCCT as a means to be more competitive at the national and international levels,
2. Vertically align proficiency expectations for students on the OCCT test for Grades 3-8,
3. Align student expectations on the OCCT more closely with student expectations for the National Assessment of Educational Progress (NAEP) (Defehr, 2009, p. 12).

Many important decisions facing the accountability systems of Oklahoma and other states are based on comparisons of state tests to NAEP. Researchers and academicians have offered extensive criticism and limitations of using the NAEP proficiency level as a benchmark for NCLB (Yeager, 2007; Loveless, 2007).

The results of this study may be beneficial to the Oklahoma legislature and policymakers who strive to set benchmarks for “proficiency” as required by NCLB.

Previously, no research beyond comparison to NAEP has been done in this area for Oklahoma schools. This study offers a comparison to another recognized test that offers a predictive value of future academic success. Comparisons to NAEP indicate that the bar is set too low, yet no direction in research is given to where the bar needs to be set. This research provides guidance on where the bar needs to be set based on college readiness and national percentile rankings. The study also offers evidence the NCLB requirement of all students attaining proficiency by 2014 is unrealistic.

Theoretical Frameworks

The study operated under two theoretical frameworks. The first framework is based on predictive validity. deKlerk (2008) describe predictive validity as the transferability of test scores to “the real world.” Test scores should theoretically transfer to real-world success, such as completing college. The study used the concept of predictive validity to predict EXPLORE scores from OPI scores. Oklahoma PASS and EPAS standards are consistent, thus congruent validity between the two assessments, OSTP and EXPLORE should exist. Furthermore, performance on the OSTP assessments should have predictive validity for performance on the EXPLORE tests due to this congruency. The EXPLORE test has tracked student progress from the eighth grade through college to determine benchmarks that predict success in college courses. The EXPLORE also provides national percentile ranks and college readiness benchmarks that will provide guidance to policy makers in setting standards of proficiency as required by the testing under NCLB.

The congruency between standards of the PASS and EPAS was supported by The Oklahoma State Regents for Higher Education Student Preparation Team. Five curriculum committees were formed in the summer of 2004 to compare the Educational Planning and Assessment System-EXPLORE/PLAN/ACT (EPAS) Standards for Transition and the Oklahoma Priority Academic Student Skills (PASS) in their curriculum areas. The goal was to provide information for classroom teachers and administrators regarding the match between EPAS and PASS standards/objectives for instructional, curriculum alignment and district planning purposes. The findings from the English Curriculum Committee were that consistency existed between the EPAS English Standards for Transition and the PASS Language Arts objectives. The Social Studies Curriculum Committee believed there was consistency between the EPAS Reading Standards for Transition and the PASS Social Studies Process Skills. Also, the Mathematics Curriculum Committee believed there was consistency between the EPAS Mathematics Standards for Transition and the PASS Mathematics objectives (Oklahoma Educational Planning and Assessment System [OEPAS], 2004).

The NCLB Act requires all students to be proficient by 2014, but it does not define proficiency; rather, it refers to NAEP. The law says proficiency must be "challenging," a term taken from NAEP. Rothstein, Jacobson & Wilder (2006) state:

> By ignoring the inevitable and natural variation amongst individuals, the conceptual basis of NCLB is deeply flawed: no goal can simultaneously be challenging to and achievable by all students across the entire achievement distribution. A standard can either be a minimal standard which presents no challenge to typical and advanced students, or it can be a challenging standard which is unachievable by most below-average students. No standard can serve

> both purposes—this is why we call “proficiency for all” an oxymoron-but this is what NCLB requires. (p. 2)

The second theoretical framework is based on normal probability distribution. Normal distribution, often called the bell curve, occurs frequently in statistics, economics, and the natural and social sciences. A normal distribution is not typical for CRT tests, such as the OSTP. “If the score distribution for a CRT did look like a normal distribution, depending on the location of the passing score, it would probably suggest that only a small proportion of the examines displayed mastery” (Professional Testing, 2006, p. 2). Criterion-referenced tests commonly have a simple classification decision reported as pass/fail (Professional Testing, 2006). The OSTP has four categories of student performance (advanced, satisfactory, limited knowledge, and unsatisfactory) allowing the tests to differentiate between student abilities, although the test may not be sensitive enough to student differences to create a normal distribution of scores. The study examines the distribution of scores to offer evidence the known variability in human ability makes it unlikely a test can be developed that is challenging to average and above average student and attainable by lower performing students.

Definition of Terms

Accountability-NCLB requires each state to develop and implement a single, statewide accountability system that will be effective in ensuring that all districts and schools make adequate yearly progress and will hold accountable those who do not.

ACT Test-The ACT test is a national college admission and placement examination that is curriculum-based. It is neither an aptitude nor an IQ test. Instead, the questions are directly related to what students have learned in high school courses in English, mathematics, reading, and science.

Adequate Yearly Progress (AYP)-Under NCLB, each state must establish a definition of AYP that each district and school are expected to meet. States must establish annual benchmarks to measure progress of the total population and nine specific subgroups so that all students reach proficient performance in 12 years. Benchmarks for AYP are based on students' scores on the 2001-2002 state assessments.

Cut Scores-Cut scores are selected points on the score scale of a test. Cut scores are used to determine whether a particular test score is sufficient for some purpose. Setting cut scores involves policymakers, educators, measurement professionals, and others in a multi-state, judgmental process (Yeager, 2007).

Educational Planning and Assessment System (EPAS)-This system was designed by ACT as an integrated series of assessment and career planning programs (EXPLORE grades 8 and 9), PLAN (grade 10), and ACT (grades 11 and 12) to help students increase their academic readiness for college.

Elementary and Secondary Education ACT (ESEA)-This legislation, enacted by the United States Congress in 1965 to improve educational opportunities for poor children was based on the principle that children from lower socio-economic environments required more education services than children from affluent environments (Association for Educational Communications and Technology [AECT], 2001).

English Language Learner-A student whose home language is one other than English and is learning English as an additional language.

EXPLORE – A test taken during the eighth or ninth grade with content closely tied to that of the achievement in the ACT.

Improving America's Schools (IASA)-IASA is an amendment to the ESEA of 1965, recognizing that the achievement gap between disadvantaged children and others had decreased but was still sizable. It required the allocation of at least an additional $750,000,000 over baseline each fiscal year (1996-1999) and thereby increasing the percentage of eligible children served in each fiscal year with the intent of serving all eligible children by fiscal year 2004 (United States Department of Education [USDOE], 1994).

National Assessment of Educational Progress (NAEP)-NAEP is the only nationally representative and continuing assessment of what America's students know and can do in various subject areas. The assessments are conducted periodically in mathematics, reading, science, writing, the arts, civics, economics, geography, and U.S. History. NAEP, sometimes referred to as the "Nation's Report Card," provides results on subject-matter achievement, instructional experiences, and school environment for populations of students, but does not provide scores for individual students or schools.

No Child Left Behind (NCLB) -Federal Legislation enacted in 2001, it replaces the ESEA of 1965 and the IASA of 1994 and focuses on individual school success as measured by student achievement data. NCLB also provides severe consequences for schools whose students do not meet prescribed achievement levels.

Oklahoma State Testing Program (OSTP)-Also called The Oklahoma Core Curriculum Test (OCCT), is a criterion referenced test covering the Priority Academic Student Skills. Math and reading tests are given in grades (3-8) and End of Instruction (EOI) tests are given at the secondary level. Also grade 5 tests writing, and social studies; grade 7 tests geography; and grade 8 tests writing, science and U.S. History. The secondary level requires EOI testing in algebra I, algebra II, biology I, English II, English III, geometry, and U.S. History. Oklahoma law requires, beginning with the 2009 ninth grade class, students to pass four of seven EOI tests, with algebra I and English II being two of the required tests to graduate.

Predictive validity-The degree to which the score on a test predicts the individual's score or performance in some other test or area.

Priority Academic Student Skills (PASS)-PASS serves as Oklahoma's specific school standards covering all areas of a student's academic growth. Oklahoma's PASS documents were developed by and for educators. The standards guide teachers and school leaders as they plan curriculum, instruction, and assessment for students.

Proficient-All students are expected to achieve at the proficient level under NCLB requirements. Proficiency is arbitrarily determined by each state. NCLB requires two levels of "high" achievement: proficient and advanced, and a third lower level of achievement: basic. States have the flexibility to give different names to these levels. Oklahoma's achievement levels are advanced, satisfactory, limited knowledge, and unsatisfactory.

Title I-The funding source for additional educational services under ESEA, IASA, and NCLB. Title I funds are based on a school district's poverty level as well as an individual school's poverty level.

The Trends in International Mathematics and Science Survey Test (TIMSS)-A national test that provides reliable and timely data on the mathematics and science achievement of U.S. fourth and eighth grade students compared to that of students in other countries. TIMSS data have been collected in 1995, 1999, 2003, and 2007 (National Center of Educational Statistics [NCES], 2009).

Summary

Chapter II is a review of the literature related to the NCLB legislation with particular emphasis on the accountability component, the history of standards based reform, particularly national and Oklahoma reform, and the testing response to both the legislation and the reform movement. Chapter III describes the study's population data, the instruments used to collect data, and the procedures for collecting and analyzing data based on the study's research questions. Chapter IV presents a statistical analysis of the quantitative data related to the research and a summary of the findings. Chapter V includes the summary, conclusions, and recommendations of the study.

Chapter II

REVIEW OF THE LITERATURE

This chapter begins by describing the political events that led up to No Child Left Behind (NCLB). Researchers', academicians', and politicians' philosophies are reviewed on the NCLB requirement that all students be proficient in math and reading by 2014 and the legislation's failure to define proficiency. NCLB allowed each state to determine the test and acceptable scores for determining proficiency. Analyses of literature discussing critics' accusations that many states have set the criteria for reaching proficiency too low to avoid the penalties for failing to reach Adequately Yearly Progress (AYP) each year is conducted. Research studies that have viewed the proficiency requirement of various states and the practicality of a single definition of proficiency for all students will be discussed from the views of researchers and academicians.

NCLB requires children to reach proficiency on challenging academic standards and assessment, paralleling the language of NAEP. Because of this relationship between the NCLB and the NAEP definition of proficiency, many politicians and advocates of education reform promote the proficiency definition of the National Assessment of Educational Progress (NAEP) for all states. Studies of NAEP are reviewed to determine researcher's views of the feasibility of using the NAEP definition of proficiency.

Studies on the findings of Oklahoma's accountability system are discussed. Since the study used the EXPLORE test to validate the current Oklahoma State Testing Program (OSTP) testing system, a history of the EXPLORE test is included. The EXPLORE test, a part of the ACT, is a predictor of college readiness. The history of the OSTP will be presented to allow readers an understanding of the many changes in Oklahoma's accountability system over time.

Two statistical benchmarks guide theoretical frameworks for the study. The first framework is based on predictive validity. The study used the concept of predictive validity to predict EXPLORE scores from OPI scores. Because the standards/objectives of the OSTP (PASS) and the standards/objectives of EXPLORE (EPAS) are consistent (OEPAS, 2004), congruent validity between the two assessments, OSTP and EXPLORE should exist. This predictive validity is essential to determine a benchmark for setting cut scores for performance levels that are meaningful on the OSTP. The study used predictive validity to estimate student performance as a national percentile rank and college readiness based on the relationship between OSTP and EXPLORE (national percentile ranks and college readiness standards are given on the EXPLORE test). The importance of college readiness is supported by Somerville and Yi (2002) who indicate the skill and knowledge employers require in the workplace look very much like those in higher education.

The second theoretical framework is based on the principal of normal probability distribution. Although criterion referenced test such as the OSTP typically do not result in a normal distribution of scores, the performance levels allow the exams to differentiate between student abilities. The study will evaluate the distribution of scores in order to provide evidence the known variability in human ability makes it unlikely a test can be developed that is challenging to average and above average students, yet attainable by lower performing students. This offers evidence that the NCLB requirement of all students attaining "proficient" by 2014 is unrealistic.

No Child Left Behind

The NCLB Act of 2001 is a part of the Elementary and Secondary Education Act (ESEA), originally passed by Congress in 1965 as part of Lyndon Johnson's "War on Poverty." This Act provided public schools with additional money to improve academic achievement levels of students from lower socio-economic backgrounds. Under Title I of ESEA, public schools receiving federal dollars had to monitor the progress of students through evaluation (Helfant, 2005).

Jorgensen & Hoffman (2003) contend that a movement towards standards-based education and evolution in achievement testing began with a 1983 report from the U.S. Department of Education, "*A Nation at Risk*" which cited concern about the poor quality of America's public schools. Jorgenson & Hoffman (2003) report that findings and recommendations of *A Nation at Risk* centered on four important aspects of the educational process: (1) content, (2) expectations, (3) time, and (4) teaching. Content had become diluted and was without central purpose. Expectations were deficient including declining homework, lack of required mathematics and science courses, increased enrollment in less demanding electives, and lack of challenge to student textbooks. American students were spending less time on schoolwork than prior generations and used classroom time and homework ineffectively. Students were not encouraged by schools to develop study skills to use time effectively or the willingness to spend more time on homework. The report also indicated that the field of teaching was not attracting enough academically able students and that teacher preparation programs needed substantial improvement.

Fuller, Gesicki, Kang & Wright (2006) contend prior to the 1990's, state testing did not meet minimal criteria for yielding valid and reliable data on student achievement over time. To accomplish this, assessment systems needed to offer annual scores, equate scores to make them comparable over time, and test comparable groups of students over time. The 1990 National Governors Association had reform designers that created the rise of "systemic reform" based on a model of organizational change that aligned assessments to transparent standards.

Jorgenson & Hoffman (2003) explain the movement towards standards-based education and assessment that began with *A Nation at Risk*, became even more prominent on the national scene with the passage of Improving America's Schools Act of 1994 (IASA), a reauthorization the ESEA. Another important 1994 law, Goals 2000: Educate America Act, focused on the needs of all students, not just the disadvantaged students. According to Jorgenson & Hoffman (2003), the IASA amendments required all states to have:

1. Content and performance standards;
2. Assessments aligned with those standards in one grade of each of three spans: 3-5, 6-9, and 10-12; and

3. An accountability system to identify the schools that were not helping all students perform as expected on those assessments. (p. 4)

Kahlenberg (2008) states that the transformation of our nation from a manufacturing economy to a new knowledge based economy was the basis for early advocates of standards based reform. However, the advocates "recognized they were asking schools to do something they had never done before—educate all students to high levels" (p. 3). He explains:

> In the past, under a manufacturing economy, it was acceptable to hold a small group of students to high standards and let the majority slip by. But in the new, knowledge-based economy, educators needed genuinely to educate far more students, and advocates realized that in order to reach new performance standards, more funding would be required. The essential bargain, then, was more funding for greater accountability. (p. 6)

Kahlenberg (2008) contends that even though NCLB was an outgrowth of the standards-based reform movement it departed from the early ideas of the movement. He cites flaws in funding, standards, testing, and accountability scheme along with the limitations in the transfer provision for students in low performing schools.

NCLB requires states to define proficiency and mandates all students meet the performance standard in math and reading by 2014. Kahlenberg (2008) states:

> Most serious educators believe this goal to be a fantasy because it denies the reality of human variability. No society throughout history has ever achieved 100% proficiency in education. To suggest that all students, including those

> who are severely disabled, will reach a meaningful standard of proficiency is a nice political slogan, but it is absurd to punish schools, principals, and teachers—and ultimately, students and their families—for failing to reach an impossible goal. A single performance standard that is impossibly high for certain special education students to meet may at the same time be too low for the vast majority of students who will not be challenged sufficiently. The problem, therefore, is not that the performance standard is set too high or too low—it is that a single standard will be both. Finally, a single performance standard necessarily will lead to an emphasis on helping children who are on the cusp of being proficient. Teachers will concentrate on the students who are almost proficient and ignore those judged to be well above the proficiency mark, and those who are so far below it that there is little chance of helping them to reach the mark in time for the test. (p. 7)

Jorgenson & Hoffman (2003) contend that the testing industry moved towards a standard based high-stakes assessment as a necessary component of standards-based reform. States instituted content and performance standards, collected longitudinal data, and demanded high-quality, custom-designed, and error-free testing materials. The NCLB Act of 2001 reauthorized ESEA in dramatic ways; "This landmark event certainly punctuated the power of assessment in the lives of students, teachers, parents, and others with deep investments in the American educational system" (Jorgenson & Hoffman, p. 6).

NCLB requires states to build assessment systems that monitor all students against high standards. States are required to assess third through eighth grade students each year in reading and math. Schools have until the 2013-2014 school year to have all students perform at proficient levels on statewide tests. States are required to develop and implement a single, statewide accountability system that will be effective in ensuring that all districts and schools make adequate yearly progress and hold accountable those schools that do not (United States Department of Education [USDOE], 2002). Furthermore, schools must show that specific subgroups of students, economically disadvantaged, a variety of ethnic groups, special education students, and limited English proficient students, are also making progress and at least 95% of each subgroup has taken the prescribed tests. One subgroup failing to meet the 95% standard or not making AYP as defined by state benchmarks, results in the entire school considered in need of improvement (No Child Left Behind [NCLB], 2001; USDOE, 2002).

NCLB allowed states wide discretion to define what students should learn, how tests were designed, and what scores are considered "proficient." Carey (2006) contends many states have taken advantage of this autonomy to make their progress appear inflated.

NCLB State Accountability Systems

Cronin, Dahlin, Adkins & Kingsbury, (2007) studied 26 states to answer three research questions and three sub questions.

1. How consistent are the various states' expectations for "proficiency" in reading and mathematics?

2. Is there evidence that states' expectations for "proficiency" have changed over time, in particular during the years immediately following the enactment of NCLB?

    A. If so, have they become more or less difficult to meet?

    B. Is it getting easier or harder to pass state tests?

3. How closely are a state's proficiency standards calibrated across grades? Are the standards in earlier grades equivalent in difficulty to proficiency standards in later grades (taking into account the obvious differences in subject content and children's development from grade to grade)?

Data from schools where students took both the state testing and the Measures of Academic Progress (MAP) developed by the Northwest Evaluation Association (NWEA) assessment were analyzed to estimate proficiency cut scores (for NCLB purposes) in 26 states. This analysis indicated that state tests varied greatly in their difficulty. In third grade reading, scores ranged from the 7% (Colorado) to the 61st percentile (California) on the NWEA scale. In 24 of the 26 states examined, the third grade proficiency cut scores were below the 50th MAP percentile, with 19 of the 26 cut scores falling in the 20% to 40% range (Cronin et al., 2007).

The second question of the Cronin et al. (2007) study had two parts. Is there evidence that states' expectations for proficiency have changed over time? If so, have state proficiency cut scores become more or less difficult? Proficiency cut score estimates were generated at two points in time for 19 states. The first estimate was in the spring 2002 and the second in the spring 2005. Nineteen states were studied with eight revising their scales or adjusting their proficiency cut scores. Five adopted new measurement scales, while the other three changed the cut score on their existing scale in at least one grade. The remaining 11 states announced no changes to their proficiency cut scores during the period of the study. Most state tests have not changed the test difficulty. Observed changes were typically in the direction of less difficulty, particularly in states with the highest standards.

The third question of the Cronin et al. (2007) study also had two parts. "How closely are proficiency standards calibrated across grades? Are the standards in earlier grades equal in difficulty to proficiency standards in later grades?" The findings indicated that reading and math tests in upper grades were consistently more difficult to pass than those in earlier grades (even after taking into account obvious differences in student development and curriculum content). Another finding was that mathematics tests were consistently more difficult to pass than the reading tests. These researchers concluded about NCLB:

> There is no common understanding of what "proficiency" means. Its definition varies from state to state, from year to year, from subject to subject, and from grade level to grade level. This suggests that the goal of achieving "100% proficiency" has no coherent meaning, either. Indeed, we run the risk that children in many states may be nominally proficient, but still lacking the education needed to be successful on a shrinking, flattening, and highly competitive planet. (p. 7)

The (National Center of Educational Statistics [NCES], 2007) employed a methodology for mapping state standards onto the national NAEP scale based on combining data from the National Longitudinal School-Level State Assessment Score Database with data from NAEP. State test data from the 2004-05 academic year and NAEP 2005 results were used for comparison purposes.

NCES (2007) analyzed 32 states on fourth grade reading. The estimated NAEP cut-score equivalents ranged from 161 (Mississippi) to 234 (Massachusetts). Twenty-three states' estimated state cut-scores fell below the NAEP basic cut-score of 208. Ten states' estimated state cut-scores fell in the NAEP basic cut-score range of 208-237 while none of the states fell in the NAEP proficiency cut-score range. Oklahoma had an estimated fourth grade reading score of 182 which ranked 28 of the 32 states analyzed on their stringency level.

Thirty-four states were used in the eighth grade reading analysis. Scores ranged from 217 (North Carolina) to 278 (Wyoming). Nine estimated cut scores fell below the NAEP basic cut score with the other 25 falling in the basic range. Oklahoma fell in the NAEP basic range with an estimated cut score of 244 with 243 as the cut-off for eighth grade reading basic level. Oklahoma ranked 20 out of the 34 states on the stringency of cut scores (NCES, 2007).

Thirty-three states were analyzed on fourth grade mathematics. NAEP score equivalents ranged from 200 (Tennessee) to 265 (Massachusetts). Six of the estimated cut-scores fell below the NAEP basic cut-score of 214, 25 states fell in the NAEP basic cut-score range of 214-249, while two states had fourth grade state math scores above 249, the NAEP proficiency cut score. Oklahoma fell in the NAEP basic range and ranked 23 of the 34 states in the stringency of their state test in fourth grade mathematics (NCES, 2007).

NCES (2007) analyzed 36 states on eighth grade mathematics. The estimated NAEP score equivalents ranged from 230 (Tennessee) to 310 (Missouri). Eight states fell below the NAEP basic cut score of 262. Twenty-five states fell in the NAEP basic cut score range of 262-298 while three had state proficiency scores in the NAEP proficiency range. Oklahoma fell below the NAEP basic cut score with an estimate scale score of 258 and ranked 30 out of 36 states studied in the stringency of their state test in eighth grade mathematics.

Carey (2006) created a ranking of states based on 11 measures contained in the March 2006 report that states submitted to the U.S. Department of Education detailing their progress under NCLB. The measures included student proficiency rates in elementary, middle, and high school; the percent of schools and districts making adequate yearly progress; high school graduation and dropout rates; school violence ratings; teacher and paraprofessional qualifications; and teacher access to high-qualify professional development. "For every measure, the pattern was the same: a significant number of states used their standard-setting flexibility to inflate the progress that their schools were making and thus minimize the number of schools facing scrutiny under the law" (Carey, 2006, p. 3). The rankings were called the "Pangloss Index" after the character in Voltaire's Candide.

> Dr. Pangloss was an inveterate optimist, a man who insisted in the face of all evidence to the contrary, that we live in the best of all possible worlds. Far too many states are using their discretion under NCLB to follow Pangloss' lead." (Carey, pp. 4-5)

Carey (2006) based this claim on how some states have a much higher percent of students meeting their proficiency requirement on their state tests than what the results of NAEP indicate as proficient. Other areas that Carey (2006) accuses states of manipulating data include confidence intervals, large group size requirements before data are collected on subgroups of NCLB, defining highly qualified teachers, graduation and drop-out rates, and defining high quality professional development.

Fuller, Gesicki, Kang & Wright (2006) studied 12 diverse states to compare improvements on NAEP results and their own state tests from 1992-2005, wanting to see if NCLB had actually increased test scores measured by NAEP. Findings indicated that states estimated a much higher share of students reaching proficiency compared with the NAEP results. Students did make greater progress in math proficiency over this period on NAEP, but not as great as indicated by state tests results. Fuller et al. (2006) contend the gap does not stem simply from NCLB's incentive for states to set low cut scores for determining student proficiency and offer evidence that states have long claimed a higher percentage of students meeting the proficiency standard relative to NAEP results even before NCLB. The study found that three years following the enactment of NCLB, some states had maintained their apparent momentum in raising the percentage of fourth graders proficient in math, while reading performance leveled-off or slipped in several states, as gauged by state and NAEP exams. The NAEP of 2005 offered three years of data from the inception of NCLB, indicating reading achievement for fourth graders remained flat, with 31 percent of the nation's children proficient in 2002-2005. Eighth graders proficient or above in reading fell two percentage points. Fourth-graders' proficiency in math climbed between 2003 and 2005, while eighth grade math scores plateaued.

Fuller et al. (2006) claim states have changed their exam systems during the time frame studied. This results in a "jagged saw-tooth" trend line as scores typically fall following the test change. Fuller et al. (2006) explain:

> Factors that inflate results are temporarily suspended. That is, teachers don't know the test items to which they must teach, questions likely align with a new set of curricular domains and constructs, and the novel format of a new test may constrain student performance. (p. 6)

Fuller et al. (2006) further found that the mean share of fourth-graders proficient or above in reading was 68% based on state testing compared to 31% on NAEP results. Fourth grade math scores averaged 65% of fourth-graders proficient on state tests compared to just 30% proficient based on NAEP results. Oklahoma was one of the 12 states analyzed by the data since 1996. The historical gap between the state and the NAEP test averaged 49 percentage points in reading and 60 in math. One argument about the differences of the NAEP and the state test is that the NAEP is insufficiently sensitive to learning gains which are somehow uniquely detected by state tests based on state education officials creating tests that are differentially sensitive to improvements at the low level.

Some researchers urge caution when comparing year-to-year changes in student performance between state and NAEP test results, if the data series begins at a low or high point in the distribution of raw test scores. This caution is based on proficiency scores that may be set near either tail of the distribution; the comparative proportions of students who must cross over the cut-point score can vary considerably (Fuller et al., 2006). On the other hand,

> If the problem was simply rooted in where states set initial cut-points, we would still see a tighter correspondence between the movement of state and NAEP scores over time. But the fact that states scores are moving independently suggest either that inflation in state scores is occurring, or that the kind of learning tapped by state tests is largely missed by the NAEP assessment. The latter argument seems unlikely. (p. 18)

The Oklahoma State Senate created The Achieving Classroom Excellence II Task Force (ACE II) pursuant to SB 921 out of concern over the difference in test scores of fourth and eighth grade students in reading and mathematics between the OSTP and the NAEP assessment. Students performed significantly lower on the NAEP than did the same age students on the OSTP. This task force was charged with five issues.

The first was comparison of the Priority Academic Student Skills (PASS) with other states' curricular standards, primarily states that score highest on the NAEP. The recommendations were that the State Department of Education should work in partnership with an independent, third party, contractor, such as Achieve, Inc. to perform a comprehensive crosswalk of Oklahoma's PASS compared to other states' standards. This independent and comprehensive study would allow not only a comparative analysis to other states standards but would also be anchored by an analysis of state content and process standards to other national standards such as those seen in NAEP and the American Diploma Project (Oklahoma, 2007, p.7).

The second issue was alignment of the PASS with the NAEP standards. The committee recommended the State Board of Education adopt all of the recommendations of the Workgroup on Curriculum Alignment, Assessment and Cut Scores of the ACE I Steering Committee approved in April of 2007. Another recommendation included the addition of constructed response questions that would align with NAEP and other national examinations with out-of-state graders to grade the examination to yield a higher level of reliability and validity to score results. Further recommendations included additional funding for the costs of creating and grading constructive response questions and that all Oklahoma content and process standards at all grades be revised to achieve a high degree of alignment with national standards including NAEP (Oklahoma, 2007).

The third issue was the feasibility of realigning the state performance level standards to NAEP performance level standards. The recommendation urged the Oklahoma State Board of Education to adopt the following definitions of performance level descriptors when setting OSTP test cut scores and reporting results:

1. Advanced-The student demonstrates superior performance on challenging subject matter.
2. Proficient-The student demonstrates mastery of appropriate grade-level subject matter and that students are ready for the next grade, course, or level of education, as applicable.

3. Limited Knowledge-The student demonstrates partial mastery of the essential knowledge and skills at the appropriate grade level, course, or level of education as applicable.
4. Unsatisfactory-The student does not perform at least at the limited knowledge level (Oklahoma, 2007, p. 14).

Some ACE II Task Force members believed that the term "mastery of appropriate grade-level subject matter" should be a requirement for proficiency. All task force members agreed the descriptor should require that students were ready for the next grade or course.

The fourth issue, the difference in achievement levels among states based on exclusion rates on the NAEP, was addressed through a recommendation that Oklahoma's minimum subgroup number be reduced to no more than 30 students and the fifth and final item was the feasibility of aligning the cut scores on state-mandated tests to NAEP cut scores. The task force recommended that Oklahoma panels setting cut scores should have the same composition in membership as seen on NAEP, reflecting a broader scope of participants including more members of the business community and higher education. Further recommendations included making the cut scores more transparent in terms of determining student achievement levels. The report was presented to the Senate for further review. SB 1880 of the 2008 session would have extended the task force for six additional months for further study; however, the governor vetoed the bill.

Finn, Julian & Petrilli (2006), of the Thomas B. Fordham Foundation, studied all states' academic standards and compared them to their own findings in 2000. Three states stood out with perfect scores: California, Indiana, and Massachusetts. Common themes from the states appear: "If you want great standards, you can't leave the process to committees." "It takes strong visionary leadership and a willingness to fight (and win the curricular battles); at the same time, bipartisanship is essential" (p. 6). The importance of standards is their relationship to increasing student achievement. Finn et al. (2006) state several analyses suggest a link between strong state standards and gains on the NAEP:

1. Ten states made statistically significant progress in the percentage of students (or the percentage of poor and minority students) reaching proficiency in fourth grade reading on NAEP from 1998 to 2005. Nine of these 10 states received at least a C from Fordham for their English/language arts standards.
2. Five states achieved statistically significant gains on the science NAEP between 2000 and 2005 at both the fourth and eighth grade levels, and three of these had among the best sets of science standards in the nation, according to Fordham's reviewers.
3. The relationship is less clear in mathematics, though four of the six states that received honors grades from Fordham also posted statistically significant gains on the eighth grade NAEP from 2000 to 2005, either for the state as a whole or for their poor or minority students (Many other states made progress too, however). (p. 6)

The lack of relationship of quality math standards and improved NAEP scores is due to NAEP and most states having substantially aligned with the standards promoted by the National Council for Teachers of Mathematics. Because of this high correlation of standards between the states and NAEP, NAEP scores rise. Criticism is leveled at the math standards of NAEP and the states through a contention that America is moving away from the kind of solid mathematics practiced around the world (Finn et al., 2006).

Finn et al. (2006) admit critics of their reviews complained that the grades on academic standards bear no relationship to NAEP performance and the criticism is partly right. No relationship between Fordham's grades on state academic standards and the percentage of students who are proficient exists. However, Finn et al. (2006) state:

> But this is no surprise. It's well known that state NAEP scores are tied most directly to the state's demographics. One could fairly say that the goal of standards-based reform in general and NCLB in particular, is to *break* that link. Hence, what matters is whether any reform, including the adoption of rigorous standards, leads to progress over time. Viewed through that lens, the picture looks more promising. From 1998 to 2005, only seven states made statistically significant progress in the percentage of their students reaching proficiency in fourth grade reading, and just six states made such progress for their poor or minority students. All of these states except for one received at least a "C" from Fordham for their English/Language Arts standards. That's not iron-clad proof that good standards boost achievement, but it seems to

> indicate that really bad standards make it much less likely. Still, lots of states received a "C" or higher from us but did not make progress on NAEP. So having decent standards could be considered "necessary," but not sufficient. (pp. 13-14)

According to the rankings of Finn et al. (2006), Oklahoma received a C in English, C in math, F in science, B in U.S. History and a B in world history. This ranked Oklahoma tied for eleventh in the nation in their overall state standards across all subjects. In 2000, the same study ranked Oklahoma 21st in the nation on state standards across all subjects.

The United States Chamber of Commerce released a report titled "Leaders and Laggards" in 2007 that gave a state-by-state report card on educational effectiveness. (United States Chamber of Commerce [U.S. Chamber], 2007) graded each state for academic achievement, academic achievement of low-income and minority students, return on investment, truth in advertising about student proficiency, rigor of standards, postsecondary and workforce readiness, 21st century teaching force, flexibility in management, and policy and data quality.

Grades were issued for academic achievement based on the percent of student's proficient or above on fourth and eighth grade math and reading on the 2005 NAEP test. The 10 states with the highest percent of students proficient or above received the letter grade A, states ranking 11-20 received the letter grade B, states ranking 21-31 received the letter grade C, states ranking 32-41 received the letter grade D, while states ranking 42-51 (the District of Columbia was included in the report) received an F. Oklahoma received an F, ranking 42nd in the nation on the percent of students proficient or above on fourth and eighth grade math and reading.

U. S. Chamber (2007) also assigned Oklahoma the letter grade of F for truth in advertising about student proficiency. The authors calculated a grade for each state based on the difference between the percentages of students deemed both proficient by the state and by NAEP in 2005. States with large gaps did poorly while those with small gaps received higher scores.

U. S. Chamber (2007) issued grades for return on investment by using an index created by dividing the percentage of students scoring at or above the proficient level on fourth and eighth grade math and reading in 2003 by 2004 state expenditures. Adjustments were made for cost of living, poverty students, and special need students. Oklahoma received the letter grade of "C" in this category and ranked 27th in the nation.

Martin (2007) in a rebuttal to "Leaders and Laggards," provided evidence to support Finn et al.'s statement that, "It's well known that state NAEP scores are tied most directly to the state's demographics" (p.13). In Figure 1 Martin compared the relationship between academic letter grades given by (U. S. Chamber, 2007) to each state and the percent of students on free and reduced lunches (proxy for poverty used by NCES).

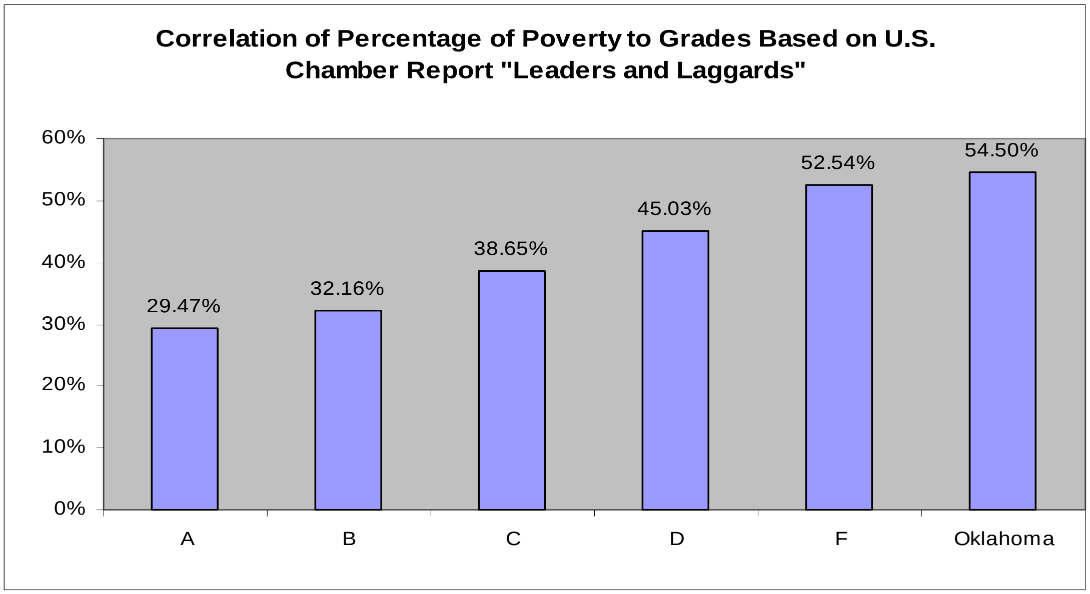


Figure 1

Martin (2007) also found a strong relationship between the letter grades given by U. S. Chamber (2007) and their adjusted per student expenditures. Figure 2 illustrates the relationship:

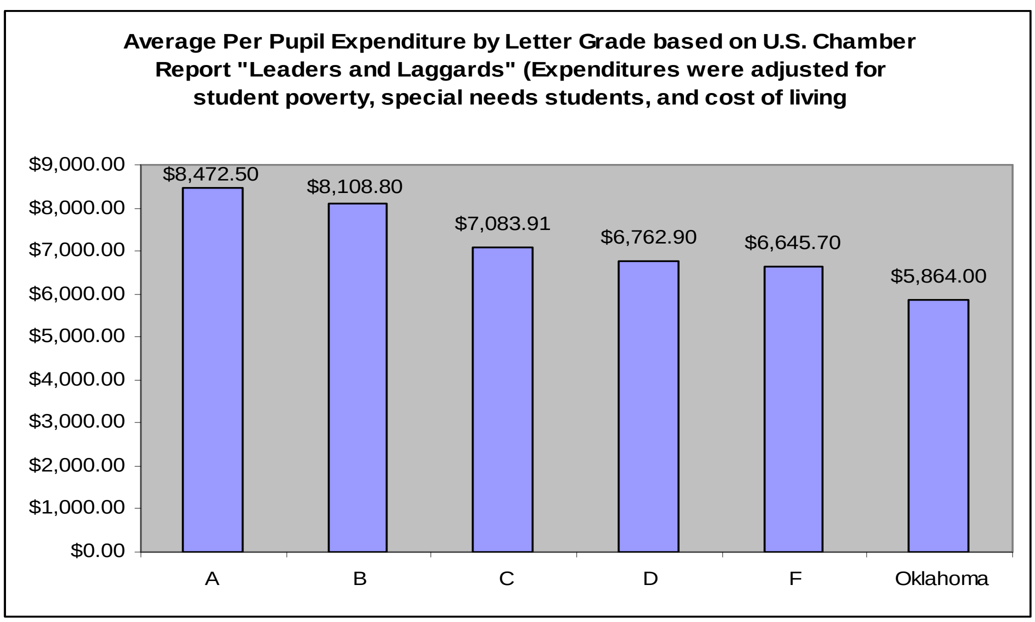


Figure 2

Rothstein, Jacobsen & Wilder (2006) used statistical analysis of student achievement across the United States and world to support their claim that the conceptual basis of NCLB is deeply flawed:

> No goal can simultaneously be challenging to and achievable by all students across the entire achievement distribution. A standard can either be a minimal standard which presents no challenge to typical and advanced students, or it can be a challenging standard which is unachievable by most below-average students. No standard can serve both purposes—this is why we call 'proficiency for all' an oxymoron-but this is what NCLB requires. (p. 2)

NCLB requires children to reach proficiency on challenging state academic standards and assessments. The standards must have rigorous content. Although the law does not further define challenging standards, NCLB uses language to describe proficiency that parallels NAEP. Cross, appointed by the U.S. Department of Education in 2002 to coordinate rulemaking for NCLB, noted, NAEP "is supposed to be the benchmark for states, and that is why its use was expanded" in the act (Rothstein et al., 2006, p. 5).

The study claims that reaching proficiency for all is an even higher and more unreachable aspiration than being first in the world, because of the wide range of performance levels, even in the highest achieving nations. They support this claim citing a study by NCES in 1993 that computed an approximate equation of performance between American students on the eighth grade NAEP test, given in 1992, and an international exam, the Second International Assessment of Educational Progress (IAEP), given the previous year. Taiwan was first in the world in math in

1991. If Taiwanese 13 year-olds had taken the U.S. NAEP exam the following year, their estimated average NAEP score would have been 285, compared to American eighth graders' average score of 262. The NAEP requirement for proficiency was 299 at that time. Taiwanese students were first in the world in math; however, approximately 60% of them scored below what NAEP defines as proficient (Rothstein et al., 2006)

Rothstein et al. (2006) further support their claim that reaching proficiency for all is a higher aspiration than being first in the world by citing 2003 data from the Third International Mathematics and Science Survey (TIMMS) that found Singapore, the highest scoring country in the world in mathematics, would still have 25% of its students failing to meet the NAEP definition of proficiency while Korea, the second highest scoring country in the world in mathematics would have one-third of its students less than proficient in math.

These researchers further support their claim that being first in the world would be much easier than having all students reach the proficiency level of NAEP citing performance on a 2001 reading test administered by the International Association for the Evaluation of Educational Achievement. America's 10 year-olds scored ninth highest in the world. Sweden was the highest scoring country on this reading test. All of the countries were closely bunched together on scores with the U.S. only 0.2 standard deviation below Sweden. Only 30% of U.S. 10-year olds met the proficient category in reading the next year on the NAEP test. Rothstein et al. (2006) caution that this is only a rough approximation, but "by comparing NAEP scale to scores to the international reading test, they estimated two-thirds of all

Swedish students, the highest scoring students in the world, were not proficient in reading as NAEP defines it" (p. 12).

Rothstein et al. (2006) claim the federal education administration uses the term grade level interchangeably with proficiency. They cite past Secretary of Education Margaret Spellings describing NCLB in this way: "We've set a historic goal to ensure every child—regardless of race, income or zip code—can read and do math at grade level. And we've given ourselves a deadline to do it by 2014 because parents have waited long enough" (p. 14). However, the Department of Education has not defined what grade level means any more than it has defined proficiency.

The authors contend that the term grade level describes performance that is considerably below the standard of performance as defined by NAEP or NCLB. Rothstein et al. (2006) state that grade level performance:

> Usually means the average performance of students currently in a given grade. It is usually established by administering a standardized test to a national random sample of students in that grade. The average score is, by definition, grade-level performance in the base year in which the national sample was tested. And also by definition, approximately half of the students in the nation perform below grade level to some degree. Increasing numbers of students, of course, can get above a previously established grade level standard if achievement rises subsequent to the base year. But no matter how high average achievement becomes, approximately half of all students will demonstrate below grade-level performance for the year in which it is measured. When the Department of Education posits a goal of having all

> students at grade level, it presumably intends a contemporary grade-level standard, not an historic and obsolete one. If this is its intent, then all students at grade level is a logical impossibility. (p. 15)

Furthermore they state that even if we interpret the NCLB goal as all students achieving grade level standards set in 2001 with the drafting of NCLB, the task is still unrealistic. They support this with comparisons internationally. Taiwan, the highest scoring nation in the world in eighth grade mathematics in 1991, had 25% of its students score below the average performance of U.S. students. A 2003 TIMSS comparison shows 10% of eighth grade students in Singapore, the highest scoring country in the TIMSS assessment, were similarly below grade level for the U.S. in that year (Rothstein et al., 2006).

These researchers concluded that inevitable individual variability in the human population made proficiency for all an oxymoron:

> There is no aspect of human performance or behavior that is not achieved in different degrees by individuals in a large population. There is an average level of math performance for eighth graders, but some perform above or below that level. There is an average level of teaching ability for eighth grade math teachers, but some perform above or below that level. There is an average susceptibility to influenza, an average pace to run a mile, an average height and weight for adults, an average inclination to attend church each week, an average skill in the operation of motor vehicles. In each of these areas, some individuals are considerably above average, many are slightly above average, many are slightly below average, and some are considerably

below average. In most of these areas, the distributions are close to what statisticians call normal (when plotted, the resulting grade looks bell-shaped), but perfect normality is not the rule. In general, however, for distributions that are close to normal, we say that roughly two-thirds of all humans perform reasonably similarly on any characteristic—statistically speaking, we say that the approximate one-third who perform slightly below average are within one standard deviation of the mean, and the approximately one-third who perform slightly above average are also within one standard deviation of the mean. But this still leaves about one-sixth who are considerably below average, as well as about one-sixth who are considerably above. In its administration of NCLB, the U.S. Department of Education barely acknowledges this human variability. It permits the lowest performing 1% of all students to be held to a vague "alternate" standard of proficiency, and the next lowest performing 2% to be held to a "modified" standard of proficiency, which still must lead to "grade level" achievement and a regular high school diploma. Let's be clear about what this means: Under NCLB, children with I.Q. s as low as 65 must achieve a standard of proficiency in math which is higher than that achieved by 60% of students in Taiwan, the highest scoring country in the world (in math), and a standard of proficiency in reading which is higher than that achieved by 65% of students in Sweden, the highest scoring country in the world (in reading) (Rothstein, et al., 2006, pp. 17-18).

Linn (2006) concurs that the requirement for all students to reach proficiency is unrealistic and this feature of NCLB that is a critical piece of the law threatens to undermine other praiseworthy aspects of the legislation. Furthermore; he states that the expectation for all students performing at the proficient level or above by 2013-2014 school year is unrealistic, unobtainable, proficient student achievement is poorly defined, varies from state to state, thus becoming a meaningless concept.

Although the definition of proficient was left to states, "It is clear that the intent of the law is that the proficient and advanced levels should be set at ambitious levels" (Linn, 2006, p. 5). He asserts that when NAEP achievement levels were set initially in 1990, they were set at quite high levels. When NAEP mathematics results were first reported in terms of achievement levels in 1990, a fourth grade student had to be at the 87th percentile to be counted as proficient while an eighth grade student at the proficient level corresponded to the 85th percentile (Braswell, Lutkus, Grigg, Santapau, Tay-Lim & Johnson, 2001).

Linn (2006) suggests that since NAEP is the only common measure of student achievement and the only uniform definition of proficient across states, a potential use of NAEP is to evaluate the reasonableness of the expectations that all students will achieve a proficient level or above by 2013-2014. Since 1990, students performing at the proficient level or above on the NAEP mathematics assessments have made fairly substantial gains. The gains are more significant at the fourth grade, but still noticeable at the eighth grade level. On the 2005 NAEP mathematics assessment, 36% of fourth grade students were at the proficient level or above compared to only 13% in 1990. Eighth grade students on the mathematics NAEP assessment increased from 15% in 1990 to 30% in 2005. Although these rates of increase are encouraging (1.53% per year at fourth grade and 1.0% per year at eighth grade), a continuation of these trends would result in 50% of the fourth graders proficient and 39% of eighth graders proficient or above by 2014, a long way from the requirement of 100% (Linn, 2006).

Linn (2006) uses the NAEP reading assessment trends to support his argument that the same holds true for reading. In 1992, when NAEP reading assessments were set, students had to be at the 71st percentile nationally to reach the proficiency reading level in fourth and eighth grades; that is, 29% of students were at the proficient level or above. Only a slight increase has occurred since then with 31% of public school students meeting the NAEP proficiency level at both fourth and eighth grade reading. "Radical changes clearly would be needed in the trends to even come close to the NCLB 100% goal and there is no reasonable basis for thinking that such changes are feasible" (Linn, 2006, p. 8).

Individual states were examined to determine if any are remotely on track to reach the 100% level of proficiency as proscribed by NCLB. The results indicate larger increases in mathematics than reading, yet only three states (Idaho, Massachusetts and Montana) had increases that averaged four or more percentage points per year at grade 4 and only one state, Texas, had an increase at eighth grade that averaged as much as three percentage points per year (Linn, 2006).

> If the proficient achievement standard is set at a high level as prescribed by NCLB and reflected on NAEP, then the goal of having all students performing at that level by 2014 is out of reach despite intensive efforts of educators and students. Unobtainable goals might not be a bad thing if there were no consequences for failing to reach them. Failure to meet NCLB goals,

> however, has serious consequences for schools, educators, and students. Setting unrealistically ambitious goals and sanctioning schools that fail to meet them does more to demoralize than to motivate educators. This is not an argument that goals should not be ambitious, but they also should be realistically obtainable given sufficient effort, especially when there are sanctions for not reaching them. At the very least there should be an "existence proof" that goals are obtainable (Linn, 2006, pp. 9-10).

NAEP

Yeager (2007) states that NAEP, often referred to as the "Nation's Report Card," is the only test that promises to measure student achievement across the country based on time and demographic groups. NAEP began in 1969, following enactment of the Elementary and Secondary Act (ESEA) in 1965. Upon NAEP's inception, state and local officials were resistive because they saw it as a federal intrusion on state policy. Resistance began to wane by the mid 1980's, following the 1983 report *A Nation at Risk,* as state leaders started to compare their assessment data to NAEP to show the impact of educational reform. "Today NAEP has two primary goals: comparing student achievement across states and tracking changes in national educational achievement over time" (Yeager, 2007, p. 2). NAEP is administered by the National Center of Educational Statistics (NCES), located in the U.S. Department of Education, and the National Assessment Governing Board (NAGB), a bipartisan board composed of governors, state and local education officials, business leaders, teachers, principals, measurement experts, and parents (Yeager, 2007).

Prior to NCLB, NAEP was voluntary for students, school districts, and states, but NCLB requires states to participate in the biennial fourth and eighth grade reading and math main NAEP assessments to receive their federal Title I funds. Critics have questioned the low stakes nature of NAEP and if students perform their best on the test due to lack of incentives typically available on other assessments. However, studies conclude that results may be slightly affected by the lack of student motivation on the NAEP test, but that any effect is likely small (Linn & Baker, 1996).

Yeager (2007) explains that NAEP's achievement levels (basic, proficient, advanced) came out in 1990 following the appointment of NAGB as a guide for what students should know and be able to do, with basic representing partial mastery of a subject, proficient representing solid performance, and advanced representing superior performance. "It marked a key transition for NAEP, the test shifted from strictly reporting performance, to judging performance against a standard, thus expanding its focus in measurement to include evaluative and interpretive functions" (Yeager, 2007, p. 10). Yeager (2007) quotes Diane Ravitich, research professor and former NAGB member, "No single aspect of NAEP has been more valuable to the public—nor more controversial" (p. 10).

Critics argue that the levels give a false sense of accuracy, when in fact, like all standard-setting processes, are subjective. Those involved in the process contend that the levels were carefully implemented and consumers want to attribute more to the results than the results can provide (Yeager, 2007). Reckase (2001), a consultant to NAGB describes the standard-setting process as "the most thoroughly planned, carefully executed, exhaustively evaluated, completely documented, and most visible of any standard-setting process he has encountered" (p. 231).

NAGB used a modified "Angoff method" to determine the "cut" or passing scores for each level. This method uses panels of teachers, business leaders, state and local education officials, and testing experts to evaluate test questions to determine the probability that a student just reaching each achievement level could correctly answer the item. The collective responses of the group are averaged to determine a cutoff score for each level. The panel evaluates the test as a whole and adjusts the achievement levels accordingly. "This method, like all standard-setting processes relies on human judgment and therefore, is subjective, leading to disagreement among researchers" (Yeager, 2007, p. 10).

Yeager (2007) lists researchers who have questioned NAEP's standard setting methods. The Government Accountability Office (GAO) commented in 1993 that NAGB's approach (to setting achievement levels) is unsuited for NAEP, and characterized the resulting levels as misleading. The National Academy of Science (NAS) analyzed the development of achievement levels for the 1996 science test and concluded "the process was fundamentally flawed because of the difficult and confusing task given to judges, inconsistencies in their judgments of items, lack of evidence for cut scores, and the unreasonable results that came out of the process" (p. 10).

Yeager (2007) admits "that without individual level scores, it was difficult to prove that students scoring at the advanced level were, in fact, later successful in college classes" (p. 11). In 1994, when Congress was considering ESEA reauthorization, the controversy surrounding the achievement levels reached such a point that the House Education and Labor committee voted to abolish the NAGB; however, the Senate restored its full authority in the final law. The 1994 ESEA reauthorization language described the achievement levels as developmental. The 2001 reauthorization of ESEA (NCLB) contained a similar provision specifying that the achievement levels should be used on a "trial" basis until determined through evaluation to be "reasonable, valid, and informative to the public" (Yeager, 2007, p. 11).

> NAEP has some serious limitations that impede its ability to provide comprehensive information to policymakers. State-level NAEP, for instance, provides snapshot data of achievement at a particular point in time for grades 4 and 8. It does not track a single set of students over time, which makes it difficult to use in tracking the outcomes of education reforms and policies, or to measure individual school performance. Yet policymakers, eager to cite NAEP as evidence of the success of education reform, are quick to provide specious interpretations that can cast favorable light on their state or district. Unfortunately, NAEP is widely used to support or oppose educational reforms, a role for which it is ill-suited (Yeager, 2007, pp. 11-12).

Loveless (2007), argues that the accusations that states are "dumbing down" their tests or setting cut scores too low to inflate student numbers reaching proficiency rests on the comparison of the number of students reaching proficiency on state tests compared to the percentage reported by NAEP. In almost all cases, states report a significantly higher percent proficient than the NAEP results verify. The first assumption these critics make is that few students reach proficient on NAEP because it is a rigorous test. A content analysis of the NAEP tests does not support this belief as the eighth grade mathematics NAEP test is dominated by problem solving with whole numbers, a concept usually taught by the end of the third grade. The eighth grade NAEP math test assesses two-step word problems with whole-number arithmetic which leaves little to evaluate about knowledge of complex numbers such as fractions, decimals, and percents. He cites a National Validity Study of NAEP that verified that less than 15% of the eighth grade NAEP math test is devoted to fractions. "Raw computation items that require students to add, subtract, multiply, or divide fractions or decimals-or even whole numbers for that matter–are virtually absent from NAEP" (Loveless, 2007, p. 10).

Loveless (2007) asserts the second assumption is that "states are racing to the bottom in response to NCLB" (p. 10). He says that this is not supported by evidence as states reported larger percentages of proficient students than NAEP before NCLB, with no appreciable change since NCLB's enactment.

The third assumption is that NAEP performance levels–where cut scores for performance levels have been set–are valid. Phillips (2007) linked The Trends in International Mathematics and Science Survey (TIMMS), which also assesses eighth graders' knowledge of mathematics and NAEP. He mapped the NAEP achievement levels onto the TIMSS scale and found that only five nations in the world would meet the NAEP proficiency level (Singapore, Korea, Hong Kong, Chinese Taipel, and Japan), 22 nations scored at the NAEP basic level, and 19 nations scored below the NAEP basic level. The United States' score was at the basic level, a little below the mid-point of the category. The five highest achieving nations in the world did not meet the goal of 100% proficiency based on the NAEP definition. Singapore had 73% of its students meeting the NAEP proficiency level with Hong Kong at 66%, Korea at 65%, Chinese Taipel at 62% and Japan at 57%.

An analysis of the United States' NAEP scores on eighth grade mathematics indicates two years' learning gains from 1990 to 2007. At this rate, the U.S. can join the group of the five proficient nations in 21 years and catch up with Singapore in about 41 years assuming they stay at their current achievement levels. Loveless (2007) further states:

> Such a gain would also represent more than four additional years of learning since 1990, on top of the 2.3 year gains that eighth graders have already accomplished. If you believe the NAEP scales and achievement levels, an eighth grader of 1990 needed to know about six more years of mathematics, equivalent to a 1990 sophomore in college, to be proficient at eighth-grade mathematics. (p.12)

Loveless (2007) contends several prestigious organizations have reviewed the NAEP achievement levels and questioned their validity, including the Government Accountability Office, the National Academy of Sciences, and the National Academy of Education. The National Academy of Education report concluded that NAEP cut scores were too high and attributed this to weak content of the tests.

He concludes that NAEP achievement levels need to be changed, especially with their increased importance due to NCLB. He suggests linking NAEP to other assessments (e.g., TIMMS). "If 25% to 50 % of students in the top-scoring countries in the world fail to meet an American standard of proficiency, one wonders how realistic that standard is as a universal expectation" (Loveless, 2007, p. 13).

Study of the Oklahoma's Accountability System

Brewer and Killeen (2009) studied Oklahoma's accountability system and defined accountability "as a contractual relationship between two parties–a provider of a good/service and a director with the power to reward, punish, or replace the provider" (p. 9). The basic building blocks of standards-based accountability are:

1. Goals embodied in a set of content and performance standards that schools and teachers use to guide curriculum and instruction,
2. Student assessments for determining if students have mastered the standards,
3. Consequences (incentives) with improved performance leading to rewards and poor performance leading to sanctions. (Brewer & Killeen, 2009, p. 9)

The report states that general sequencing of content standards may be similar among states; however, performance standards defining how much of the standards students should master differ greatly across states. Oklahoma's content standards were considered to be "fairly well developed" while their performance standards were described as very weak and lacking rigor. The low performance standards ratings were due to the "large discrepancy between NAEP achievement levels and Oklahoma's own assessments along with the stakeholders' perceptions of little transparency in the way standards are set (Brewer & Killeen, 2009).

Oklahoma created a statewide education accountability system a decade before NCLB mandated one. SB 183 and HB 1017 in 1989 laid the foundation for Oklahoma's accountability system with the former creating the system of declaring school districts academically at risk because of low test scores. Another important piece of legislation was HB 1466 in 1985 creating the OSTP, the main student assessment for the state's accountability system (Brewer & Kileen, 2009). They further describe the history of Oklahoma's accountability system:

> SB 183, which required norm-referenced tests in grades 3, 5, 7, 9, and 11, expanded the OSTP in 1989. In 2000 all norm-referenced tests were dropped, and were replaced by four EOI (course specific Criterion Reference Tests) for high school: English II, U.S. History, Algebra I, and Biology I, and the complete Iowa Test of Basic Skills for third graders. In 2002 the Math and Reading sections of the Stanford 9 replaced the Iowa tests. CRT's were added to fourth grade in 2005, and sixth and seventh grades in 2006. Since its inception, the OSTP has changed both the format of the tests (i.e., norm-referenced, curriculum referenced, end-of-instruction) and testing companies hired. Riverside provided the norm-referenced tests from 1985 to 2001. Harcourt-Brace provided the CRTs from 2001 to 2005. CTB McGraw-Hill provided the CRTs from 1998-2000, and then the EOI tests starting in 2001. Data Recognition Corporation took over the CRTs in 2005 (p. 26).

EXPLORE

According to the (Oklahoma Educational Planning and Assessment System [OEPAS], 1999) the Oklahoma State Regents for higher Education (OSRHE) partnered with ACT in 1993 to create EPAS. OSRHE has fully funded this initiative, provided to Oklahoma school districts at no cost.

ACT (2003) states EPAS now serves over 95% of Oklahoma's eighth and tenth grade public school students and, in 2003, Oklahoma EPAS served 489 school districts (including 42 private districts and two Bureau of Indian Affairs schools), with nearly 85,000 students taking the assessments and benefitting from the system's early interventions (p. 2).

OEPAS (1999) contends it is a comprehensive student preparation system that contains the EXPLORE test in eighth grade, the PLAN test in tenth grade, and the ACT in eleventh or twelfth grade. Furthermore, EPAS is linked to the PASS, Oklahoma's core curriculum framework (p. 4). EPAS is a voluntary program over and above Oklahoma's mandated K-12 testing requirements. The three assessments of EPAS—EXPLORE (eighth grade), PLAN (tenth grade) and ACT (eleventh or twelfth grade)—comprise the only assessment system in the state that measures student readiness along a continuum of college readiness benchmarks (ACT, 2003, p. 1).

EXPLORE (2008/2009) contends its report can help answer three important questions about a student's future.

1. Where do I stand right now? EXPLORE shows your strengths and weaknesses in four subject areas: English, math, reading, and science. You can see how your scores compare to those of other students like you who have taken EXPLORE. Your report also shows you the skills you likely have.

2. What are my plans and goals for after high school? When you took EXPLORE, you answered questions about your plans after high school and about your interests in several kinds of activities. This information can help you learn more about careers, clarify your goals, and begin to plan your future.
3. Am I on track for college? For most students, reaching their goals includes getting a college education--Preparing for college means taking the right courses in high school. Make sure the courses you plan to take in high school match those recommended for college success. (p. 1)

EXPLORE (2008/2009) explains the composite score range of EXPLORE is 1 to 25 and that no test can measure educational development with absolute precision. An example given is that a score of 16 on one of the four tests, such as reading, means that educational development is probably somewhere from 14 to 18 (16 plus or minus 2). College readiness benchmarks are also given on the EXPORE test. ACT has identified college readiness benchmark scores that indicate if students have scored at or above the benchmark scores, they are likely to be on track and do well in entry-level college courses in these subjects. This assumes that these students will continue to work diligently and take challenging courses throughout high school. The college readiness benchmark scores for mathematics are 17 and the benchmark for reading is 15 (EXPLORE, 2008/2009, p. 4).

Oklahoma School Testing Program (OSTP)

The (Oklahoma School Testing Program [OSTP] Test Interpretation Manual, Grades 3-8, Oklahoma Core Curriculum Tests, 2008) states that:

1. Beginning in 1995, the State Department of Education (SDE) administered Oklahoma Core Curriculum Tests (OCCT) at Grades 5, 8, and 11 in Mathematics and Science; Grade 8 also tested in Reading and Writing.
2. In 1996, tests in Reading and Writing were added in Grades 5 and 11. In 1997, U.S. History, Constitution, and Government was added; in 1998 Oklahoma History was added in Grade 11, and Geography was added in all three grades.
3. In 1999, Fine Arts was included in the OCCT's.
4. In the spring of 2001, OCCTs were administered to students in Grades 5 and 8 in Reading; Mathematics; U.S. History, Constitution, and Government; Science; Geography; and Fine Arts. (p. 1)

On June 7, 2003, House Bill 1414 was signed into law changing the OSTP. Many of the changes were to comply with the requirements of NCLB:

1. During the 2003-2004 school year, the third grade norm-referenced test was administered for the last time. Beginning with the 2004-2005 school year, there was no state-level, norm-referenced testing in the OSTP.
2. During the 2003-2004 school year, the third grade norm-referenced test was replaced with Reading and Mathematics criterion-referenced tests aligned to Oklahoma's standards and the PASS. Field testing for these criterion-referenced tests occurred during the 2003-2004 school year and the tests were implemented statewide in the 2004-2005 school year.
3. Grade 4 Reading and Mathematics criterion-referenced tests were field tested during the 2003-2004 school year and implemented statewide in the 2004-2005 school year.

4. Beginning in the 2003-2004 school year, the Grade 5 U.S. History, Constitution, and Government criterion-referenced test and the Grade 5 Geography criterion-referenced Social Studies test were combined into one test.
5. Beginning in 2003-2004, the Grade 8 criterion-referenced Geography test was eliminated. A Grade 7 criterion-referenced Geography test was field tested during the 2003-2004 school year and implemented statewide in the 2004-2005 school year. This test was administered online except for those students needing accommodations.
6. Beginning in the 2003-2004 school year, the criterion-referenced Fine Arts test was eliminated at Grades 5 and 8. District-level fine arts assessments were implemented in Grades 3 through 8 beginning with the 2004-2005 school year.
7. Grades 6 and 7 Reading and Mathematics criterion-referenced tests were tested during the 2004-2005 school year and were implemented statewide during the 2005-2006 school year. (OSTP, 2008, p. 2).

In 2007-2008, the OSTP tested Mathematics and Reading in Grades 3-8, Grade 5 added Science, Social Studies and Writing; Grade 7 added Geography, and Grade 8 added Science, U.S. History and Writing.

OSTP (2008) states the key components of the OCCTS are the PASS, performance level descriptors, and the Oklahoma Performance Index (OPI). These components ensure the validity and reliability of the testing program, as well as the reports that are produced. The purpose of the OCCT is to gather information about student performance to ensure students meet high standards and to evaluate success as presented in the PASS standards.

The manual classifies student performance into the four performance levels of advanced, satisfactory, limited knowledge and unsatisfactory and advises that to interpret OCCT results correctly, it is important to understand the specific knowledge and skills required of a student at each performance level. The performance descriptors identify the specific knowledge and skills that a student must demonstrate at each level. These performance level descriptors use competency-based, content-specific phrases developed by panels of Oklahoma educators and approved by the Oklahoma State Board of Education. NCLB mandates that states define performance levels on statewide assessments (OSTP, 2008).

Performance is scored by a method called the Oklahoma Performance Index (OPI) reported on a scale from 400-990. It contends that this system is more accurate than percent correct because of the added factor of the difficulty level of questions and the factoring of the possibility of guessing correctly. The OPI scores are reported on a scale because test questions change from year-to-year, creating a test to be slightly more or less difficult than the prior year. OPI scores account for this difference in difficulty, and scores are reported on a common scale so they have the same meaning from year- to-year. An example given in the manual was that one year a student may need to answer 37 questions correctly to obtain an OPI score of 750 while the next year 35 questions may meet the same threshold due to a slight increase in the difficulty of the test. OPI scores are useful for comparing student scores for the same grade and subject area, but cannot be used to accurately compare scores across grades or subjects (OSTP, 2008).

Summary and Theory

The definitions of proficiency as required by NCLB and the mandate that all students meet that standard by 2014 have created a great dilemma for public education in Oklahoma and across the nation. Critics such as (Yeager, 2007; Carey, 2006; Cronin et al., 2007), accuse states of lowering their standards to avoid the punitive actions of NCLB for failing to reach prescribed achievement levels each year. The concern of states setting low standards has led advocates to promote the proficiency definition of NAEP as the required standard for all states.

Other researchers such as (Kahlenberg, 2008; Rothstein et al., 2006; Linn, 2006), have provided evidence that the goal of all students reaching proficiency on NAEP or any other rigorous test is impossible due to the variability in human ability. NAEP also has many critics. Yeager (2007) admits that without tracking individual scores it is difficult to determine if NAEP scores are predictors of future success in college classes. Loveless (2007) contends that NAEP is a simple test of content but counters this with a very high accuracy requirement to compensate for the lack of rigorous questions. Phillips (2007) provides evidence that the highest achieving nations in the world could not meet the requirement of all students attaining proficiency.

Cronin et. al (2007) advise that states need an external yardstick to examine their cut scores. They state that NAEP is the most commonly used, but agree that it is a less than desirable tool due to the concerns mentioned above. The EXPLORE test is a predictor of future academic success (ACT, 2008) and is given to many eighth grade students across Oklahoma in the early fall. The OSTP is given to all Oklahoma students in the third-eighth grades each spring in mathematics and reading. The EXPLORE test will provide an external yardstick to give guidance on the meaning of OPI scores and performance level to allow parents, educators, and policy makers to interpret OPI scores and performance levels as national percentile ranks and college readiness standards. Somerville and Yi (2002) support that all students need college readiness skills regardless of their plans to attend college or enter the workforce.

The study used theory supported by the concept of normal distribution, discussed in Chapter 1, as its basis for the difficulty of meeting this requirement. Score distributions on the OSTP tests were analyzed to illustrate if known variations in human traits and abilities were present. Finally, student performance was tracked from grades 3-7 to see if sufficient progress was being made to meet the deadline of 2014 set by NCLB for all students to attain proficiency.

Chapter III

METHODOLOGY

Introduction

The study had two purposes with the first purpose being to define "proficiency levels" (advanced, satisfactory, limited knowledge, and unsatisfactory) and OPI scores on the third-seventh grade math and reading OSTP in a manner that is meaningful to parents, educators, and policy makers. Defining proficiency levels and OPI scores in a meaningful manner was accomplished by determining a relationship between student achievement on the OSTP and the EXPLORE test in math and reading. The relationship allowed the researcher to estimate national percentile ranks and college readiness based on the relationship to the EXPLORE test. A second purpose was to examine realism of the NCLB requirement of 100% of students reaching the proficiency level by 2014. Two methods were employed to evaluate this proposition. OPI scores distributions within each grade and subject were examined to determine if the known variability in human ability was demonstrated. The study tracked performance levels of students from third to seventh grade in math and reading and examined the effectiveness of current remediation practices to determine if student achievement is improving, particularly for students who fall in the unsatisfactory and limited knowledge categories.

This chapter describes the study's population data, the instruments used to collect data, and the procedures for collecting and analyzing data relating to two research questions on predictive validity and two OSTP research questions.

Predictive Validity Research Questions

1. Are OPI scores in the OSTP third-seventh grade math and reading statistically significantly related to performance in math and reading on the eighth grade EXPLORE test?
2. Do the OPI scores and performance levels on the OSTP third-seventh grade math and reading assessment predict national percentile ranks and college readiness standards as defined by the EXPLORE test?

OSTP Research Questions:

1. Does OSTP performance at each grade (3-7) and subject (math and reading) follow a normal distribution?

Distribution of scores will be used to offer evidence the known differences in student ability make it difficult to create a test that is challenging to average and above average students, yet attainable by the lower performing students.

2. Do student performance levels on OSTP math and reading remain consistent from year to year (third-seventh grade)?

A comparison of performance levels was used to evaluate if remediation efforts improved student achievement. Particular focus was on students who scored at the limited knowledge and unsatisfactory levels to determine if Oklahoma schools were moving towards the NCLB requirement of all students reaching proficiency by 2014.

Sample Population

The study's population data were derived from approximately 1150 eighth grade students using their third through seventh grade OSTP math and reading test and their eighth grade EXPLORE math and reading test. Students selected were from 12 public school districts in Oklahoma with an aggregate average API score of 1310 compared to the state average of 1279. Enrollments of the districts ranged from 650 to 2,860 students.

All 12 school districts had a middle school (sixth through eighth grade) configuration with the schools ranging in size from 120 students to 650 students. The average mobility rate of the schools was 7.2% compared to the state average of 10.2%, indicating a population less mobile than the state average. Student attendance rate was 94.4% compared to the state average of 94.2%, or a difference of only .2%.

Instruments

Instruments included the third through seventh grade OSTP math and reading assessments and the eighth grade EXPLORE math and reading test. OSTP uses third-seventh grade criterion-reference tests to measure student attainment of skills established in Oklahoma's Core Curriculum, PASS. The EXPLORE test is a part of ACT's Educational Planning and Assessment System (EPAS). The Oklahoma State Regents for Higher Education Student Preparation team formed committees in the summer of 2004 to compare EPAS Standards to the PASS standards in curriculum areas. They found consistency between EPAS and PASS standards (OEPAS, 2004). If tests cover the same standards, their related assessments should have congruent validity between them.

The OSTP scale scores are given in a format called Oklahoma Performance Index (OPI) which factors in the difficulty level of questions along with the possibility of guessing correctly to determine scores. Performance levels (advanced, satisfactory, limited knowledge, and unsatisfactory) are determined by the Oklahoma State Department of Education (OSDE) and vary from subject-to-subject and grade-to-grade (OSTP, 2008). Table 1 lists the OPI performance level cut scores for each grade, subject, and year of the data involved in the study.

Table 1

OPI Performance Level Cut Scores

Grade	Year	Advanced	Satisfactory	Limited Knowledge	Unsatisfactory
			Math		
3	2004-05	807-990	700-806	604-699	400-603
4	2005-06	824-990	700-823	600-699	400-599
5	2006-07	806-990	700-805	615-699	400-614
6	2007-08	779-990	700-778	652-699	400-651
7	2008-09	766-990	700-765	667-699	400-666
			Reading		
3	2004-05	881-990	700-880	623-699	400-622
4	2005-06	879-990	700-878	631-699	400-630
5	2006-07	839-990	700-838	635-699	400-634
6	2007-08	831-990	700-830	656-699	400-655
7	2008-09	802-990	700-801	668-699	400-667

To comply with NCLB, Oklahoma began to test students in grades 3-8 with the Oklahoma Core Curriculum Test (OCCT) in math and reading in 2004-2005. The mathematics section of the test contains multiple choice questions to measure knowledge and skills in five performance standards: patterns and algebraic reasoning, number sense, number operations and computations, geometry and measurement, and data analysis and probability. The reading test also has five performance standards: vocabulary, comprehensive/critical literacy, literature, and research and information (OSTP, 2008).

Data Collection Procedures

Data related to the OSTP math and reading test as well as the EXPLORE test were stored in each district's administrative offices. Telephone calls were made to each superintendent describing the method and purpose of the study. Permission was received for the data from each school district. By assigning an identification number to each student, data received by the researcher were anonymous, absent student names and identifying information.

A review of the database was performed to identify any missing assessments. All identification numbers not having all math and reading test scores (third, fourth, fifth, sixth, and seventh grades, and eighth grade EXPLORE) associated with them were eliminated from the study. This process resulted in a reduction of the sample population to 586 sets of scores. The number of missing scores was higher than anticipated when the study began. Several factors can be considered as reasons for the missing data. The most likely cause was the studies sample had a mobility rate of 7.2% each year. Although the data should have followed the student to the next district, the data either wasn't sent or stored in an area not assessable to the person responsible for providing the data to the researcher. Student absenteeism is another potential reason for the missing data. Make-up tests were available for students missing the original exam date; however, students absent for extended periods during the testing window may have missed the exams. A final reason considered is students originally taking the regular OSTP assessment but classified later on in a learning disability category that allowed them to take an alternative assessment. Alternative assessment scores were eliminated from the study. All data were placed in a SPSS program for analysis.

Data Analysis Procedures

According to the (OEPAS, 2004), curriculum committees formed by the Oklahoma State Regents for Higher Education Student Preparation Team in the summer of 2004 found that consistency existed between the EPAS Mathematics Standards for Transition and the PASS Mathematics objectives, the EPAS Reading Standards for Transition and the PASS Social Studies Process Skills, and the EPAS English Standards for Transition and the PASS Language Arts objects. The consistency of standards is an indication that both OSTP math and reading assessments and the EXPLORE have content validity. Content validity is usually determined by a group of experts who decide if the test items are reflective of the larger body of knowledge (Sattler, 1992). Therefore, evidence is not supported by data and is qualitative. Due to dependence on human judgment, content validity is insufficient to determine correlation between assessments (Lally, 2000). To establish such correlation, some form of construct or criterion validity is necessary. Criterion validity draws an inference from a test score to performance (Sattler, 1992) and one form of criterion validity is predictive validity.

In this study, the OSTP assessment scores represented the independent or predictor variable for performance on the eighth grade EXPLORE which represented the dependent or criterion variable. Measurement of the correlation between the two variables established a correlation coefficient. Correlation coefficients range from -1.00 to +1.00 with -1.00 representing a perfect negative correlation and +1.00 representing a perfect positive correlation. A value of 0.0 represents no correlation. The theoretical framework supported by predictive validity was that OSTP (PASS objectives) and EXPLORE (EPAS objectives) had consistent math and reading standards, objectives, and process skills. Therefore, OPI scores on the OSTP should predict performance on the EXPLORE. The theoretical framework would be accepted if the correlation coefficient reached statistical significance at the .01 level.

Regression analyses were also performed on the OSTP math and reading/ EXPLORE math and reading data to support further the findings in this part of the study. Regression analysis is a statistical tool that uses the relationship between quantitative variables in order for other variable(s) to predict the dependent variable. In this study the dependent variable was performance on the EXPLORE and the independent variable is the scores on the OSTP assessment. Regression analysis serves to find the line that best predicts the relationship between variables, the line of "best fit," when points are plotted from the data. Since most relationships are not perfect, there is deviation from this line and it is referred to as the residual value. The smaller the variability of the residual values, the better the prediction. If the variables are perfectly related, there is no residual variance and the ratio of variance would be one. Most ratios fall between zero and one. This ratio is called the R-squared (r^2)

value. The r^2 value indicates how well the model fits the data, and therefore, how much of the variance is accounted for.

To determine if OPI scores for each grade and subject met the normal distribution criteria, the Kolmogorov-Smirnov (one-variable) test was used. Kurtosis and skewness were analyzed to confirm normal distribution or to determine the manner the sample scores failed to be normally distributed. Kurtosis is the manner in which observations cluster around a central point. Skewness is the measure of the asymmetry of a distribution.

Finally to determine if student performance levels in the third-seventh grade math and reading tests remained consistent from year-to-year, a statistical analysis was run using Cross Tabulations. In this analysis the relationship between students at each performance level in grades 3-7 in both math and reading was investigated.

CHAPTER IV

FINDINGS AND ANALYSES

Introduction

The first of the study's two purposes was to define "proficiency levels" (advanced, satisfactory, limited knowledge, and unsatisfactory) and OPI scores on the third-seventh grade math and reading OSTP in a manner meaningful to parents, educators, and policy makers. Defining proficiency levels and OPI scores in a meaningful manner was accomplished by determining a relationship between student achievement on the OSTP and the EXPLORE test in math and reading. Such a relationship allowed the researcher to estimate national percentile ranks and college readiness based on the relationship to the EXPLORE test. The second purpose was to examine if the NCLB requirement of all students attaining "proficiency" by 2014 was realistic. Two methods were used to evaluate this proposition. Examined was the distribution of OPI scores within each grade and subject. A distribution of scores illustrating the known variations in student ability offers evidence a test challenging to average and above average students is unattainable by lower performing students. Finally, performance levels of students were tracked from third to seventh grade in math and reading to examine the effectiveness of current remediation practices in improving student achievement, particularly students who fell in the unsatisfactory and limited knowledge categories.

Findings Predicting Validity Questions

Research Question #1: Are OPI scores in the OSTP assessments in third-seventh grade math and reading statistically significantly related to performance in math and reading on the eighth grade EXPLORE test?

To determine the relationship between performance on the OSTP math and reading assessments and performance on the EXPLORE, Pearson correlations between assessments were conducted. The correlations examined student performance in terms of OPI scores on the OSTP math and reading at each grade to the EXPLORE math and reading (Table 2). All of the correlations were significant at the .000 level (one-tailed). The direction of the correlations indicated a positive relationship between performance on OSTP third-seventh grade math and reading assessments and performance on EXPLORE math and reading. The correlation size increased each year, with the exception of the sixth grade, as the timing of the administration of the OSTP test came closer to the timing of the administration of the EXPLORE test, indicating each progressive grade (with the exception of the sixth grade) was a slightly better indicator of performance on the EXPLORE test than the prior grade. The math positive correlation coefficients (OSTP grades 3-7/Explore respectively), r=.600, .606, .622, .595, and .645 increased slightly at each grade level with an exception at the sixth grade, which had the lowest correlation coefficient. However, the increase in the size of the correlation relationship, excluding sixth grade was small indicating that the relationship between performance on the OSTP math and EXPLORE math was relatively stable.

The reading positive correlation coefficients (OSTP grades 3-7/EXPLORE respectively), r=.566, .535, .582, .592, and .666 also increased slightly at each grade level with an exception at the fourth grade, which had the lowest correlation coefficient. However, the increase in the size, although greater than in math, of the correlation relationship was small, indicating that the relationship between performance on the OSTP reading and EPLORE reading was also relatively stable.

Table 2:

Correlations Between OSTP Math and EXPLORE Math

Year	Grade	Coefficient	Correlation Significance
2004-05	3	.600	.000
2005-06	4	.606	.000
2006-07	5	.622	.000
2007-08	6	.595	.000
2008-09	7	.645	.000

Table 3:

Correlations Between OSTP Reading and EXPLORE Reading

Year	Grade	Coefficient	Correlation Significance
2004-05	3	.566	.000
2005-06	4	.535	.000
2006-07	5	.582	.000
2007-08	6	.592	.000
2008-09	7	.666	.000

Choudhury (2009) describes r-values of 0 .5 to 1.0 as strong, 0.3 to 0.5 as moderate, and .1 to .3 as weak (disregarding sign of r-value), although he cautions many experts would somewhat disagree on the choice of boundaries. Calkins (2005) classifies correlation coefficients between 0.9 to 1.0 as very highly correlated, correlation coefficients between 0.7 and 0.9 as highly correlated, correlation coefficients between 0.5 and 0.7 as moderately correlated and correlation coefficients between 0.3 and 0.5 as low correlation. Based on these guidelines, the correlation of OSTP math and reading and EXPLORE math and reading were considered as moderate to strong.

In response to Research Question #1, "Are OSTP assessments in third-seventh grade math and reading statistically significantly related to performance in math and reading on the EXPLORE test?" the findings indicated that the OPI scores at each grade and subject were statistically significantly associated with performance on the EXPLORE test at the .000 level of significance and considered moderate to strong in degree of correlation.

The study then examined the second predictive validity research question, "Do the OPI scores and performance levels on the OSTP third-seventh grade math and reading assessment predict national percentile ranks and college readiness as defined by the EXPLORE test?

Analyzing OSTP third-seventh grade combined on each subject to the respective EXPLORE test would account for the largest percentage of variance due to the test; however, OPI scores are based on the difficulty of the test and cannot be compared across grades. Grade levels with similar OPI scores to performance levels were combined and grade levels with dissimilar OPI scores to performance levels were separated. In math, grades 3-5 had similar OPI scores to performance levels, and were therefore combined to predict EXPLORE scores, grade 6 math was used alone to predict EXPLORE scores, and grade 7 math was used alone to predict EXPLORE scores due to both having dissimilar OPI scores to performance levels. Because of the level of variance gained due to the combining grades 3-5, those analyses were run at p=.05. Single grade comparisons were analyzed at p=.01 to limit the possibility of a Type I error.

In reading, third and fourth grade had similar OPI scales to performance levels and were combined to predict EXPLORE reading scores while fifth and sixth grade had similar OPI scores and were combined to predict EXPLORE reading scores. Seventh grade had dissimilar OPI scores in relationship to performance levels which led to a single comparison to predict EXPLORE scores. For combined grades, p=.05 was used for analyses while for the single grade p=.01 was used to limit the possibility of a Type I error. A series of linear regression analyses were performed to predict EXPLOE math and reading scores from OSTP math and reading scores. All of the regressions were significant at the .000 level of significance, and all models showed a positive direction in the ability of the variables to predict performance on the EXPLORE test.

Table 4:

Relationship of OSTP Math and Reading with Performance on the EXPLORE Math and Reading Assessments

Year	Grade	r^2	F	Sig	Beta	df	b-wt.	t	sig.
				Math					
2004-05	3	.462	166.430	.000	.251	3	.013	5.538	.000
2005-06	4	.462	166.430	.000	.212	3	.010	4.291	.000
2006-07	5	.462	166.430	.000	.294	3	.014	6.099	.000
2007-08	6	.366	320.757	.000	.595	1	.033	17.910	.000
2008-09	7	.416	416.305	.000	.645	1	.028	20.404	.000
				Reading					
2004-05	3	.359	162.970	.000	.375	2	.020	8.095	.000
2005-06	4	.359	162.970	.000	.273	2	.017	5.898	.000
2006-07	5	.414	206.031	.000	.338	2	.019	7.977	.000
2007-08	6	.414	206.031	.000	.367	2	.022	8.643	.000
2008-09	7	.443	464.823	.000	.666	1	.035	21.560	.000

In the regression between the third-fifth grade OSTP math and EXPLORE math the model had an F value of 166.430 and was significant at the .000 level of significance with three degrees of freedom. The r^2 value of .462 indicates that the OSTP third-fifth grade math assessments accounted for 46% of the variance related to performance on the EXPLORE math test.

The sixth grade OSTP math assessment model had an F value of 320.757 and was significant at the .000 level of significance with one degree of freedom. The r^2 value of .355 indicated that the grade 6 OSTP math assessment accounted for almost 36% of the variance in performance on the EXPLORE math test.

Examining the seventh grade OSTP math assessment model revealed an F value of 416.305 which was also significant at the .000 level of significance with one degree of freedom. The r^2 value of .416 indicated that the OSTP seventh grade math assessment accounted for almost 42% of the variance in performance on the EXPLORE math test.

In the regression between the third and fourth grade OSTP reading scores and the EXPLORE reading scores the model had an F value of 162.970 and was significant at the .000 level with two degrees of freedom. The r^2 value of .359 indicates that the OSTP third and fourth grade reading assessment accounted for almost 36% of the variance related to performance on the EXPLORE reading test.

The fifth and sixth grade OSTP reading assessment model had an F value of 206.031 and was significant at the .000 level with two degrees of freedom. The r^2 value of .414 indicates that the fifth and sixth grade OSTP reading assessment accounted for over 41% of the variance in performance on the EXPLORE reading test.

Examination of the seventh grade OSTP reading assessment model revealed an F value of 464.823 which was also significant at the .000 level with one degree of freedom. The r^2 value of .443 indicates that the seventh grade OSTP reading assessment accounted for 44% of the variance in performance on the EXPLORE reading test.

The standardized Beta Coefficient is used to compare the strength of the effect of each independent variable on the dependent variable. The independent variable with the largest standardized Beta (independent of sign) has the strongest effect. As expected, the Beta values of the regression analysis were consistent with the correlation coefficients found in the correlations analysis of OSTP and EXPLORE, and the effects of OSTP assessments to performance on the EXPLORE were relatively stable.

The Unstandardized Regression Coefficient (b-weight) indicates the amount of change in the dependent variable associated with a one unit change in the independent variable. The third-fifth grade math resulted in a constant value of -13.207 with the third grade having a b-weight of .013, the fourth grade had a b-weight of .01 and the fifth grade b-weight was .014. The sixth grade had a b-weight of .033 with a constant value of -9.424 while the seventh grade had a b-weight of .028 with a constant of -4.455.

Third and fourth grade reading had a constant of -14.139 with the third grade having a b-weight of .02 while the fourth grade b-weight was .017. Fifth and sixth grade reading combined had a constant of -16.762 with a fifth grade b-weight of .019 and a sixth grade b-weight of .022. Seventh grade reading had a b-weight of .035 and constant of -10.564.

The study further used linear regression to specify a functional relationship between OSTP and EXPLORE math and reading based on the grade comparison listed above so that EXPLORE math could be predicted from OSTP math and EXPLORE reading could be predicted from OSTP reading. SPSS software was used to calculate a regression equation by fitting a straight line or regression line to the points in a joint distribution, using the criterion of least squares. A prediction was accomplished in that for each value of OSTP math, the regression line gave the predicted value of EXPLORE math and for each OSTP reading, the regression line gave a predicted value of EXPLORE reading. Since the regression line represented the predicted value of EXPLORE math for each OPI score of OSTP math and the predicted value of EXPLORE reading for each OPI score of OSTP reading based on a prediction and not the actual values, the discrepancy between the actual values and the regression line constitutes an error in prediction of EXPLORE math/reading from OSTP math/reading. Therefore, the standard error in predicting EXPLORE math scores from OSTP math scores ranged from 2.28557 to 2.49862 and the standard error for predicting EXPLORE reading from OST reading scores ranged from 2.65337 to 2.85024. Table 5 gives a summary of this data.

Table 5:

Summary of Regression Analysis for Predicting EXPLORE Math and Reading Scores from OSTP Math and Reading Scores

Variable	B	SE B	ß
	Math		
Third-Fifth Grade	-13.182	1.289	.251, .212, .294
Sixth Grade	-9.424	1.396	.595
Seventh Grade	-4.455	.983	.645
	Reading		
Third-Fourth Grade	-14.139	1.671	.375, .273
Fifth-Sixth Grade	-16.762	1.573	.338, .367
Seventh Grade	-10.564	1.194	.666

Regression lines for specified subject and grades were used to predict EXPLORE scores from OPI scores. Table 6 illustrates the equations:

Table 6:

Equations for Predicting EXPLORE Math and Reading scores from OSTP Math and Reading scores

Grades	Equation
	Math
3-5	EXPLORE= -13.207 +.013(OPI) +.01(OPI) +.014(OPI)
6	EXPLORE= -9.424 + .033(OPI)
7	EXPLORE= -4.455 + .028(OPI)
	Reading
3-4	EXPLORE= -14.139 + .02(OPI) + .017(OPI)
5-6	EXPLORE= -16.762 + .019(OPI) + .022(OPI)
7	EXPLORE= -10.564 + .035(OPI)

The SPSS software program was used to estimate EXPLORE scores from appropriate OSTP scores. Table 7 demonstrates the findings. EXPLORE math score of 17 or above is considered college readiness. The national percentile rankings are also given by the EXPLORE Profile Summary Report for each numeric score received on the EXPLORE for each respective subject.

Table 7:

OSTP Third-Seventh Grade OPI Math Score ranges Predicting EXPLORE Math Scores Using Linear Regression

Grade 3-5 OPI Score	Grade 6 OPI Score	Grade 7 OPI Score	Explore Score	National Percentile Rank
400-640	400-603	400-534	1-10	1-12
641-667	604-634	535-569	11	17
668-694	635-664	570-605	12	23
695-721	665-694	606-641	13	31
722-748	695-724	642-676	14	41
749-775	725-755	677-712	15	53
776-802	756-785	713-748	16	64
803-829	786-815	749-784	17	75
830-856	816-846	785-819	18	83
857-883	847-876	820-855	19	89
884-910	877-906	855-891	20	92
910-990	907-990	892-990	21-25	95+

Based on these OPI scores, performance levels for the math portion of each grade could also be estimated. Some score ranges overlap performance levels as indicated in Table 8.

Table 8

OSTP Math Performance Levels Predicting EXPLORE Score Ranges and National Percentile Rank

OSTP Performance Level	EXPLORE Score Range	National Percentile Rank
	Grades Three-Five	
Unsatisfactory	1-9	1-9
Limited Knowledge	9-13	9-31
Satisfactory	13-17	31-75
Advanced	17-25	75-100
	Grade Six	
Unsatisfactory	1-12	1-23
Limited Knowledge	12-14	23-41
Satisfactory	14-16	41-64
Advanced	16-25	64-100
	Grade Seven	
Unsatisfactory	1-14	1-41
Limited Knowledge	14-15	41-53
Satisfactory	15-17	53-75
Advanced	17-25	75-100

A cross-tabulation analyses of seventh grade math and EXPLORE math was analyzed to check for accuracy of the prediction. Unsatisfactory on the seventh grade math test was predicted to relate to an EXPLORE score range of 1-14. Of the 103 students who scored unsatisfactory on the seventh grade OSTP math test, 85 scored in the predicted EXPLORE range; however, 11 students scored a 15 on the EXPLORE, 6 students scored in the 16-17 range and 1 student scored 18+ on the EXPLORE. Three students scoring unsatisfactory on the OSTP seventh grade math test and seven scoring limited knowledge on the same assessment scored a 17 or above on the EXPLORE math test, which is considered college readiness and in the top quartile in the nation. The OSTP and EXPLORE tests have a standard error of measurement, the relationship between OSTP and EXPLORE has a standard error of measurement as well as other variables not identified affect student performance from time to time.

Next, reading scores were evaluated. SPSS software was again used to determine the appropriate estimated EXPLORE scores from OPI scale scores. As stated above, due to OPI scales similar to performance levels, OSTP third and fourth grade reading were used to predict EXPLORE scores; OSTP fifth and sixth grade reading were used to predict EXPLORE scores; and, through a different OPI score range, OSTP seventh grade reading was used to predict EXPLORE scores. Table 9 illustrates the predicted EXPLORE reading scores from the OSTP reading assessment.

Table 9

OSTP Third-Seventh Grade OPI Reading Score Ranges Predicting EXPLORE Scores using Linear Regression. College Readiness Standard is 15+ for EXPLORE

Grades 3-4 OPI Score	Grades 5-6 OPI Score	Grade 7 OPI Score	EXPLORE Score	National Percentile Rank
400-665	400-664	400-601	1-10	1-16
666-692	665-689	602-630	11	29
693-719	690-713	631-658	12	42
720-746	714-738	659-687	13	54
747-774	739-762	688-716	14	64
775-801	763-786	717-744	15	73
802-828	787-811	745-773	16	79
829-855	812-835	774-801	17	85
856-882	836-860	802-830	18	88
883-909	861-884	831-858	19	91
910-936	885-908	859-887	20	94
937-990	909-990	888-990	21-25	95+

EXPLORE score ranges can also be predicted from OSTP performance levels for reading. Table 10 illustrates those predictions. An EXPLORE score of 15 or above is considered the College Readiness standard. National percentile ranks are given by EXPLORE for each numeric score on the EXPLORE. Some overlap exists in performance levels on the OSTP and EXPLORE scores as indicated below.

Table 10

OSTP Reading Performance Levels Predicting EXPLORE Score Ranges and National Percentile Rank-EXPLORE Score of 15 or Above Indicates College Readiness

OSTP Performance Level	EXPLORE Score Range	National Percentile Rank
	Grades Three-Four	
Unsatisfactory	1-9	1-7
Limited Knowledge	9-12	7-42
Satisfactory	12-18	42-88
Advanced	18-25	88-100
	Grades Five-Six	
Unsatisfactory	1-10	1-12
Limited Knowledge	10-12	12-42
Satisfactory	12-18	42-88
Advanced	18-25	88-100
	Grade Seven	
Unsatisfactory	1-13	1-54
Limited Knowledge	13-14	54-64
Satisfactory	14-17	64-85
Advanced	18-25	88-100

Therefore, in response to Research question #2, “Do the OPI scores and performance levels on the OSTP third-seventh grade math and reading predict national percentile ranks and college readiness standards as defined by the EXPLORE test? it was determined that OPI scores at each grade and subject were significantly associated with performance on the EXPLORE test. A comparison of the actual number of students to the predicted number of students indicates a higher level of accuracy in the middle range of scores (EXPLORE 13-18 and related OSTP OPI scores) than the lower end of the scale (EXPLORE 1-12 and related OPI score ranges) and the upper end of the scale (EXPLORE 19-25). The higher accuracy in the middle range of scores than in the lower and upper range of scores may indicate that the norm-referenced EXPLORE test is more sensitive in detecting lower and higher performing students than the criterion-references OSTP test.

The study continued into the OSTP Research questions, “Does OSTP performance at each grade (3-7) and subject (math and reading) follow a normal distribution?

The first method used to answer this question was the One-Sample Kolmogorov–Smirnov test on SPSS software. The findings of this test indicated that only third grade reading had a normal distribution of scores. Kurtosis and skewness were analyzed to determine the shape of the distribution of scores. SPSS defines kurtosis as the extent to which observations cluster around a central point. For a normal distribution, the value of the kurtosis statistic is zero. Positive kurtosis indicates that, relative to a normal distribution, the values are more clustered around the center of the distribution and have thinner tails until the extreme values of the distribution, at which point the tails are thicker relative to a normal distribution. Negative kurtosis indicates that, relative to a normal distribution, the observations cluster less and have thicker tails until the extreme values of the distribution, at which point the tails are thinner relative to a normal distribution.

SPSS defines skewness as a measure of the asymmetry of a distribution. The normal distribution is symmetric and has a skewness value of 0. A distribution with a significant positive skewness has a long right tail. A distribution with a significant negative skewness has a long left tail. As a guideline, skewness values more than twice its standard error is taken to indicate a departure from symmetry.

All grade and subject OSTP scores were analyzed with SPSS graphical plots and frequency distribution histograms were run with normal probability plots. Finally, distribution percentages in each standard deviation of the study were compared to the standard deviation distribution of normal distribution.

As expected, the data, charts and distributions were consistent in explaining how the data were non-normally distributed. The data revealed that all of the kurtosis values were in a positive direction ranging from .587 to 4.898 with a standard error of kurtosis of .202. These values indicate that the data set was clustered more about the center, have thinner tails until the extreme values, at which point the tails of the distribution are thicker relative to a normal distribution.

The results further revealed that skewness was slightly to moderately skewed across all grades and subjects. Grades 5 and 7 math were negatively skewed at -.345 and -.407 respectively while grades 4-6 reading were also negatively skewed at -.135, -.388 and -.183 respectively. The remaining grades and subjects were positively skewed with grade 3 math at .398, grade 4 math was skewed at .116 and sixth grade math was skewed at .178. Grade 3 reading was skewed at .286 and seventh grade reading was skewed at .133. Based on SPSS guidelines, a skewness value more than twice its standard error is taken to indicate a departure from symmetry; fourth grade math, fourth grade reading, sixth grade math, sixth grade reading, and seventh grade reading did not indicate a departure from symmetry. Third grade math, third grade reading, fifth grade math, fifth grade reading, and seventh grade math all had skewness less than ± ½ indicating the distribution was moderately skewed.

Although not a part of the research question, the EXPLORE math and reading data were also examined using the one-sample Kolmogorov-Smirnov test for a normal distribution of scores. EXPLORE math and reading also were determined to have a non-normal distribution of scores. EXPLORE math had a 1.091 kurtosis with a moderately negative skewness of .231 which was comparable to OSTP data for normal distribution. EXPLOE reading had a kurtois of -.230 indicating slightly less cluster around the center and having a thicker tail until the extreme values, at which point the tails are thinner relative to a normal distribution. EXPLORE reading had a positive skewness of .471, indicating a moderate right tail.

The mean and standard deviation of the sample data was compared to the state data for each OSTP grade and subject and EXPLORE subjects. This comparison revealed the sample group had a higher mean score and a lower standard deviation, as compared to the overall state population. Mean scores higher than the state mean and a standard deviation smaller than the state sample, offers evidence that although only 1 of 10 grade and subject analyses ended with a normal distribution of scores, it is possible that the sample group contained more average and above average scores than the overall state population.

Table 11 displays skewness, standard error of skewness, kurtosis, and standard error of kurtosis for the sample data. Table 12 displays the means and standard deviations of the data in this study compared to the means and standard deviations of the OSTP and EXPLORE statewide.

Table 11

Descriptive Statistics of OSTP and EXPLORE Math and Reading.

Grade	Skewness	SE of Skewnesss	Kurtosis	SE of Kurtosis
		Math		
3	.398	.101	1.056	.202
4	.116	.101	.867	.202
5	-.345	.101	1.364	.202
6	.178	.101	4.898	.202
7	-.407	.101	2.146	.202
EXPLORE	-.231	.101	1.091	.202
		Reading		
3	.28	.101	.587	.202
4	-.135	.101	1.633	.202
5	-.388	.101	.999	.202
6	-.183	.101	3.112	.202
7	.133	.101	.621	.202
EXPLORE	.471	.101	-.230	.202

Table 12

Means and Standard Deviations of OSTP Third-Seventh Grade Math and Reading Test and EXPLORE Math and Reading for this Study and Oklahoma Population.

Grade	M of this study	M of Oklahoma	SD of this study	SD of Oklahoma
		Math		
Three	749.08	734.68	61.54	80.26
Four	773.40	751.90	63.94	84.90
Five	777.51	764.20	63.07	75.40
Six	754.87	748.00	56.02	64.00
Seven	719.74	716.50	72.27	82.50
EXPLORE	15.51	14.60	3.11	3.40
		Reading		
Three	777.24	752.34	65.51	80.42
Four	791.96	770.10	57.40	79.60
Five	769.97	754.60	61.87	78.60
Six	759.98	753.50	58.71	67.80
Seven	740.02	729.60	68.28	72.10
EXPLORE	15.07	14.00	3.55	3.50

The distribution of scores was analyzed to determine the percent of students that fell within one standard deviation of the mean, two standards deviations from the mean, and three standard deviations from the mean on the OSTP math and reading and expectations on any normal distribution of scores. Table 13 illustrates the findings.

Table 13

Distribution of Scores for OSTP by Percentage in each Standard Deviation from the Mean and Comparison to Normal Distribution of Scores

Grade	-3 SD	-2SD	-1SD	M	+1SD	+2SD	+3SD	Total
				Math				
3	1.4	13.0	33.3	7	33.5	8.1	3.1	99.4
4	2.8	12.1	33.8	0	34.4	13.4	2.9	99.4
5	1.9	11.6	36.5	0	33.8	13.3	2.1	99.2
6	1.5	12.5	33.9	.2	40.1	9.7	.9	98.8
7	2.8	8.5	36.3	0	38.4	11.5	1.0	98.5
				Reading				
3	2.0	11.3	33.6	0	39.3	11.6	.7	98.5
4	2.2	10.1	38.4	0	32.4	13.1	2.6	98.8
5	2.5	10.5	32.7	0	42.7	9.5	1.2	99.1
6	1.7	9.9	37.6	0	35.3	13.8	.9	99.2
7	2.8	8.5	36.3	0	38.4	11.5	1.0	98.5
Normal	2.14	13.59	34.13	0	34.13	13.59	2.14	99.7

Although the scores were non-normally distributed, the percentages in each standard deviation from the mean are similar to the expectations of a normal distribution.

In response to OSTP Research Question #1: "Does OSTP performance at each grade (3-7) and subject (math and reading) follow a normal distribution"? only third grade reading met the Kolmogorov-Smirov test of normal distribution. Criterion test are typically not designed to have a normal distribution of scores; however an analyses of the data indicated the sample was relatively close in the percent of students in each standard deviation from the mean to the expected percent in each standard deviation from the mean for scores meeting the criterion for normal distribution. The data further indicated that the study sample had a higher mean and lower standard deviation than the Oklahoma population in general. Kurtosis analyses revealed more scores than normal hovered around the mean. This offers evidence either the sample was not representative (more average or above average students) of the state population or the OSTP criterion referenced test is not designed to be sensitive enough to separate scores in a manner that meets the criterion for normal distribution. Further investigation into the overall scores of the state is needed to make a final determination on this question. However, it does offer evidence that the known variability in abilities is present in the score distribution.

The last OSTP Research Question, “Do student performance levels on OSTP math and reading (grades 3-7) remain consistent from year-to-year?” was addressed next. To achieve a more specific understanding of the relationship between performances from year-to-year on the OSTP math and reading assessment, a descriptive statistical analysis was run using Cross Tabulations. This analysis investigated the relationship between students at each performance level from grades 3-7 in both math and reading beginning with third grade math. Two students originally performed in the unsatisfactory category, 99 at the limited knowledge level, 384 scored at the satisfactory performance level, and 101 at the advanced level. When these students were tested at the end of the fourth grade, one of the two students in the unsatisfactory level in the third grade moved into the limited knowledge category and one moved to satisfactory. Of the group's 99 students in the limited knowledge performance level, 2 dropped to the unsatisfactory level, 37 remained in the limited knowledge category, 59 moved up to the satisfactory level, and 1elevated to the advanced level. The 384 students scoring at the satisfactory level had 307 remain in that category, 23 drop to limited knowledge, and 54 classified as advanced. The 101 advanced students had 32 drop to the satisfactory level while 69 remain at the advanced level.

Following the fifth grade testing year, the two students that completed the third grade as unsatisfactory had one remain in this category while one continued on to the satisfactory level. The 99 students that began in the limited knowledge category had 44 remain in this category, 3 drop to unsatisfactory, 49 finish the fifth grade at the satisfactory level while 3 made it to the advanced category. The 384 students finishing the third grade at the satisfactory level had 271 remain in this category, 13 drop to limited knowledge, 1 fall to the unsatisfactory level and 99 categorized as advanced. The 101 students that ended the third grade as advanced math students, had 28 finish the fifth grade as satisfactory math students and 73 remain in the advanced category.

The two students finishing the third grade as unsatisfactory moved into the satisfactory category at the end of the sixth grade. Of the 99 students who finished the third grade in the limited knowledge category in math, 33 remained in this category at the end of the sixth grade, 13 dropped to the unsatisfactory level, 50 were considered satisfactory, and 3 were advanced following their sixth grade year. The 384 students originally testing satisfactory had 236 still in this category following their sixth grade year, 31 in limited knowledge, 6 unsatisfactory, and 111 advanced. The 101 third graders beginning as advanced had 20 drop into the satisfactory level and 81 remain as advanced following the sixth grade.

To summarize the data, 101 students or 17.2 % completed third grade math as less than proficient. Fifty-five of the 101 students, or 54%, attained the proficient category following the sixth grade. Four hundred eighty-five students or 82.8% began as satisfactory or advanced following the third grade. Four hundred forty eight or 92.4% of these students remained as satisfactory or advanced following the sixth grade. The 485 students originally testing satisfactory or advanced had 37 or 7.6% drop into the limited knowledge or unsatisfactory level.

As these third grade students completed the seventh grade, the standards were increased because of political pressure. The two students ending the third grade as unsatisfactory were back in this category. The 99 students scoring limited knowledge had 24 remain in this category, 46 drop to the unsatisfactory level, 25 advance to satisfactory, and 4 move to the advanced level. The 384 students completing the third grade at the satisfactory level had 55 drop to unsatisfactory, 62 limited knowledge, 193 remain satisfactory, and 74 move into the advanced classification. The 101 students classified as advanced following their third grade year had 71 remain as advanced and 30 drop to the satisfactory mark. Based on tougher standards, of the 101 students that first tested as less than satisfactory at the end of the third grade, only 29% were able to reach satisfactory standards by the end of the seventh grade. Despite the tougher standards, of the 485 students that first tested as satisfactory or above at the end of the third grade, 395 or 81% of them still had satisfactory or above math skills at the end of the seventh grade.

Further findings of the study revealed progress was made in mathematics overall for students in the study, at least through the sixth grade; however, the satisfactory and above students made the most progress. In mathematics, 18 fewer students were considered as less than satisfactory in the sixth grade compared to the third grade and 94 more students were classified as advanced based on the same grade comparison. The more rigorous standards put in place following the students seventh grade year resulted in 88 more students classified as less than satisfactory compared to their third grade testing year; however, despite the more rigorous requirements, 48 more students were considered advanced following their seventh grade year compared to their third grade year.

The crosstab calculations display a standardized residual number. If numbers remain in categories as expected, the residuals should fall between ±2. If residuals fall outside of ±2, they are considered unusual. A positive standardized residual number indicates a stronger likelihood of falling in a category and a negative number indicates a less likelihood of being in a particular category. A comparison of third to seventh grade students indicated that students that began as unsatisfactory had a 2.8 residual or strong likelihood of remaining in that category. Students who began as limited knowledge had a 6.9 residual or very strong likelihood of being unsatisfactory at seventh grade, a 2.5 residual or strong likelihood of still being limited knowledge, a -2.6 residual or less likely of being satisfactory, and a -4.2 residual or highly unlikely of being advanced. Students who began as satisfactory had a 2.4 residual or strong likelihood of still being satisfactory as seventh graders and a -2.4 residual or less likely to be advanced as seventh graders. Third grade students who began as advanced as third graders had a -4.2 residual to be unsatisfactory and a -3.9 residual to be limited knowledge, or highly unlikely to drop to these categories while they had an 8.9 residual or highly likely to remain as advanced as seventh graders.

Table 14:

Cross Tabulation Analyses of Third and Seventh Grade Performance Levels on OSTP Math

Performance Level		Unsatisfactory	Limited Knowledge	Satisfactory	Advanced	Total
		Seventh Grade OSTP Math				
Unsatisfactory	Count	2.0	0.0	0.0	0.00	2
	Expected	.4	.3	.8	.50	2
	St. Resid.	2.8	-.5	-.9	-.70	
Limited Knowledge	Count	46.0	24.0	25.0	0.0	99
	Expected	17.4	14.5	41.8	25.20	99
	St. Resid.	6.9	2.5	-2.6	-4.42	
Satisfactory	Count	55.0	62.0	193.0	74.00	384
	Expected	67.5	56.4	12.5	97.60	384
	St. Resid.	-1.5	.8	2.4	-2.40	
Advanced	Count	0.0	0.0	30.0	71.00	101
	Expected	17.8	14.8	42.7	25.70	101
	St. Resid	-4.2	-3.9	-1.9	8.90	

A descriptive analysis using cross tabulation was next run on OSTP reading scores. The sample used in this study had 5 students categorized as unsatisfactory at the end of the third grade, 59 as limited knowledge, 483 as satisfactory, and 39 recognized as advanced on the OSTP scale.

The end of the fourth grade revealed that the five students classified as unsatisfactory had four classified as satisfactory while one moved up to limited knowledge. The 59 students classified as limited knowledge at the end of the third grade had 3 drop to unsatisfactory, 15 remain at the same level, and 41 in the satisfactory category. The 483 students classified as satisfactory at the end of the third grade tested had the following fourth grade results: 1 was unsatisfactory, 8 were limited knowledge, 461 remained at satisfactory, and 13 moved to advanced. Twenty-three of the 39 students testing as advanced in the third grade remained advanced and 16 dropped to satisfactory.

As the third graders tested at the end of the fifth grade, the five that were originally unsatisfactory had two as satisfactory, two as limited knowledge and one as unsatisfactory. The 59 limited knowledge students dispersed to 6 as unsatisfactory, 24 as limited knowledge, and 29 as satisfactory. The 483 students beginning as satisfactory had 4 drop to unsatisfactory, 34 categorized as limited knowledge, 440 remain as satisfactory and 5 move to the advanced level. The 39 advanced third grade students had 36 drop to satisfactory and 3 remain as advanced.

As the third graders moved into the sixth grade and middle school, the five original students classified as unsatisfactory had three categorized as limited knowledge and two as satisfactory. The 59 students that began their OSTP testing career as limited knowledge had 1 drop to unsatisfactory, 27 remain as limited knowledge, and 31 advance to satisfactory. The 483 students in the satisfactory category had 441 remain as satisfactory, 37 drop to limited knowledge and 3 drop to unsatisfactory while 2 met the advanced requirement. The original 39 advanced students had 34 classified as satisfactory and 5 remaining as advanced at the end of the sixth grade.

The 64 students classified as less than satisfactory following the third grade OSTP exam had 36 or 56% attain the satisfactory or advanced category following the sixth grade. The 522 students originally classified as satisfactory or above readers in the third grade had 482 or 92% of them remain as satisfactory or advanced readers following the sixth grade.

The more rigorous standards implemented following these students' seventh grade year resulted in the five students originally classified as unsatisfactory having three move up to satisfactory, one remain as limited knowledge, and one still considered unsatisfactory. The 59 students originally classified as limited knowledge had 11 fall to unsatisfactory, 33 remain as limited knowledge, and 15 move to satisfactory. The 483 beginning their testing career as satisfactory readers had 14 drop to unsatisfactory, 79 fall to limited knowledge while 383 remain as satisfactory readers and 7 reach the advanced category. The 39 readers first classified as advanced, had 1 classified as limited knowledge, 34 satisfactory, and 4 advanced at the end of the seventh grade.

A summary of the students over four years reveals that of the 64 students originally considered as less than satisfactory readers following the third grade, only 18 or 28% were considered satisfactory or advanced readers following the seventh grade. The 522 students ending the third grade as satisfactory readers or above, had, 428, or 82% considered satisfactory o advanced following the seventh grade. This was with more rigorous standards in place.

Reading scores from third to sixth grade revealed a slight decline in the number of students satisfactory (522 students or 89% were satisfactory or above following the third grade compared to 515 students or 87.8% following the sixth grade) with the number declining to 436 or 76% following the more rigorous standards put in place following the seventh grade testing year.

Third to seventh grade OSTP reading was analyzed to determine the categories that had a strong likelihood of students continuing in that category. Students who began the third grade as limited knowledge category had a 5.2 or strong likelihood that they would drop to the unsatisfactory level as seventh graders, a 6.4 or strong likelihood to remain as limited knowledge and a -4.4 or highly unlikely to be categorized as satisfactory as seventh grade students.

Table 15

Cross Tab Analyses Between Third and Seventh Grade Reading Performance Levels on the OSTP

Performance Level		Unsatisfactory	Limited Knowledge	Satisfactory	Advanced	Total
Seventh Grade OSTP Reading						
Unsatisfactory	Count	1.0	1.0	3.0	0.0	5
	Expected	.2	1.0	3.7	.1	5
	St. Resid.	1.7	.0	-.4	-.3	
Limited	Count	11.0	33.0	15.0	0.0	59
Knowledge	Expected	2.6	11.5	43.8	1.1	59
	St. Resid.	5.2	6.4	-4.4	-1.1	
Satisfactory	Count	14.0	79.0	383.0	7.0	483
	Expected	21.4	94.0	358.5	9.1	483
	St. Resid.	-1.6	-1.5	1.3	-.7	
Advanced	Count	0.0	1.0	34.0	4.0	39
	Expected	1.7	7.6	29.0	.7	39
	St. Resid	1.3	2.4	.9	3.8	

In response, to the second OSTP Research Question, "Do student performance levels on OSTP math and reading remain consistent from year-to-year (grades 3-7)?" many students change performance levels as they move from the third-seventh grade in both math and reading. However, the changes of most interest were the third grade students who tested less than satisfactory on the OSTP and those who tested satisfactory or above on the OSTP.

The math and reading numbers were very close in these categories. Fifty-four percent of the third grade students scoring below satisfactory in math reached satisfactory or above by the end of sixth grade year. However, when the tougher standards were applied in the seventh grade, only 29% of the third grade students classified as less than satisfactory reached the satisfactory or above level following their seventh grade year. The third grade math students classified as satisfactory or above had better outcomes. Ninety-two percent testing as satisfactory or above in the third grade were still considered satisfactory or above by the end of the sixth grade. Following the tougher standards being applied in the seventh grade, 81% of the students that began as satisfactory or above at the end of the third grade in math were still considered satisfactory or above at the end of the seventh grade,

The third grade students classified as less than satisfactory in reading had 56% of these students reach satisfactory or above by the end of the sixth grade. However, when the tougher standards were applied in the seventh grade, only 28% of the third grade students classified as less than satisfactory were able to reach the satisfactory or above level at the end of the seventh grade. The third grade reading students that were classified as satisfactory or above in the third grade had better outcomes. Identical to the outcomes in math, 92% of the students testing as satisfactory or above in third grade reading were still considered satisfactory or above by the end of the sixth grade. Following the tougher standards being applied in the seventh grade, 82%

of the students that began as satisfactory or above at the end of the third grade in reading were still considered satisfactory or above at the end of the seventh grade,

Summary of Findings

A correlation analysis between the OSTP math and reading assessment and the EXPLORE math and reading found a statistically significant positive association between the assessments. The strength of the association, with slight exception, increased each year as students got closer to the time of the EXPLORE assessment. The correlations ranged from .535 to .666 which is considered moderate to strong in association. These correlations led to the use of linear regression to predict EXPLORE scores from OSTP assessments with the purpose to allow some concrete meaning to be applied to the OSTP OPI scores in terms of national percentile rankings and college readiness standards.

Students categorized as satisfactory in math for grades 3-5 fell in the national percentile range of 31%-75% while advanced students were in the top quartile nationally; sixth grade math students categorized as satisfactory fall in the 41%-64% range while advanced students were 64%-100%. The more rigorous standards put in place following these students seventh grade year resulted in students in the satisfactory range falling in the national percentile range of 53%-75% while advanced students were in the top quartile. In most grades in math, college readiness standards were met in the upper range of satisfactory.

Students categorized as satisfactory in reading for grades 3-6 fell in the national percentile range of 42%-88% with advanced readers in the 88%-100% range. The more rigorous standards put in place following these students seventh grade year resulted in satisfactory students falling in the 64%-85% range and advanced students ranged from 88%-100%. In reading, college readiness standards were met in the middle range of the satisfactory category for grades 3-6 while seventh grade college readiness standards were met in the lower range of the satisfactory category.

A Komogorov-Smirnov test was applied to each OSTP grade and subject assessment study to determine the distribution of scores. Only third grade reading met the normal distribution of scores. Skewness and kurtosis test were conducted to determine the criteria for failing to meet normal distribution. These tests indicated the sample had a higher cluster of scores around the mean than normal distribution, a fact confirmed by SPSS graph analyses (See Appendix E). However, a comparison of the means and standard deviations of the studies sample to the overall state population indicated the sample had a slightly higher mean score and lower standard deviation from the mean than the overall state population. Criterion referenced test such as the OSTP typically do not have a normal distribution of scores. The OSTP ranks students into four performance levels indicating the test is designed to separate student abilities, although possibly not sensitive enough to result in a normal distribution. Although, the scores were non-normally distributed, the distribution of scores does offer evidence of the known variability in human abilities.

Finally, the study tracked student performance levels both in math and reading from grades 3-7 to determine the progress being made, particularly students below the satisfactory level. The findings indicated slightly over 50% of the students who tested below satisfactory in the third grade had obtained a satisfactory or above score by the end of the sixth grade. Over 90% of the students who tested as satisfactory or above at the end of the third grade were still at that level by the end of the sixth grade. As these students completed the seventh grade, the criteria for reaching each performance level increased. The tougher standards resulted in only slightly over 28% of the students who ended the third grade as less than satisfactory attaining the satisfactory or above level by the end of the seventh grade. Despite the increased standards, over 80% of the students who ended the third grade as satisfactory or above, remained at those levels at the end of the seventh grade.

Progress was made in math achievement from the third to the sixth grade with 18 fewer students classified as less than satisfactory and 94 more students classified as advanced following the sixth grade testing year compared to the original test at the end of the third grade. The more rigorous standards required following the students seventh grade year resulted in 88 more students classified as less than satisfactory compared to the third grade year; however, 48 more students attained the advanced classification than when they first tested. Reading scores did not reveal progress from the third to the sixth grade in terms as the number of students classified as less than satisfactory increased slightly.

Chapter V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary

No Child Left Behind (NCLB) requires each state to establish challenging academic standards for all students in reading and math and to test students annually to see if they are reaching these standards. The term “proficiency” is used by NCLB to describe the level of achievement students are required to meet. The bar is raised each year until 2014 when all students are required to reach the “proficiency” level as defined by each state (Carey, 2007). Critics of NCLB (Kahlenberg, 2008; Rothstein et al., 2006; and Linn, 2006) contend that no standard can be challenging to typical and advanced students and achievable by lower performing students.

NCLB gives states broad discretion to determine what students must learn, how to test, and how to define “proficiency.” Carey (2006) and Cronin et al. (2007) contend many states have taken advantage of this autonomy to make their educational performance look much better than it really is, based on a comparison of the National Assessment of Educational Progress (NAEP) and state testing in math and reading. Many state test results indicate a higher percent of students reaching proficiency levels than NAEP tests confirm.

External validity is necessary for purposes of accountability. Such validity is the extent test performance is related to another valued, independent, and direct measure of which the test is designed to assess (Lally, 2000). One form of external validity is predictive validity, a measure of whether achievement on one assessment can predict achievement on some future assessment of the same content and skills (Sattler, 1992). NAEP has been the predominant instrument to determine the external validity of state test. Since one of the limitations of NAEP is that it does not track individual students and looks only at scores during grades 4 and 8, it is difficult to use in tracking the outcomes of education reforms and policies (Yeager, 2007).

Of the study's two purposes, the first was to define "proficiency levels" (advanced, satisfactory, limited knowledge, and unsatisfactory) and Oklahoma Performance Index (OPI) scores on the Oklahoma State Testing Program (OSTP) third-seventh grade math and reading tests in a manner meaningful to parents, educators, and policy makers. Defining proficiency levels and OPI scores in a meaningful manner was accomplished by determining the relationship between student achievement on the OSTP and the EXPLORE test in math and reading. The relationship allowed the researcher to use linear regression to estimate OPI scores and performance levels as national percentile ranks and college preparedness (EXPLORE gives national percentile ranks and college readiness benchmarks for each score on the exam).

The second purpose examined if the NCLB requirement of all students attaining "proficiency" by 2014 is realistic. Two methods were employed to evaluate this proposition. OPI score distribution within each grade and subject were examined to determine if a normal distribution was present. Finally, the study tracked students OPI scores from third to seventh grade in math and reading and examined the effectiveness of current remediation practices to determine if student achievement is improving, particularly students classified as unsatisfactory and limited knowledge.

The first purpose of the study, defining proficiency levels and OPI scores for grades 3-7 math and reading in a meaningful manner (national percentile ranks and college readiness), was based on a theoretical framework offered by the statistical concept of predictive validity. OEPAS (2004) found that consistency existed between the Oklahoma Priority Academic Student Skills (PASS) standards/objectives and the Educational Planning and Assessment System (EPAS) standards/objectives. The study used predictive validity to postulate that if the PASS standards and their related assessment (OSTP), and the EPAS standards and their related assessment (EXPLORE) were consistent, congruent validity between the two assessments should exist. Furthermore, performance on the OSTP assessments should have predictive validity for performance on the EXPLORE tests due to this congruency. The EXPLORE test offers benchmarks for success in college.

The study's second purpose, examining the realism of the NCLB requirement of all students reaching the proficiency level by 2014, was guided by a theoretical framework supported by normal probability distribution. Normal distribution occurs frequently in statistics and in the natural and social sciences and infers that 68% of all values lie within one standard deviation from the mean, 95.4% of all values lie within two standard deviations from the mean and 99.7% of all values lie within three standard deviations from the mean (Experiment Resources, 2010). A distribution of scores illustrating the known variability in human ability offers evidence it is unlikely that a test challenging to average and above average students would be attainable by lower performing students.

Two predictive validity research questions were used to guide the study's first purpose, defining OPI scores and performance levels in a meaningful manner (national percentile ranks and college readiness standards):

1. Are OPI scores in the OSTP third-seventh grade math and reading statistically significantly related to performance in math and reading on the eighth grade EXPLORE test?
2. Do the OPI scores and performance levels on the OSTP third-seventh grade math and reading assessment predict national percentile ranks and college readiness standards as defined by the EXPLORE test?

Two OSTP research questions guided the study's second purpose, examining the realism of the NCLB requirement of all students reaching the proficiency level by 2014:

1. Does OSTP performance at each grade (3-7) and subject (math and reading) follow a normal distribution?
2. Do student performance levels on the OSTP math and reading remain consistent from year-to-year (grades 3-7)?

A comparison of performance levels was used to evaluate if remediation efforts improved student achievement, particular students scoring limited knowledge and unsatisfactory, to determine if students were moving towards the NCLB requirement of all students reaching proficiency by 2014.

The study's population was derived from eighth grade students' third-seventh grade OSTP math and reading scores and related eighth grade EXPLORE scores. Students selected were from 12 public school districts in Oklahoma with a middle school (sixth-eighth grade) configuration ranging in size from 120 to 650 students with an aggregate average API score of 1310 compared to the state average of 1279. The selected districts had less mobility of students (7.2% compared to the state average of 10.2 %) and slightly better attendance (94.4% compared to the state average of 94.2%).

Data related to the OSTP math and reading test as well as the EXPLORE test were stored in each district's administrative offices. Permission to obtain data was received from each district's superintendent. By having an identification number assigned to each student, data received by the researcher was anonymous and absent student names and identifying information.

A review of the database was performed to identify any missing assessments. All identification numbers that did not have math and reading test scores for each grade (3-7) of OSTP and EXPLORE for grade 8 associated with them were eliminated from the study. This process resulted in a reduction of the sample population from 1,150 scores to 586 sets of scores. Factors possibly causing the missing data included student mobility (data either did not follow the student as it should have or was placed in an area not discovered by the person responsible for providing the data to the researcher), absenteeism, and students originally taking the regular OSTP assessment and later categorized into a learning disability category that allowed them to take an alternative assessment. Only students taking the regular

OSTP math and reading exam were included in the study. All data were placed in a SPSS program for analysis.

Based on the assessment data, the research found OSTP three-seventh math and reading assessment had a moderate to strong effect in predicting performance on the EXPLORE math and reading assessment. Typically, as the OSTP assessments came closer to the timing of the EXPLORE assessment the ability to predict EXPLORE scores from OSTP scores became stronger. Grades within subjects were combined when OPI scale ranges were similar to performance levels to explain a higher percentage of the variance between assessments. The OSTP OPI scores explained from 36%-46% of the variance based on the grade and subject examined.

The purpose of the study was to give guidance based on an estimate. In terms of national percentile ranks, the approximate unsatisfactory performance level for grades 3-5 in math ranged from 1%-9% with the sixth grade unsatisfactory category ranging from 1%-23%, limited knowledge for third-fifth grade math ranged from 9%-31% while the sixth grade limited knowledge category ranged from 23%-41%. The OSTP satisfactory performance level for third-fifth math ranged from 31%-75%, while sixth grade students in the satisfactory category ranged from 41%-64%. The students classified as advanced on the OSTP for grades 3-5 were in the top quartile nationally while sixth grade advanced students were in the 64-100 national percentile ranking. The rigor for reaching satisfactory increased following these students' seventh grade year. Unsatisfactory seventh grade students fell in the 1%-41% range, limited knowledge students were in the 41%-53% area, satisfactory students fell between 53%-75% and advanced students were in the top quartile nationally. Some

EXPLORE scores overlapped with OSTP performance levels at several grades and subjects. College readiness standards were met in the upper ranges of OPI scores in the satisfactory performance level for math in most grades.

The approximate national percentile ranking according to the association with EXPLORE for the unsatisfactory category for reading for third and fourth grade students were 1%-7%; grades 5-6 were 1%-12% while seventh grade was 1%-54%. The limited knowledge category in reading for third and fourth grade students were 7%-42%, grades 5-6 were 12%-42% while the increased rigor for the seventh grade resulted in a range of 54%-64%. The satisfactory reading students in grades 3-6 were in the 42%-88% range while seventh grade satisfactory students fall in the 64%-85%. Advanced students in reading for all grades fell in the 88%-100% range and college readiness standards were met in the middle OPI score range of the satisfactory category for grades 3-6 while seventh grade students in the lower range of satisfactory typically met the college readiness standard for reading.

The study's next purpose was to offer evidence the NCLB requirement that all students reach a proficiency level by 2014 is unrealistic. The study examined the distribution of scores at each grade and subject of OSTP math and reading. Only the OSTP third grade reading scores met the criteria on normal distribution. However, criterion referenced tests such as the OSTP typically are not designed to provide a normal distribution of scores. The OSTP does separate students into four performance categories indicating the assessment is designed to identify differences in students' abilities. The distribution of scores is non-normally distributed; however, the data indicates the distribution of scores has similar characteristics to a normal distribution. The study's sample had a higher mean OPI score and a smaller standard deviation than the overall state population. Examining kurtosis and skewness of the sample indicated failure to meet normal distribution was based on more scores clustering around the mean. Two possible explanations are the OSTP is not designed to be sensitive enough to differentiate scores into a normal distribution or the sample has more average and above average students than the state population. Further investigation is needed to determine if OSTP scores meet the criteria for normal distribution. The distribution of scores did offer substantial evidence of the known variations in human ability which was one of the goals of the study.

The second method used to offer evidence of the difficulty of all students attaining the proficiency requirement of NCLB was to track student progress from third-seventh grade in math and reading, specifically those students less than satisfactory and students satisfactory or above in third grade in math and reading. The results indicated that students moved in and out of performance levels from grade-to-grade and subject-to-subject. However, of the students identified as less than satisfactory in the third grade in both math and reading, slightly over half attained satisfactory performance levels by the end of their sixth grade year. The study further found that over 90% of the students who were satisfactory or above in the third grade were still at that level following the sixth grade.

The increased standards implemented following the seventh grade testing had the most negative effect on students originally scoring as less than satisfactory in the third grade. Slightly over 28% of the students classified as less than satisfactory following the third grade OSTP math and reading assessments were able to meet the criterion of satisfactory or above as defined by the new OSTP standards. Over 80% of the students classified as satisfactory or above following the third grade OSTP math and reading test still met this criterion following the new standards put in place after their seventh grade testing year.

The study further found progress was made in mathematics for most students in the study, at least through the sixth grade; however, the satisfactory and above students made the most progress. In mathematics, 18 fewer students were considered as less than satisfactory in the sixth grade compared to the third grade and 94 more students were classified as advanced based on the same grade comparison. The more rigorous standards implemented following the students' seventh grade year resulted in 88 more students classified as less than satisfactory compared to their third grade testing year; however, despite the more rigorous requirements, 48 more students were considered advanced following their seventh grade year compared to their third grade year. Reading did not reveal the same progress as math in academic achievement; however, the more rigorous standards had the most detrimental effect on the students classified as less than satisfactory.

Conclusions

The study's first purpose was to define OPI scores and "proficiency levels" on the third-seventh grade math and reading OSTP in a manner that is meaningful to parents, educators, and policy makers. The study was successful in this endeavor by providing estimates for OPI scores that offer a comparison to their peers nationally (national percentile ranks) and scores indicative of college readiness. The study does not define proficiency any more than NCLB (NCLB only offers language similar to

NAEP in defining “proficiency” and suggest it must be rigorous). However, it does allow parents and educators a better understanding of their students’ current progress and a benchmark for college readiness. Policy makers who will ultimately determine cut scores for proficiency have more concrete bench marks to guide their decisions. The benchmarks include using OPI scores that give a traditional definition of grade level based on a national comparison of peers and also allow guidelines on OPI scores that meet college readiness standards.

However, it is important to understand that the relationship offers only an estimate. OSTP and EXPLORE each has a standard error of measurement and the relationship between them has a standard error of measure. Also other factors not identified in this study might affect student performance at various times.

An example of this error is demonstrated with the following cross-tabulation analyses of seventh grade math and EXPLORE math. Unsatisfactory on the seventh grade math test was predicted to relate to an EXPLORE score of 1-14. One-hundred three students scored unsatisfactory on the seventh grade OSTP math test with 85 of the students scoring in the predicted EXPLORE range; however, 11 students scored a 15 on the EXPLORE, 6 students scored in the 16-17 range and 1 student scored 18+ on the EXPLORE. Three students scoring unsatisfactory on the OSTP seventh grade math test and seven scoring limited knowledge on the same exam, scored a 17 or above on the EXPLORE math test, which is considered college readiness and in the top quartile in the nation.

The second purpose of the study examined if the NCLB requirement of all students attaining “proficiency” by 2014 is realistic. The study provided considerable evidence the known variations in student ability, the current system of remediation, and the increased standards make this NCLB goal near impossible.

The study examined the distribution of scores in grades 3-7 math and reading to offer evidence the known variations in student ability make it unlikely a test challenging to average and above average students would be attainable by lower performing students. Although only the third grade reading test had a normal distribution of scores, which was unexpected, a criterion test such as the OSTP typically is non–normally distributed. However; the OSTP does classify students into four performance categories indicating the test is designed to differentiate between student abilities-although possibly not sensitive enough to provide a normal distribution of scores. The sample had a larger percentage of students clustering around the mean than a normal distribution. Two possible explanations are the sample may not be representative of the overall state sample or the OSTP is not designed to be sensitive enough to spread scores in a normal distribution. The sample in the study had a higher mean score and a smaller standard deviation than the state population which provides further evidence the study's sample was not necessarily representative of the state population. However, the data had a similar percentage of students in each standard deviation from the mean to a normal distribution of scores indicating the known difference in human abilities was present, the determination of which was one of the goals of the study. Further study is needed to determine if a sample more representative of the state population would meet the normal distribution criteria.

Due to the OSTP not having a normal distribution of scores and a larger percentage of students clustering around the mean, the outcomes were not evenly distributed along a “line of best fit.” Such a finding is not inconsistent with the goals of criterion referenced tests (Helfant, 2005), such as the OSTP. Criterion-referenced tests are intended to measure how well a person has learned a specific body of knowledge and skills (Sattler, 1992). It is possible for all students to earn a passing score if they master the material tested. Norm-referenced tests are made to compare test takers. The principle of norm-referenced tests implies that half will score above a midpoint and half score below.

The second method of offering evidence the NCLB requirement of all students attaining proficiency by 2014 is unrealistic was tracking student performance from grades 3-7. Particular focus was on students originally classified as less than satisfactory and students classified as satisfactory or above in the third grade. The study found progress was made in mathematics for most students in the study, at least through the sixth grade; however, the satisfactory and above students made the most progress. Reading did not reveal the same progress made in math in academic achievement; however, the more rigorous standards had the most detrimental affect on the less than satisfactory students.

Slow progress occurred in math towards the NCLB goal of all students being proficient by 2014. Students considered proficient increased from 82.7% to 85.8% from third to sixth grade indicating slightly over 1% growth per year. This pace would require an additional 14 years from 2008 indicating 2022 as a possible date for all students to be proficient with current levels of progress and the old testing standards in place. The more rigorous standards left only 67.7% of students considered satisfactory or above, indicating the goal of 100% proficiency is not remotely on the horizon.

Reading progress actually had a slight decline from 89% classified as satisfactory following the students' third grade testing year to 87.8% following their sixth grade year. The more rigorous standards following the students seventh grade year left only 76% classified as satisfactory or above readers, indicating the goal of all students being proficient in reading by 2014 is simply unrealistic.

The study found the vast majority of students classified as satisfactory or above following the third grade would remain satisfactory or above in middle school, despite the increased rigor of the test. The students classified as less than satisfactory in the third grade had slightly over half reach satisfactory following the sixth grade; however, the more rigorous standards following the seventh grade year pulled about half of these students below the satisfactory level.

Oklahoma may have met the goal of having tests considered rigorous and of avoiding the criticism of making scores look better than they are. However, more Oklahoma students are now considered as less than proficient than when NCLB first began. The results also reveal that students struggling by the end of the third grade are the most affected by increased standards. The study offers evidence the vast majority of students originally classified as satisfactory or above will succeed despite the increased standards. The students NCLB intended to help the most are currently the most negatively affected by the increased standards. Although not an intended purpose of the study, the high probability of future success for those classified as satisfactory or above by the end of the third grade compared to the dismal outlook for those classified as less than satisfactory by the end of the third grade provides support for early childhood education and other early intervention programs.

In drawing further conclusions from the statistical findings, limitations of the study need to be considered. Although the sample size of 586 was ample, the sample was not representative of the overall state population. A more diverse group of student abilities would have allowed for a "line of better fit" producing a higher percentage of variance between exams and, therefore, more accurate predictions of EXPLORE scores from OSTP OPI scores. A more diverse group could possible create a normal distribution of scores, offering further evidence of the known variability in human ability, making a single test challenging to the most gifted students and attainable by lower achieving students unrealistic.

One test such as the OSTP or the EXPLORE cannot reflect the total knowledge and abilities of all students. Kornhaber (2004) explains, "the decision to retain students, assign them to remedial classes, deny students their high school diploma, withhold school funds, or reconstitute a school staff should not be based on the results of a single test" (p. 60). In this study, the affects of the OSTP OPI scores accounted for 36%-46% of the variance in EXPLORE scores. Other factors, unaccounted for in this study, explain the remaining variance.

Recommendations

Based on the findings and conclusions of the research study, several recommendations are offered. First, further studies should be done with a population more representative of Oklahoma to see if the findings can be replicated. In this way, researchers can determine if the findings are irregular or representative of other similar studies' findings.

A second recommendation is that researchers consider other variables that influence student achievement as measured by the OSTP and EXPLORE tests. Specific attention might be given to the students who improved from the less than satisfactory proficiency levels to satisfactory or above proficiency levels as they moved from the third grade to the sixth and seventh grades. Over half the students improved from less than satisfactory to satisfactory by the end of the sixth grade and slightly over 28% still were considered satisfactory or above by the end of the seventh grade even with more stringent standards. The students considered satisfactory or above at the end of the third grade yet dropping below the satisfactory level at sixth and seventh grades is another area to investigate. Although over 90% of the students classified as satisfactory or above at the end of third grade were still satisfactory or above following the sixth grade, and over 80% of the students classified as satisfactory or above at the end of the third grade were still satisfactory or above following the seventh grade; almost 10% and 20% respectively, did not maintain their performance. Determining the variables associated with students failing to maintain their previous success would be worthy of further study.

Another recommendation is for researchers to investigate the benefits of increasing the rigor of tests to student performance. The study's data indicated that most of the students classified satisfactory or above continued to succeed even with higher standards. However, the group most negatively affected by increased standards was comprised of students originally classified as less than satisfactory. Amrein & Berliner (2002) investigated the 18 states with the highest stakes testing programs and found only slight indications of gains in learning on the NAEP, college entrance exams, or advanced placement tests.

Further research should be conducted on correlation of teacher grades to student performance. Findings that teachers are already identifying students that are unsatisfactory, limited knowledge, satisfactory, and advanced with teacher made assessment could save millions of dollars in testing that could be invested in instructional improvement.

The final recommendation is for researchers and policy-makers to work together to find alternatives to the requirement that all students reach proficiency by the 2013-2014 school year. Wallis & Steptoe (2007) contend, "Decreeing that all kids (except 1% with serious disabilities and an additional 2% with other issues) must be proficient by 2014 is a little like declaring that all the children are above average in the mythical town of Lake Wobegon" (p. 23).

Linn (2006) offers the following approach. Define a cut score on a state assessment equivalent to a median score in a base year. An example of 2002 is given for a base year. The percentage of students scoring above that cut score would then be used to monitor improvement in student performance with target increases set at reasonable levels, (e.g., 3%) per year. The portion of students scoring above the 2002 median would improve from 50% to 86% in 2014, a dramatic improvement in achievement that is not totally unrealistic.

Wallis & Steptoe (2007) contend “most educators prefer a more flexible measure of student improvement known as the growth model. In this approach, schools track the progress of each student year-to-year. Success is defined by a certain amount of growth, even if the student isn’t on grade level” (p. 34). Linn (2006) explains that current growth models are approved under NCLB. However, he contends the most constraining principle is the model must ensure all students be proficient by 2013-2014.

Kornhaber (2004) states standardized testing has been widely accepted because it is perceived to address a broad array of educational problems (e.g’s low standards, weak student motivation, poor curriculum and instruction, inadequate learning, and educational inequality). Business people and politicians realize an educated society is crucial to retain and attract businesses. Testing should offer accountability; be informative to students, parents, educators, and policymakers; and answer the George Bush question, “Is our children learning” (Kurtzman, n.d., p.1)?

Final Comment

NCLB set public schools up to fail by its original design. A criterion requiring all students to pass a test considered challenging and rigorous offers only two logical outcomes. If all students reach the goal, the test is too easy—if all students do not reach the goal, public schools failed. Researchers have offered more than ample evidence NCLB is unrealistic in the expectation all students pass a test considered rigorous. However, so far, politicians have refused to change NCLB to help all students succeed. Other research has based claims NCLB was unrealistic in expecting all students to pass a rigorous test on known variations in student ability, providing evidence that students from top performing countries in the world cannot meet the 100% proficiency goal, and progress that has been made on NAEP tests in past history. This research adds to the body of knowledge by providing empirical data that OPI score distribution illustrates the known differences in human abilities, and current remediation efforts and methods are insufficient to meet the 2014 timeline.

NCLB uses the term “research-based” as a criterion for using federal money to improve student or teacher performance. However, the framers of NCLB failed to follow their own advice (base decisions on proven research) when researchers provided sufficient evidence that aspects of NCLB would not work. Too many politicians choose to stick to their original beliefs when the facts indicate otherwise.

Testing needs to offer educators and parents reliable data on the strengths and weaknesses of students, teachers, and the school overall. However, unrealistic expectations, with punishment for failing, offer no logical benefits.

Oklahoma recently increased its criteria for students attaining “proficiency” in math and reading. The decision appears to be based on political pressure from groups contending the previous tests were too simple to pass; based on comparisons to NAEP. The Oklahoma State Department of Education (OSDE) decision to increase standards ignored the research showing considerable evidence that NAEP is flawed and the OSDE provided no research supporting increased standards result in improved student achievement. Oklahoma, like the framers of NCLB, makes many important decisions affecting the lives of students and educators based on opinion, without supporting research. More tragically, when research supports the opposite of their personal beliefs or agenda, often times it is ignored.

Today's technology offers unprecedented opportunities to track students from elementary school to college. Massive data banks should be created to store student information and track student progress. The data would allow the opportunity to determine appropriate test scores at milestones that are indicators of future success. Students identified as making unusual gains in achievement could be qualitatively studied to guide appropriate and successful interventions. However, the opportunity to use the data will be of no benefit, if politicians continue to ignore it because it conflicts with their personal beliefs.

The testing data has the opportunity to identify programs and methods of instruction improving student achievement. The current accountability system is used to criticize public education in hopes that highlighting their failures will shame educators into improving instruction or provide support to dismantle public education. Testing and accountability can offer great opportunities if used to identify and reinforce effective teachers and programs, eliminate ineffective teachers and instructional methods, and provide information to parents on their child's abilities and progress. Simply identifying failing students and schools is only reiterating what is already known.

References

ACT (2003). *Preparing Oklahoma students for the future.* Retrieved March 22, 2009, from http://www.act.org/research/policymakers/pdf/oklahoma.pdf

ACT (2005). *Crisis at the core: Preparing all students for college and work.* Retrieved April 2, 2009, from http://www.act.org/research/policymakers/pdf/crisis_report.pdf

ACT (2008). *College readiness standards for EXPLORE, PLAN, and the ACT.* Retrieved February 22, 2009, from http://www.act.org/standard/pdf/CRS.pdf

Amrein, A. L. & Berliner, D. C. (2002, March). High-stakes testing, uncertainty, and student learning, *Education Policy Analysis Archives, 10*(18). Retrieved April 17, 2010 from http://epaa.asu.edu/epaa/v10n18/

Association for Educational Communications and Technology. (2001). *Elementary and secondary act of 1965 (ESEA).* Retrieved March 4, 2009, from http://www.aect.org/About/History/esea.htm

Braswell, J. S., Lutkus, A. D., Grigg, W. S., Santapau, S. L., Tay-lim, B. S., & Johnson, M. S. (2001). *The nation's report card: Mathematics 2000.* Washington, DC: National Center for Education Statistics.

Brewer, D. J., & Kileen, K. M. (2009). *Organizing effective educational accountability: The case of Oklahoma.* Retrieved February 8, 2009, from http://obecinfo.com/downloads/OKEdAccountabilityReportFinal.pdf

Calkins, K. G. (2005). *Applied statistics-lesson 5: Correlation coefficients*. Retrieved April 16, 2010, from http://www.andrews.edu/~calkins/math/edrm611/edrm05.htm

Carey, K. (2006, May). Hot air: How states inflate their progress under NCLB. *Education Sector.* Retrieved January 31, 2009, from http://www.educationsector.org/usr_doc/Hot_Air_NCLB.pdf

Carey, K. (2007, November). The pangloss index: How states game the No Child Left Behind Act. *Education Sector*. Retrieved January 31, 2009, from http://www.educationsector.org/usr_doc/The_Pangloss_Index.pdf

Cronin, J., Dahlin, M., Adkins, D., & Kingsbury, G. G. (2007). *The proficiency illusion*. Washington, DC: The Fordham University.

Choudhurry, A. (2009). *Statistical correlation-Strength of relationship between variables*. Retrieved April 16, 2010, from http://www.experiment-resources.com/statistical-correlation.htm

DeFehr, J. (2009). *Oklahoma state testing program and achieving classroom excellence* u*pdate* [PowerPoint slides]. Retrieved February 13, 2010, from http://www.sde.state.ok.us/Services/Conference/Leadership/Sessions/55_76_ OSTPACEUpdate.ppt#268,12, Standard Setting deKlerk, G. (2008). *Classical test theory (CTT): Online readings in testing and assessment.* Retrieved April 4, 2010, from International Test Commission Lecture Notes Online web site: http://www.intestcom.org/Publications/ORTA/Classical+test+theory.php

Education Week (2004, September 21). No child left behind. *Education Week.* Retrieved January 9, 2009, from http://www.edweek.org/rc/issues/no-child-left-behind/

Experiment Resources (2010). *Normal distribution* [PDF document]. Retrieved April 7, 2010, from Lecture Notes Online Web site: http://www.experiement-resources.com/normal-probability-distributionn.htm.

EXPLORE (2008). *Using your EXPLORE results 2008/2009*. Retrieved May 2, 2009, from http://www.act.org/explore/pdf/UsingResults.pdf

Finn, C. E., Julian, L., & Petrilli, M. J. (2006). *The state of state standards.* Retrieved March 3, 2009, from *http://www.edexcellence.net/doc/State% 20of%20State %20Standards2006FINAL.pdf*

Fuller, B., Gesicki, K., Kang, E., & Wright, J. (2006). *Is the No Child Left Behind Act working? The reliability of how states track achievement*. Retrieved March 1, 2009, from University of California, Policy Analysis for California Education Web site: http://www.gse.berkeley.edu/research/ PACE/reports /WP.06-1.pdf

Helfant, M. T. (2005). *The relationship between third and fourth grade Everyday Mathematics assessments and performance on the New Jersey assessment of skills and knowledge in fourth grade. (NJASK/4).* Retrieved from ProQuest Digital Dissertations. (UMI No. 3226835)

Hoberock, B. (2009, April 30). Henry vetoes bill to move education testing. *Tulsa World*. Retrieved February 13, 2009, from: http://www.tulsaworld.com/site/printerfriendlystory.aspx?articleid=20090430_16_A1_OLHMIY788733

Jorgensen, M. A., & Hoffman, J. (2003). *History of the no child left behind act of 2001 (NCLB).* Retrieved March 3, 2009, from http://www.pearson assessments.com/NR/rdonlyres/D8E33AAE-BED1-4743-98A1-BDF4D49D7274/0/HistoryofNCLB_Rev2_Final.pdf

Kahlenberg, R. D. (2008). *Fixing No Child Left Behind*. New York, NY: The Century Foundation.

Kornhaber, M. L. (2004). Appropriate and inappropriate forms of testing, assessment, and accountability. *Education Policy, (18)*1, 45-70.

Kurtzman, D. (N.D.). Top 10 Bushisms: The stupidest things President George W. Bush as ever said. Retrieved, May 1, 2010 from About.com Political Humor website http://politicalhumor.about.com/cs/georgewbush/a/top10bushisms.htm

Lally, M. O. (2000). *A construct validity study of the New Jersey assessment (EWT), grade eight proficiency assessment (GEPA) and elementary school proficiency assessment (ESPA).* Unpublished doctoral dissertation, Seton Hall University, South Orange, NJ.

Linn, R. L. (2006). *Improving the accountability provisions of NCLB.* Boulder, Colorado, University of Colorado, Center for Research on Evaluation, Standards, and Student Testing.

Linn, R. L. & Baker, E. L. (1996). *Assessing the validity of the national assessment of educational progress*: *NAEP technical review panel white paper*. Los Angeles, California, University of California, National Center for Research on Evaluation, Standards, and Student Testing. Retrieved August 26, 2009, from http://research.cse.ucla.edu/Reports/TECH416.PDF

Loveless, T. (2007). *How well are American students learning?* (The 2007 Brown Center Report on American Education, Volume II Number 2). Washington, DC: The Brookings Institution.

Martin, R. D. (2007). *Administrators response to "leaders and laggards"* [Powerpoint slides]. Retrieved March 6, 2009. from Oklahoma Technical Assistance Center website:http://www.csdcotac.org/media/Response%20to %20Leaders%20and%20Laggards/Response%20to%20Leaders%20 and%20Laggards_files/Default.htm

National Center of Education Statistics. (2007). *Mapping 2005 state proficiency Standards onto the NAEP scales* (NCES Publication No. 2007-482). Washington, DC: U.S. Department of Education.

National Center of Educational Statistics (2009). *Trends in international mathematics and science study (TIMMS).* Retrieved March 10, 2009, from http://nces.ed.gov/timss/

No Child Left Behind Act. (2001). *PL107-110.*Washington DC: United States Department of Education.

Oklahoma Educational Planning and Assessment System. (2004). *Comparing the EPAS standards to the Oklahoma Priority Academic Student Skills (Pass).* Retrieved February 17, 2009 from http://www.okhighered.org/epas/pass.shtml

Oklahoma School Testing Program. (2008). *Test interpretation manual Grades 3-8 Oklahoma core curriculum tests.* Retrieved April 16, 2009, from http://www.sde.state.ok.us/AcctAssess/pdf/forms/TIMGr3_8.pdf

Oklahoma State Regents for Higher Education. (1999).*Oklahoma educational planning and assessment system (EPAS).* Oklahoma City, OK.

Oklahoma State Senate (2007). *Report of the achieving classroom excellence II task force.* Retrieved April 2, 2009 from http://www.oksenate.gov/publications/ issue_papers/education/ace_II_task_force_report.html

Philips, G. W. (2007). *Expressing international educational achievement in terms of U.S. performance standards: Linking NAEP achievement levels to TIMSS,* Washington, DC: American Institutes for Research.

Reckase, M. (2001). *The controversy over the national assessment governing board standards.* Washington, DC: Brookings Institution Press.

Rotherman, A. J. (2006, July). Making the cut: How states set passing scores on standardized tests. *Education Sector*. Retrieved February 3, 2009 from http://www.educationsector.org/usr_doc/EXPCutScores.pdf

Rothstein, R., Jacobsen, R., & Wilder, T. (2006). ‘Proficiency for all’—An oxymoron. Campaign for Educational Equity. *Examining America’s commitment to closing achievement gaps: NCLB and it alternatives* (pp. 1-76). New York: Columbia University.

Sattler, J. M. (1992). *Assessment of Children.* San Diego, CA: Jerome M. Sattler Publications Inc.

Shavelson, R. J. (1996). *Statistical reasoning for the behavioral sciences* (3rd ed.). Massachusetts: Pearson Educational Company.

Somerville, J., and Yi, Y. (2002). *Aligning k-12 curriculum and postsecondary expectations: State policy in transition*. Washington, DC: National Association of System Heads.

Study says state tests set low bar for students (2009, January 15). *Newsok.com.* Retrieved March 7, 2009, from http://www.newsok.com/article/3338205/

United States Chamber of Commerce (2007, February). *Leaders and laggards: A state by state report on educational effectiveness.* Washington, DC: Institute for a Competitive Workforce.

United States Department of Education (USDOE). (2002). *No Child Left Behind: A desktop reference.* Jessup, JD: Educational Publications Center.

United States Department of Education (USDOE). (2004). *Improving America’s Schools Act*, Public Law 103-382. Retrieved March 4, 2009, from www.ed.gov/legislation/ESEA/toc.html

Wallis, C. & Steptoe, S. (2007, June 4). How to fix No Child Left Behind. *Time*, *169,* 34.

Yeager, M. (2007). Understanding *NAEP:* Inside the nation's education report card. *Education Sector*. Retrieved April 14, 2009, from http://www.educationsector.org/usr_doc/EXPNAEP.pdf

APPENDIX A

CORRELATIONS

Correlations for Table 2, p. 74

Math (N=586)

	Mean	Std. Deviation
Grade 3	749.0836	61.54035
Grade 4	773.4044	63.94345
Grade 5	777.5119	63.06998
Grade 6	754.8720	56.01656
Grade 7	719.7440	72.27095
EXPLORE	15.5085	3.10733

Correlations

		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	EXPLORE
Grade 3	Pearson Correlation	1	.702**	.681**	.644**	.583**	.600**
	Sig. (1-tailed)		.000	.000	.000	.000	.000
	N	586	586	586	586	586	586
Grade 4	Pearson Correlation	.702**	1	.741**	.682**	.652**	.606**
	Sig. (1-tailed)	.000		.000	.000	.000	.000
	N	586	586	586	586	586	586
Grade 5	Pearson Correlation	.681**	.741**	1	.698**	.683**	.622**
	Sig. (1-tailed)	.000	.000		.000	.000	.000
	N	586	586	586	586	586	586
Grade 6	Pearson Correlation	.644**	.682**	.698**	1	.704**	.595**
	Sig. (1-tailed)	.000	.000	.000		.000	.000
	N	586	586	586	586	586	586
Grade 7	Pearson Correlation	.583**	.652**	.683**	.704**	1	.645**
	Sig. (1-tailed)	.000	.000	.000	.000		.000
	N	586	586	586	586	586	586
EXPLORE	Pearson Correlation	.600**	.606**	.622**	.595**	.645**	1
	Sig. (1-tailed)	.000	.000	.000	.000	.000	
	N	586	586	586	586	586	586

** Statistically Significant

Correlations for Table 3, p. 74
Reading (N=586)

	Mean	Std. Deviation
Grade 3	777.2440	65.50911
Grade 4	791.9590	57.40257
Grade 5	769.9693	61.87002
Grade 6	759.9761	58.71233
Grade 7	740.0154	68.28003
EXPLORE	15.0700	3.55281

Correlations

		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	EXPLORE
Grade 3	Pearson Correlation	1	.698**	.698**	.669**	.613**	.566**
	Sig. (1-tailed)		.000	.000	.000	.000	.000
	N	586	586	586	586	586	586
Grade 4	Pearson Correlation	.698**	1	.716**	.667**	.642**	.535**
	Sig. (1-tailed)	.000		.000	.000	.000	.000
	N	586	586	586	586	586	586
Grade 5	Pearson Correlation	.698**	.716**	1	.665**	.678**	.582**
	Sig. (1-tailed)	.000	.000		.000	.000	.000
	N	586	586	586	586	586	586
Grade 6	Pearson Correlation	.669**	.667**	.665**	1	.640**	.592**
	Sig. (1-tailed)	.000	.000	.000		.000	.000
	N	586	586	586	586	586	586
Grade 7	Pearson Correlation	.613**	.642**	.678**	.640**	1	.666**
	Sig. (1-tailed)	.000	.000	.000	.000		.000
	N	586	586	586	586	586	586
EXPLORE	Pearson Correlation	.566**	.535**	.582**	.592**	.666**	1
	Sig. (1-tailed)	.000	.000	.000	.000	.000	
	N	586	586	586	586	586	586

**Statistically Significant

APPENDIX B

REGRESSION

Regressions for Table 4, p. 77; Grades Three-Five Math (p=.05)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.680[a]	.462	.459	2.28557	.462	166.430	3	582	.000

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2608.198	3	869.399	166.430	.000[a]
	Residual	3040.259	582	5.224		
	Total	5648.457	585			

UNSTANDARDIZED COEFFICIENT, STANDARD ERROR AND BETA FOR Table 5 (P. 80)

Grades Three-Five Math

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-13.207	1.289		-10.246	.000
	Grade 3	.013	.002	.251	5.538	.000
	Grade 4	.010	.002	.212	4.291	.000
	Grade 5	.014	.002	.294	6.099	.000

Regressions for Table 4, p. 77; Grade 6 Math (p=.01)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	Df1	df2	Sig. F Change
1	.595[a]	.355	.353	2.49861	.355	320.757	1	584	.000

ANOVA

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	2002.507	1	2002.507	320.757	.000[a]
	Residual	3645.950	584	6.243		
	Total	5648.457	585			

UNSTANDARDIZED COEFFICIENT, STANDARD ERROR AND BETA FOR Table 5 (P. 80) Grade SIX Math

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-9.424	1.396		-6.751	.000
	Grade 6	.033	.002	.595	17.910	.000

Regressions for Table 4, p. 77; Grade 7 Math (p=.01)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.645[a]	.416	.415	2.37629	.416	416.305	1	584	.000

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2350.765	1	2350.765	416.305	.000[a]
	Residual	3297.693	584	5.647		
	Total	5648.457	585			

UNSTANDARDIZED COEFFICIENT, STANDARD ERROR AND BETA FOR TABLE 5 (P. 80)

GRADE SEVEN MATH

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-4.455	.983		-4.531	.000
	Grade 7	.028	.001	.645	20.404	.000

Regressions for Table 4, p. 77; Grades Three-Four Reading (p=.05)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.599[a]	.359	.356	2.85024

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2647.902	2	1323.951	162.970	.000[a]
	Residual	4736.229	583	8.124		
	Total	7384.131	585			

UNSTANDARDIZED COEFFICIENT, STANDARD ERROR AND BETA FOR TABLE 5 (P. 80)

GRADES THREE-FOUR READING

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	T	Sig.
1	(Constant)	-14.139	1.671		-8.462	.000
	Grade 3	.020	.003	.375	8.095	.000
	Grade 4	.017	.003	.273	5.898	.000

Regressions for Table 4, p. 77; Grades Five-Six Reading, (p=.05)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.644[a]	.414	.412	2.72411	.414	206.031	2	583	.000

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3057.816	2	1528.908	206.031	.000[a]
	Residual	4326.316	583	7.421		
	Total	7384.131	585			

UNSTANDARDIZED COEFFICIENT, STANDARD ERROR AND BETA FOR TABLE 5 (P. 80)

GRADES FIVE-SIX READING

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-16.762	1.573		-10.658	.000
	Grade 5	.019	.002	.338	7.977	.000
	Grade 6	.022	.003	.367	8.643	.000

Regressions for Table 4, p. 77, Grade 7 Reading (p=.01)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.666[a]	.443	.442	2.65337

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3272.539	1	3272.539	464.823	.000[a]
	Residual	4111.593	584	7.040		
	Total	7384.131	585			

UNSTANDARDIZED COEFFICIENT, STANDARD ERROR AND BETA FOR TABLE 5 (P. 80)

GRADES SEVEN READING

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	T	Sig.
1	(Constant)	-10.564	1.194		-8.847	.000
	Grade 7	.035	.002	.666	21.560	.000

APPENDIX C

KOLMOGOROV-SMIRNOV TEST

Nonparametric Tests (Kolmogorov-Smirnov Test)

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of opi3_math is normal with mean 749.084 and standard deviation 61.54.	One-Sample Kolmogorov-Smirnov Test	.006	Reject the null hypothesis.
2	The distribution of opi3_read is normal with mean 777.244 and standard deviation 65.509.	One-Sample Kolmogorov-Smirnov Test	.130	Retain the null hypothesis.
3	The distribution of opi4_math is normal with mean 773.404 and standard deviation 63.943.	One-Sample Kolmogorov-Smirnov Test	.027	Reject the null hypothesis.
4	The distribution of opi4_read is normal with mean 791.959 and standard deviation 57.403.	One-Sample Kolmogorov-Smirnov Test	.001	Reject the null hypothesis.
5	The distribution of opi5_math is normal with mean 777.512 and standard deviation 63.07.	One-Sample Kolmogorov-Smirnov Test	.021	Reject the null hypothesis.
6	The distribution of opi5_read is normal with mean 769.969 and standard deviation 61.87.	One-Sample Kolmogorov-Smirnov Test	.031	Reject the null hypothesis.
7	The distribution of opi6_math is normal with mean 754.872 and standard deviation 56.017.	One-Sample Kolmogorov-Smirnov Test	.005	Reject the null hypothesis.
8	The distribution of opi8_read is normal with mean 759.976 and standard deviation 58.712.	One-Sample Kolmogorov-Smirnov Test	.001	Reject the null hypothesis.
9	The distribution of opi7_math is normal with mean 719.744 and standard deviation 72.271.	One-Sample Kolmogorov-Smirnov Test	.007	Reject the null hypothesis.
10	The distribution of opi7_read is normal with mean 740.015 and standard deviation 68.28.	One-Sample Kolmogorov-Smirnov Test	.042	Reject the null hypothesis.
11	The distribution of explore_math is normal with mean 15.509 and standard deviation 3.107.	One-Sample Kolmogorov-Smirnov Test	.000	Reject the null hypothesis.
12	The distribution of explore_read is normal with mean 15.07 and standard deviation 3.553.	One-Sample Kolmogorov-Smirnov Test	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

APPENDIX D

Math SKEWNESS AND KURTOSIS

SKEWNESS AND KURTOSIS DATA FOR TABLE 11, (P. 90)

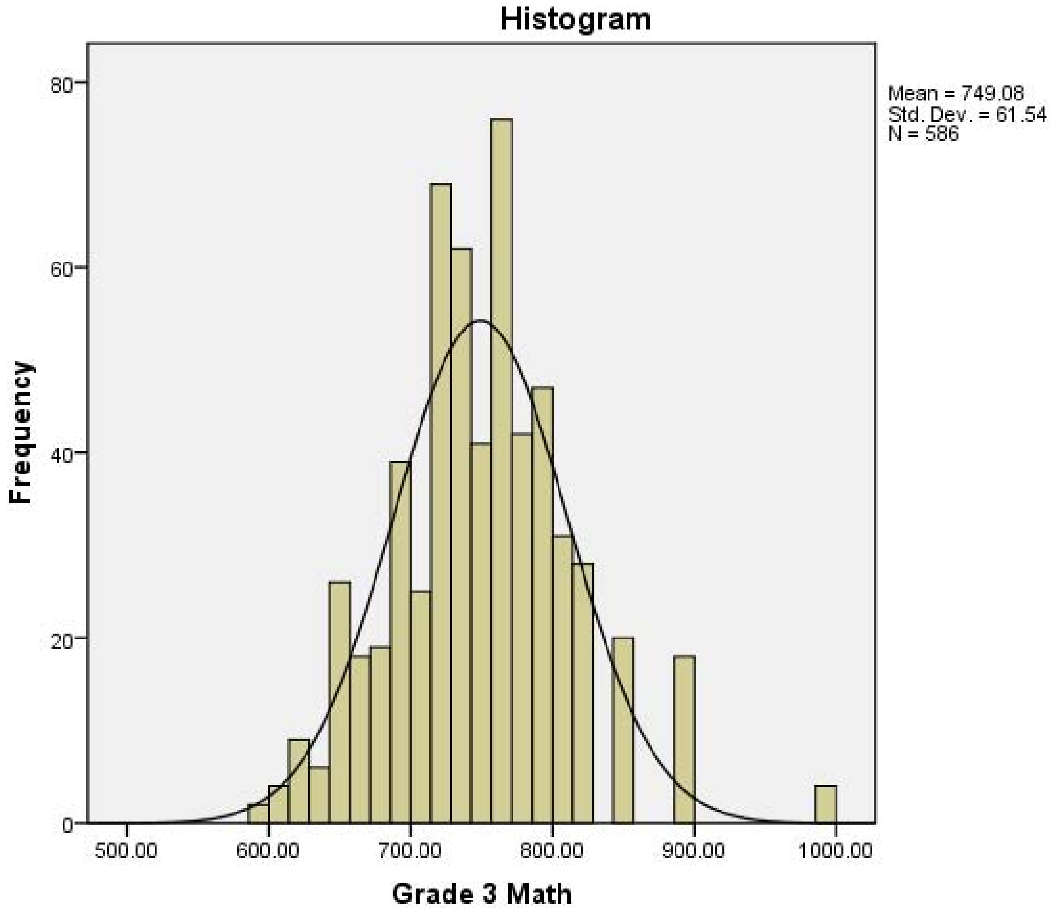
MATH

Statistics

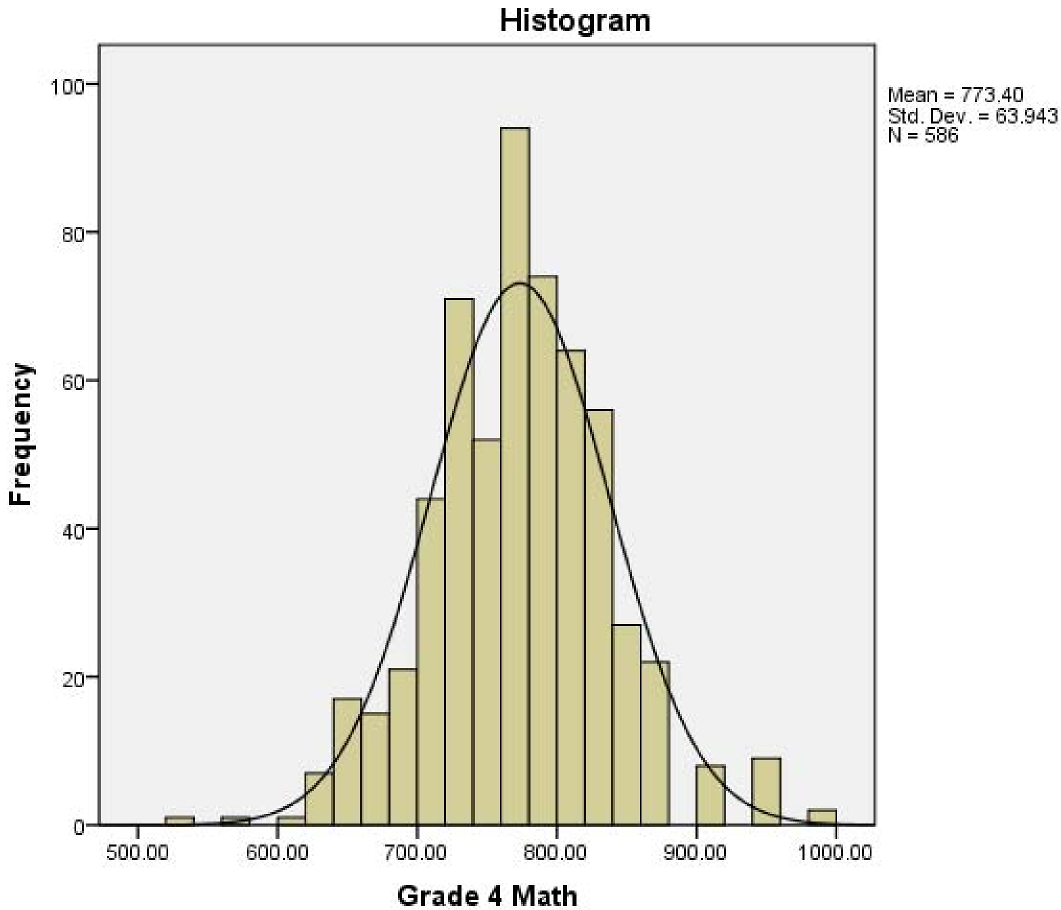
		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7
N	Valid	586	586	586	586	586
	Missing	0	0	0	0	0
Mean		749.0836	773.4044	777.5119	754.8720	719.7440
Median		749.0000	777.0000	775.0000	757.0000	724.0000
Mode		792.00	794.00	755.00	762.00[a]	708.00
Std. Deviation		61.54035	63.94345	63.06998	56.01656	72.27095
Skewness		.398	.116	-.345	.178	-.407
Std. Error of Skewness		.101	.101	.101	.101	.101
Kurtosis		1.056	.867	1.364	4.898	2.146
Std. Error of Kurtosis		.202	.202	.202	.202	.202

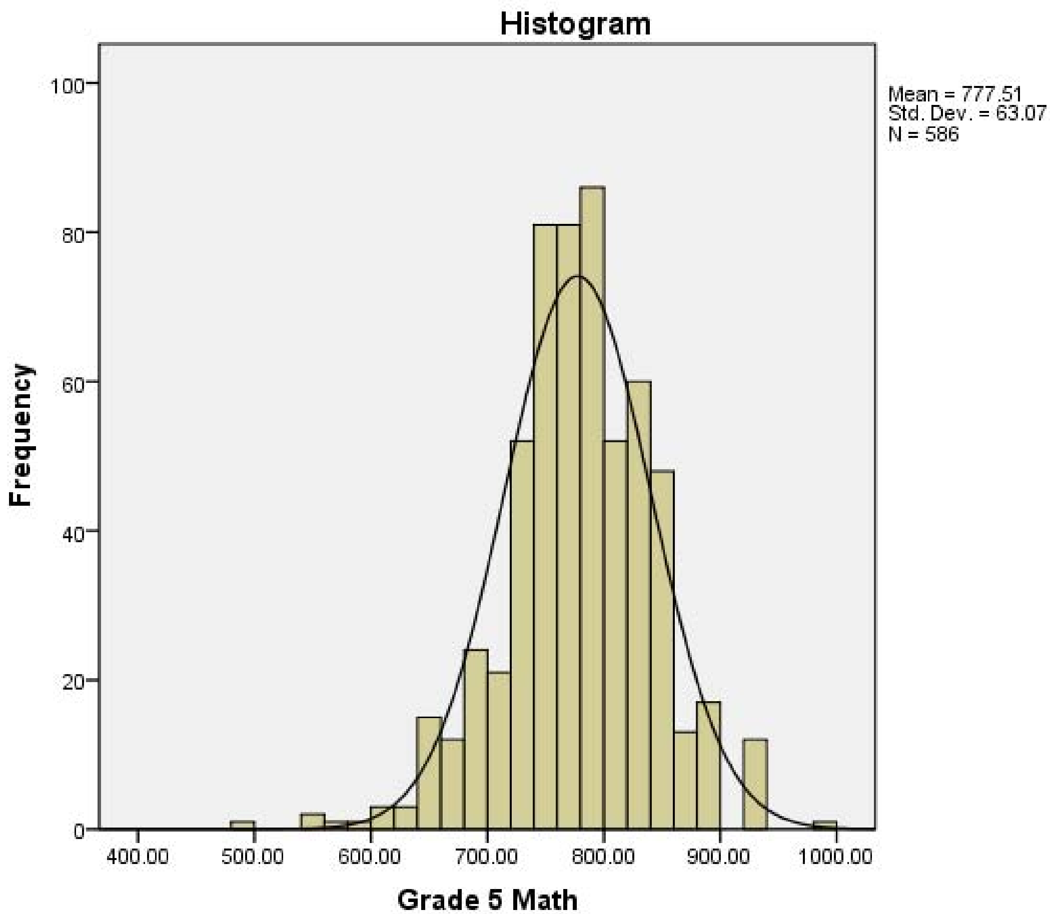
APPENDIX E

HISTORGRAMS FOR MATH, GRADES THREE-SEVEN

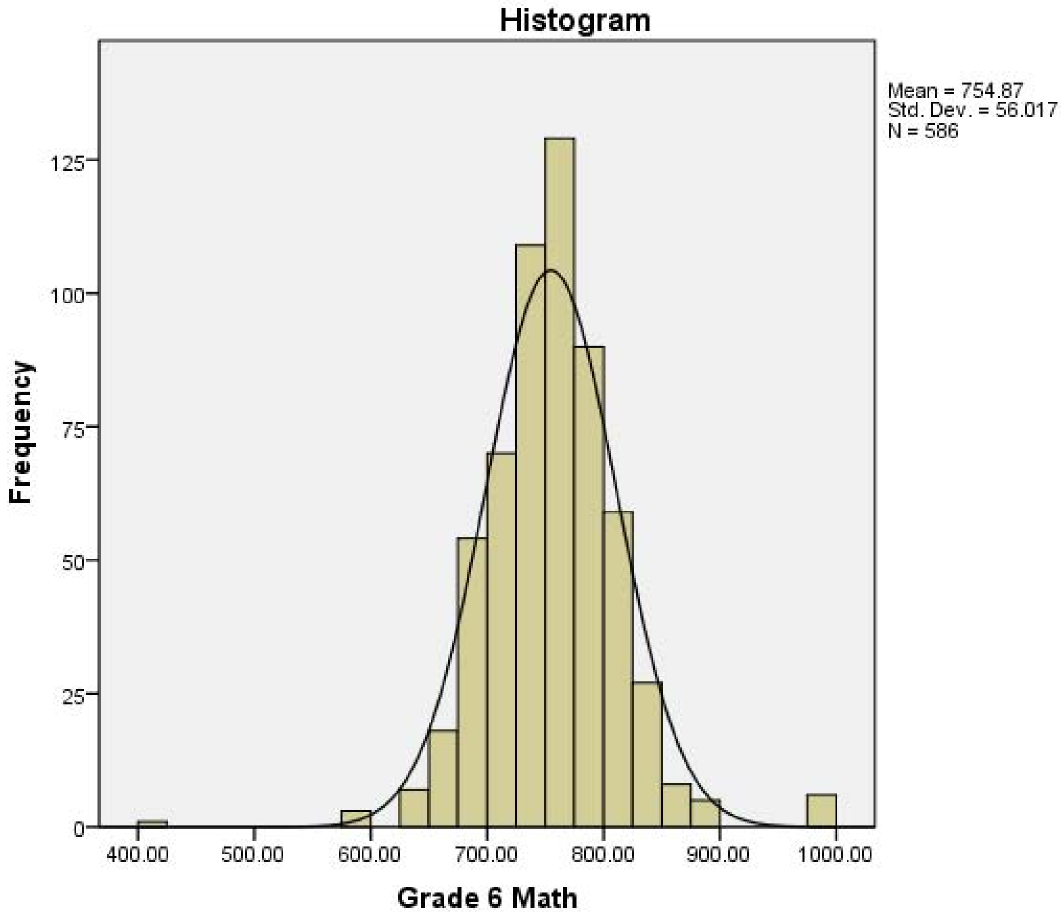

Histogram
Mean = 749.08
Std. Dev. = 61.54
N = 586
Frequency
0
20
40
60
80
500.00
600.00
700.00
800.00
900.00
1000.00
Grade 3 Math

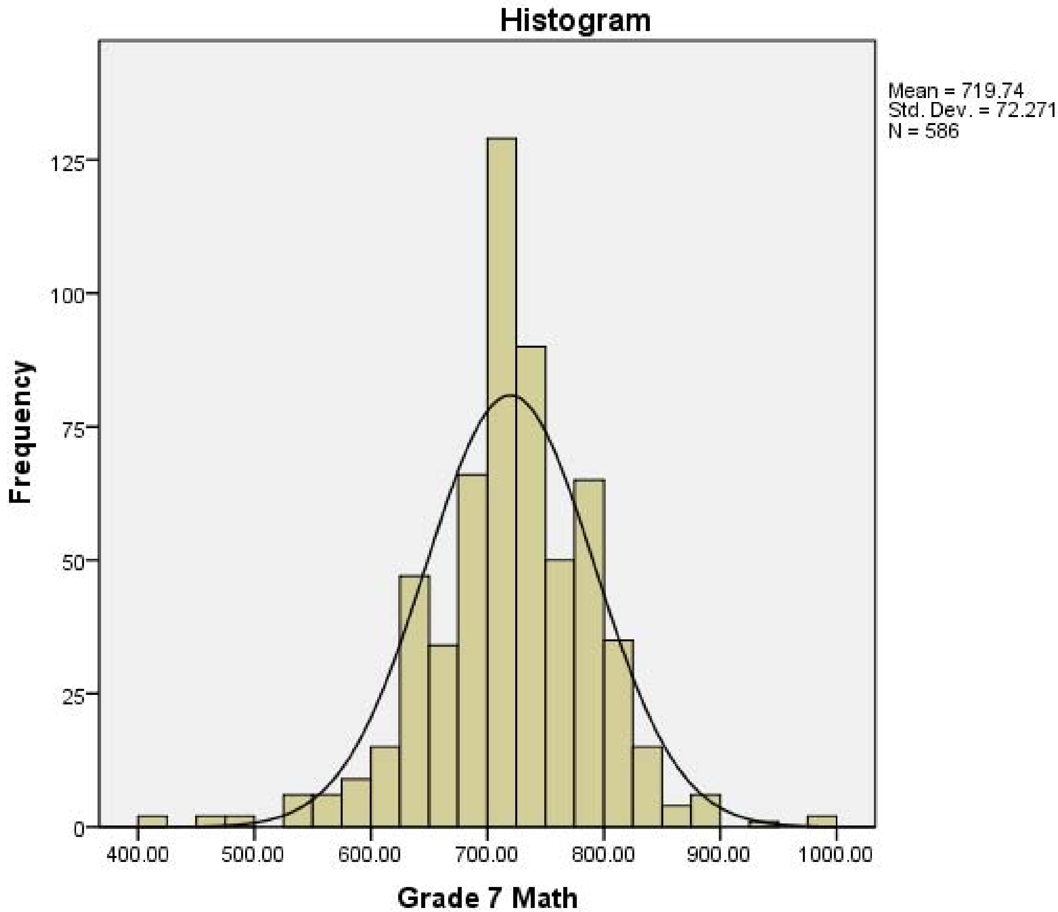
.


Histogram
Mean = 773.40
Std. Dev. = 63.943
N = 586
Frequency
100
80
60
40
20
0
500.00
600.00
700.00
800.00
900.00
1000.00
Grade 4 Math



Histogram
Mean = 777.51
Std. Dev. = 63.07
N = 586
Frequency
0
20
40
60
80
100
400.00
500.00
600.00
700.00
800.00
900.00
1000.00
Grade 5 Math


Histogram
Mean = 754.87
Std. Dev. = 56.017
N = 586
Frequency
0
25
50
75
100
125
400.00
500.00
600.00
700.00
800.00
900.00
1000.00
Grade 6 Math


Histogram
Mean = 719.74
Std. Dev. = 72.271
N = 586
Frequency
0
25
50
75
100
125
400.00
500.00
600.00
700.00
800.00
900.00
1000.00
Grade 7 Math

APPENDIX F

SKEWNESS AND KURTOSIS

GRADE THREE-SEVEN READING

SKEWNESS AND KURTOSIS DATA FOR TABLE 11, (P.90)

READING

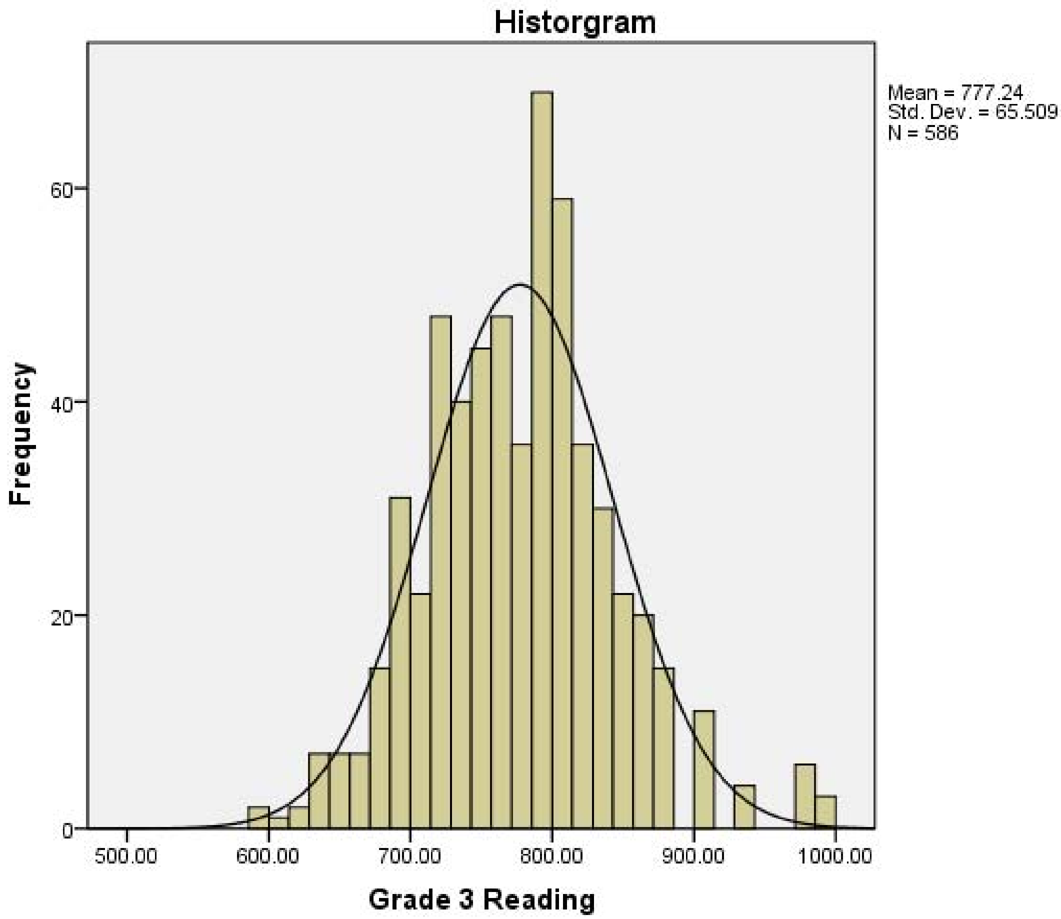
Statistics

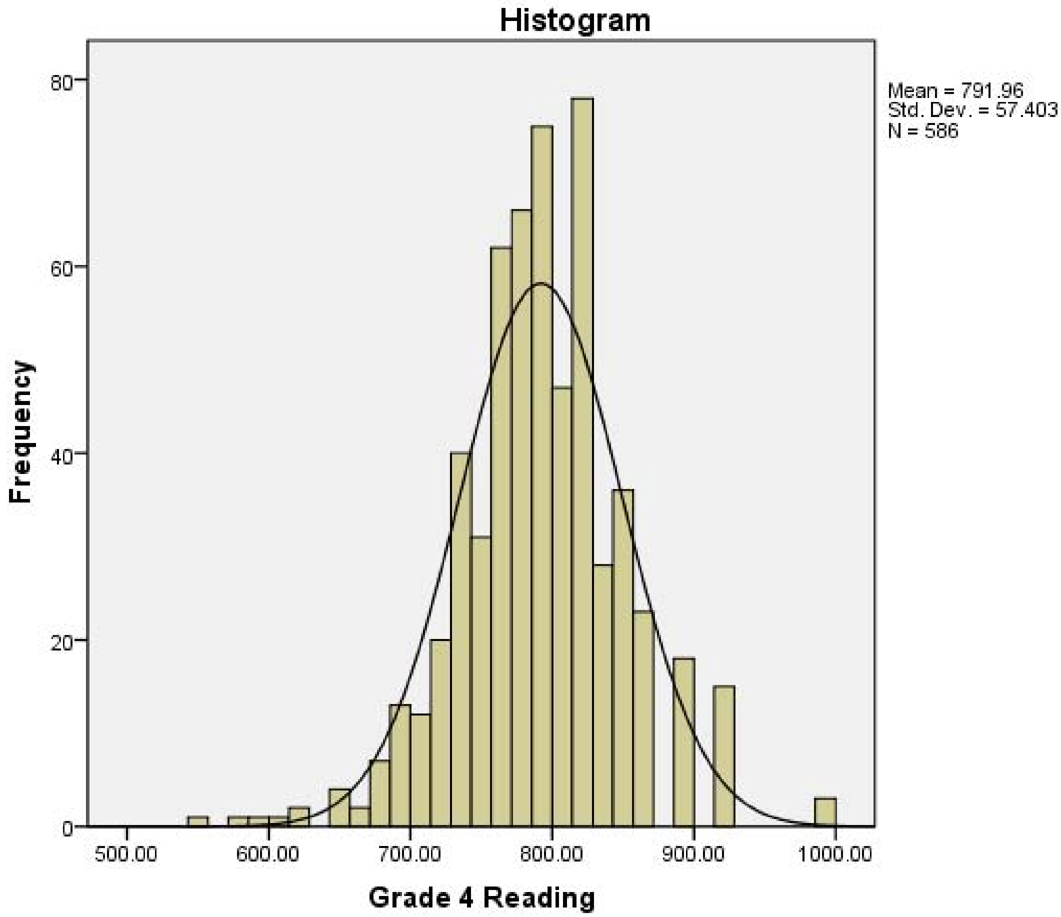
		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7
N	Valid	586	586	586	586	586
	Missing	0	0	0	0	0
Mean		777.2440	791.9590	769.9693	759.9761	740.0154
Median		778.0000	791.0000	779.0000	762.0000	737.0000
Mode		824.00	807.00[a]	817.00	793.00	775.00[a]
Std. Deviation		65.50911	57.40257	61.87002	58.71233	68.28003
Skewness		.286	-.135	-.388	-.183	.133
Std. Error of Skewness		.101	.101	.101	.101	.101
Kurtosis		.587	1.633	.999	3.112	.621
Std. Error of Kurtosis		.202	.202	.202	.202	.202

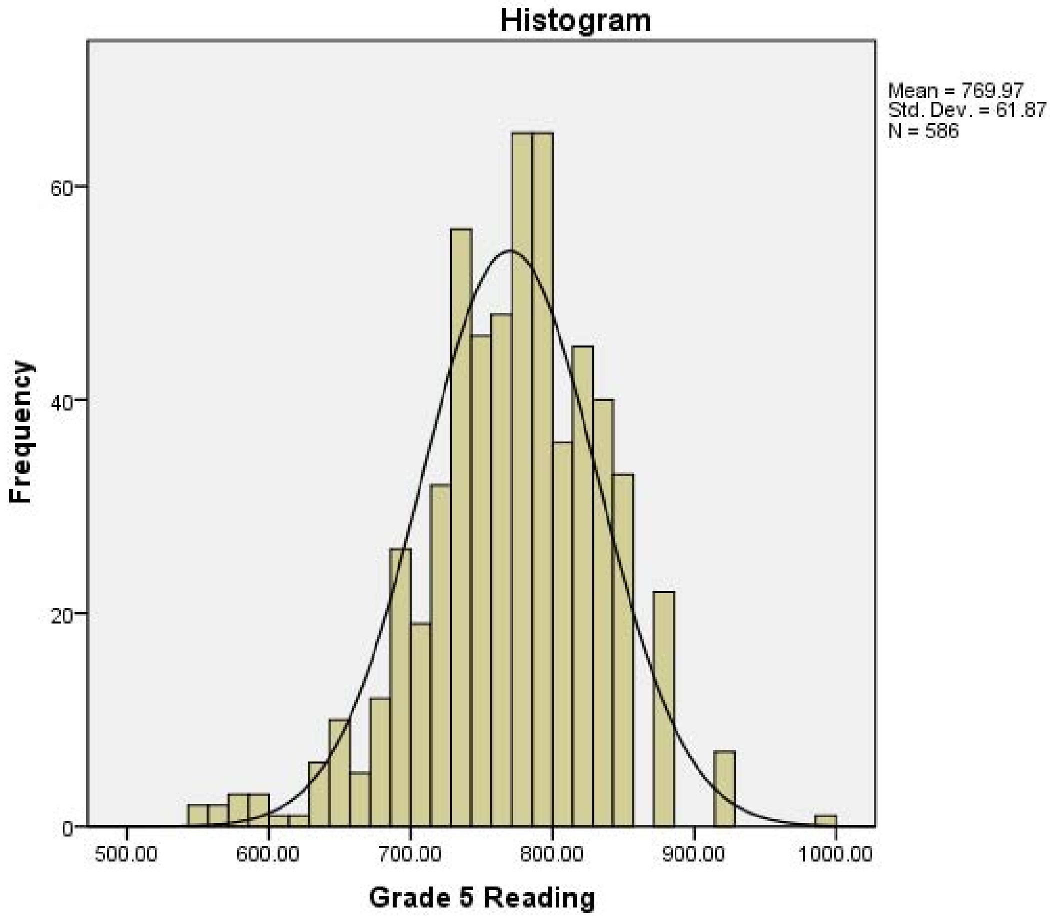
APPENDIX G

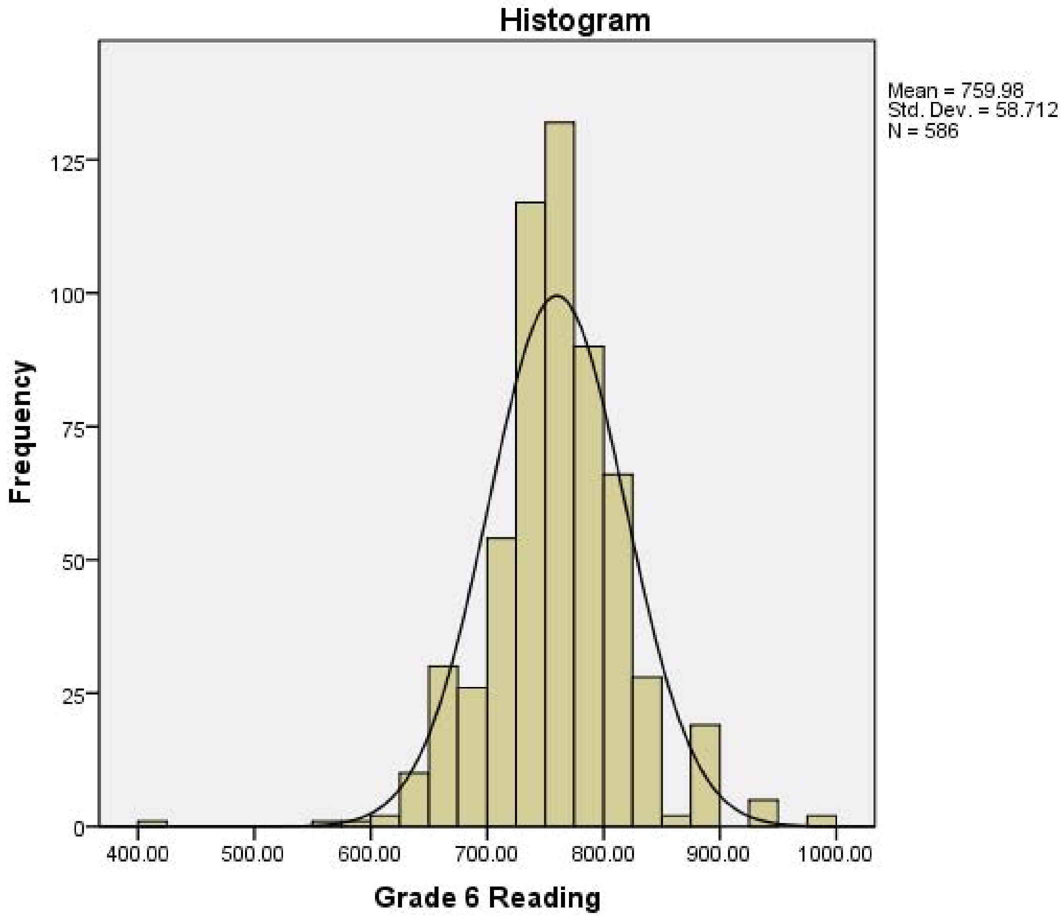
HISTOGRAMS FOR READING

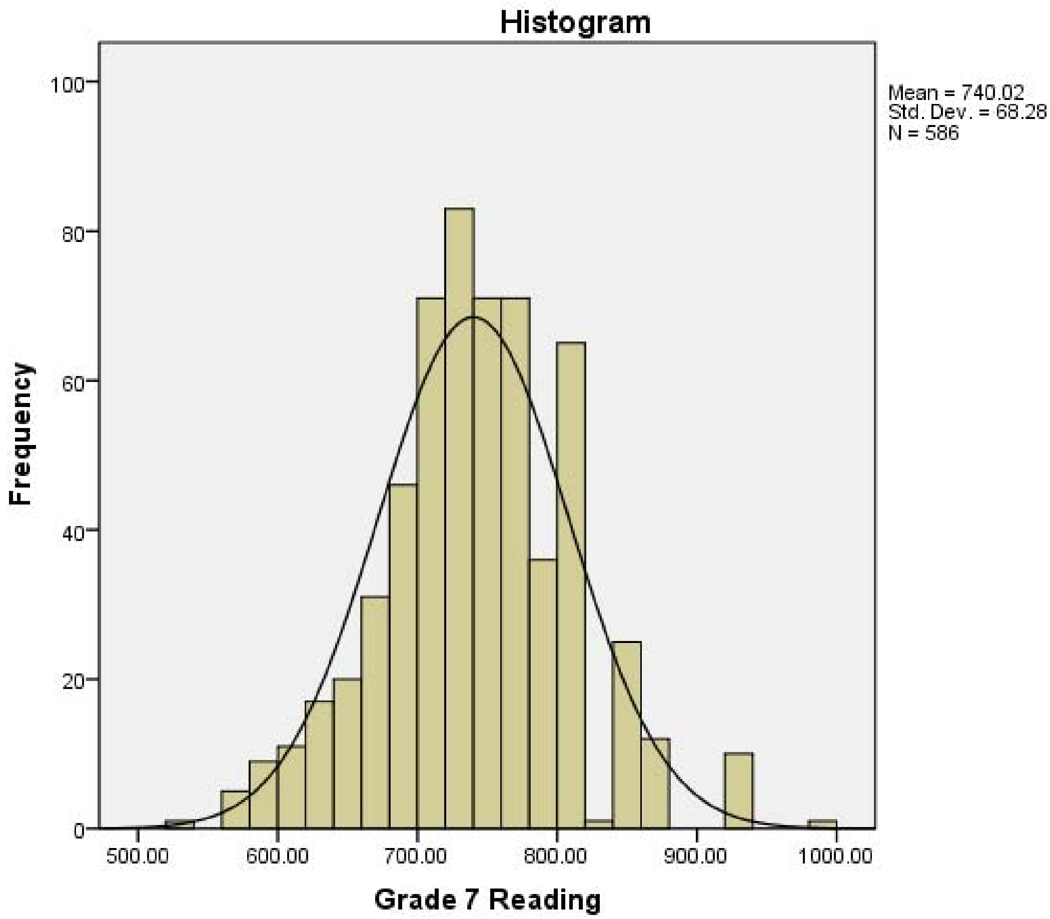
GRADES THREE-SEVEN


Historgram
Mean = 777.24
Std. Dev. = 65.509
N = 586
Frequency
60
40
20
0
500.00
600.00
700.00
800.00
900.00
1000.00
Grade 3 Reading


Histogram
Mean = 791.96
Std. Dev. = 57.403
N = 586
Frequency
0
20
40
60
80
500.00
600.00
700.00
800.00
900.00
1000.00
Grade 4 Reading


Histogram
Mean = 769.97
Std. Dev. = 61.87
N = 586
Frequency
0
20
40
60
500.00
600.00
700.00
800.00
900.00
1000.00
Grade 5 Reading


Histogram
Mean = 759.98
Std. Dev. = 58.712
N = 586
Frequency
0
25
50
75
100
125
400.00
500.00
600.00
700.00
800.00
900.00
1000.00
Grade 6 Reading


Histogram
Mean = 740.02
Std. Dev. = 68.28
N = 586
Frequency
0
20
40
60
80
100
500.00
600.00
700.00
800.00
900.00
1000.00
Grade 7 Reading

APPENDIX H

CROSS TABULATIONS

Grade 3 Math* Grade 4 Math Cross-tabulations

			Grade 4 Math				Total
			Unsatisfactory.	Limited Knowledge	Satisfactory	Advanced	
Grade 3 Math	Unsatisfactory	Count	0	1	1	0	2
		Expected Count	.0	.2	1.4	.4	2.0
		Std. Residual	-.1	1.7	-.3	-.7	
	Limited Knowledge	Count	2	37	59	1	99
		Expected Count	.3	10.3	67.4	20.9	99.0
		Std. Residual	2.9	8.3	-1.0	-4.4	
	Satisfactory	Count	0	23	307	54	384
		Expected Count	1.3	40.0	261.5	81.3	384.0
		Std. Residual	-1.1	-2.7	2.8	-3.0	
	Advanced	Count	0	0	32	69	101
		Expected Count	.3	10.5	68.8	21.4	101.0
		Std. Residual	-.6	-3.2	-4.4	10.3	
Total		Count	2	61	399	124	586
		Expected Count	2.0	61.0	399.0	124.0	586.0

grade3 Math * grade 5 Math Cross-tabulation

			Grade 5 Math				Total
			Unsatisfactory	Limited Knowledge	Satisfactory	Advanced	
Grade 3 Math	unsatisfactory	Count	1	0	1	0	2
		Expected Count	.0	.2	1.2	.6	2.0
		Std. Residual	7.5	-.4	-.2	-.8	
	Limited Knowledge	Count	3	44	49	3	99
		Expected Count	.8	9.6	59.0	29.6	99.0
		Std. Residual	2.3	11.1	-1.3	-4.9	
	Satisfactory	Count	1	13	271	99	384
		Expected Count	3.3	37.4	228.7	114.7	384.0
		Std. Residual	-1.3	-4.0	2.8	-1.5	
	Advanced	Count	0	0	28	73	101
		Expected Count	.9	9.8	60.2	30.2	101.0
		Std. Residual	-.9	-3.1	-4.1	7.8	
Total		Count	5	57	349	175	586
		Expected Count	5.0	57.0	349.0	175.0	586.0

Grade 3 Math * Grade 6 Math Cross-tabulation

			Grade 6				Total
			Unsatisfactory	Limited Knowledge	Satisfactory	Advanced	
Grade 3 Math	Unsatisfactory	Count	0	0	2	0	2
		Expected Count	.1	.2	1.1	.7	2.0
		Std. Residual	-.3	-.5	.9	-.8	
	Limited Knowledge	Count	13	33	50	3	99
		Expected Count	3.2	10.8	52.0	32.9	99.0
		Std. Residual	5.5	6.7	-.3	-5.2	
	Satisfactory	Count	6	31	236	111	384
		Expected Count	12.5	41.9	201.8	127.8	384.0
		Std. Residual	-1.8	-1.7	2.4	-1.5	
	Advanced	Count	0	0	20	81	101
		Expected Count	3.3	11.0	53.1	33.6	101.0
		Std. Residual	-1.8	-3.3	-4.5	8.2	
Total		Count	19	64	308	195	586
		Expected Count	19.0	64.0	308.0	195.0	586.0

Information for Table 14; (p. 97)

Grade3 Math * Grade 7 Math Cross-tabulation

			Grade 7				Total
			Unsatisfactory	Limited Knowledge	Satisfactory	Advanced	
Grade 3	Unsatisfactory	Count	2	0	0	0	2
		Expected Count	.4	.3	.8	.5	2.0
		Std. Residual	2.8	-.5	-.9	-.7	
	Limited Knowledge	Count	46	24	25	4	99
		Expected Count	17.4	14.5	41.9	25.2	99.0
		Std. Residual	6.9	2.5	-2.6	-4.2	
	Satisfactory	Count	55	62	193	74	384
		Expected Count	67.5	56.4	162.5	97.6	384.0
		Std. Residual	-1.5	.8	2.4	-2.4	
	Advanced	Count	0	0	30	71	101
		Expected Count	17.8	14.8	42.7	25.7	101.0
		Std. Residual	-4.2	-3.9	-1.9	8.9	
Total		Count	103	86	248	149	586
		Expected Count	103.0	86.0	248.0	149.0	586.0

Grade 3 Reading * Grade 4 Reading Cross-tabulation

			Grade 4 Reading				
			Unsatisfactory	Limited Knowledge	Satisfactory	Advanced	Total
Grade 3 Reading	Unsatisfactory	Count	0	1	4	0	5
		Expected Count	.0	.2	4.5	.3	5.0
		Std. Residual	-.2	1.8	-.2	-.6	
	Limit Know.	Count	3	15	41	0	59
		Expected Count	.4	2.4	52.6	3.6	59.0
		Std. Residual	4.1	8.1	-1.6	-1.9	
	Satisfactory	Count	1	8	461	13	483
		Expected Count	3.3	19.8	430.2	29.7	483.0
		Std. Residual	-1.3	-2.6	1.5	-3.1	
	Advanced	Count	0	0	16	23	39
		Expected Count	.3	1.6	34.7	2.4	39.0
		Std. Residual	-.5	-1.3	-3.2	13.3	
Total		Count	4	24	522	36	586
		Expected Count	4.0	24.0	522.0	36.0	586.0

Grade 3 Reading * Grade 5 Reading Cross-tabulation

			Grade 5 Reading				Total
			Unsatisfactory	Limited Knowledge	Satisfactory	Advanced	
Grade 3 Reading	Unsatisfactory	Count	1	2	2	0	5
		Expected Count	.1	.5	4.3	.1	5.0
		Std. Residual	3.0	2.1	-1.1	-.3	
	Limit. Know.	Count	6	24	29	0	59
		Expected Count	1.1	6.0	51.0	.8	59.0
		Std. Residual	4.6	7.3	-3.1	-.9	
	Satisfactory	Count	4	34	440	5	483
		Expected Count	9.1	49.5	417.9	6.6	483.0
		Std. Residual	-1.7	-2.2	1.1	-.6	
	Advanced	Count	0	0	36	3	39
		Expected Count	.7	4.0	33.7	.5	39.0
		Std. Residual	-.9	-2.0	.4	3.4	
Total		Count	11	60	507	8	586
		Expected Count	11.0	60.0	507.0	8.0	586.0

Grade 3 Reading *Grade 6 Reading Cross-tabulation

			Grade 6 Reading				Total
			Unsatisfactory	Limited Knowledge	Satisfactory	Advanced	
Grade 3 Reading	Unsatisfactory	Count	0	3	2	0	5
		Expected Count	.0	.6	4.3	.1	5.0
		Std. Residual	-.2	3.2	-1.1	-.2	
	Limited Knowledge	Count	1	27	31	0	59
		Expected Count	.4	6.7	51.1	.7	59.0
		Std. Residual	.9	7.8	-2.8	-.8	
	Satisfactory	Count	3	37	441	2	483
		Expected Count	3.3	55.2	418.7	5.8	483.0
		Std. Residual	-.2	-2.5	1.1	-1.6	
	Advanced	Count	0	0	34	5	39
		Expected Count	.3	4.5	33.8	.5	39.0
		Std. Residual	-.5	-2.1	.0	6.6	
Total		Count	4	67	508	7	586
		Expected Count	4.0	67.0	508.0	7.0	586.0

Information for Table 15, p. 101

Grade 3 Reading *Grade 7 Reading Cross-tabulation

			Grade 7 Reading				
			Unsatisfactory	Limited Knowledge	Satisfactory	Advanced	Total
Grade 3 Reading	Unsatisfactory	Count	1	1	3	0	5
		Expected Count	.2	1.0	3.7	.1	5.0
		Std. Residual	1.7	.0	-.4	-.3	
	Limited Knowledge	Count	11	33	15	0	59
		Expected Count	2.6	11.5	43.8	1.1	59.0
		Std. Residual	5.2	6.4	-4.4	-1.1	
	Satisfactory	Count	14	79	383	7	483
		Expected Count	21.4	94.0	358.5	9.1	483.0
		Std. Residual	-1.6	-1.5	1.3	-.7	
	Advanced	Count	0	1	34	4	39
		Expected Count	1.7	7.6	29.0	.7	39.0
		Std. Residual	-1.3	-2.4	.9	3.8	
Total		Count	26	114	435	11	586
		Expected Count	26.0	114.0	435.0	11.0	586.0

VITA

Ricky Dale Martin

Candidate for the Degree of

Doctor of Education

Dissertation: RELATIONSHIP BETWEEN COLLEGE READINESS, OKLAHOMA STATE TESTING PROGRAM, AND EXPLORE

Major Field: Education Administration

Biographical:

Education:

Bachelor of Science in Education: 1984; East Central University, Ada, Ok

Master of Education in Educational Administration: 1988; University of Central, Oklahoma, Edmond, Ok

Doctor of Education in Education Administration: Oklahoma State University, Stillwater, Oklahoma, July, 2010.

Experience:

1984-1988: Teacher/Coach; Chandler, Ok

1988-1996: Teacher/Coach; Geary, Ok

1996-2000: Elementary Principal/Coach; Kiefer, Ok

2000-2003: Middle School Principal; Prague, Ok

2003-2010: Superintendent; Prague, Ok

Professional Membership:

Cooperative Council of Oklahoma School Administrators

Name: Ricky Dale Martin　　　　　　　　　　Date of Degree: July, 2010

Institution: Oklahoma State University　　　　Location: Stillwater, Oklahoma

Title of Study: RELATIONSHIP BETWEEN COLLEGE READINESS, OKLAHOMA STATE TESTING PROGRAM, AND EXPLORE

Pages in the Study:169　　　　Candidate for the Degree of Doctor of Education

Major Field: Education Administration

Scope and Method of Study: The study investigated the relationship between performance on the Oklahoma State Testing Program (OSTP) for grades 3-7 and the EXPLORE in math and reading for 586 students. The EXPLORE test, a part of the ACT, is given in the eighth grade and provides college readiness benchmarks and a national percentile ranking (NPR) for each EXPLORE score. Linear regression was used to predict EXPLORE scores from Oklahoma Performance Index (OPI) scores to give a more meaningful understanding of OPI scores and performance levels in terms of college readiness and national percentile rankings. The study further analyzed the distribution of scores on the OSTP with the Kolmogorov-Smirnov test and examined kurtosis, skewness, and graphical plots of the data. Cross tabulation was used to track student performance levels from grades 3-7.

Findings and Conclusions: OPI scores were moderate to strong in their ability to predict performance on the EXPLORE assessment. The OPI scores explained from 36%-46% of the variance based on the grade and subject examined. The results indicated in grades 3-6 satisfactory performance began in the 31%-41% range while college readiness was typically met in the upper range of the satisfactory category for math and the middle range of satisfactory for reading. The Oklahoma State Department of Education increased the rigor for this study's seventh grade class resulting in satisfactory for math beginning at the 53% (NPR) and the 63% (NPR) for reading. College readiness remained in the upper range of satisfactory for math and began in the lower range of satisfactory for reading. Only the third grade reading had a normal distribution of scores; however, the distributions demonstrated the known variability in human ability. Over 50% of the students classified below satisfactory at the end of the third grade attained satisfactory following their sixth grade year. Only 28% maintained this performance level following the increased rigor implemented following the seventh grade. Over 90% of the students performing satisfactory following the third grade were still classified as satisfactory following the sixth grade and over 80% were still satisfactory considering the increased standards following seventh grade. The study provided guidelines for setting performance levels and evidence the NCLB requirement of all students attaining proficiency by 2014 is unrealistic.

ADVISER'S APPROVAL: Dr. Ken Stern