

AN ANALYSIS OF THE EFFECTS OF HIGH- AND LOW-STAKES TESTING
ON STUDENT ACHIEVEMENT

by

Frankie Eubanks Mathis

Ed.S., University of South Alabama, 1999

M.Ed., University of South Alabama, 1995

B.S., University of Mobile, 1991

A dissertation submitted to the Department of Research and Advanced Studies
College of Professional Studies
The University of West Florida
In partial fulfillment of the requirements for the degree of
Doctor of Education

2012

© 2012 Frankie Eubanks Mathis

The dissertation of Frankie Eubanks Mathis is approved:

Charles O. Aplin, Ed.D., Committee Member

Date

Matoteng M. Ncube, Ph.D., Committee Member

Date

David L. Stout, Ph.D., Committee Chair

Date

Accepted for the Department/Division:

Joyce C. Nichols, Ed.D., Chair

Date

Accepted for the University:

Richard S. Podemski, Ph.D., Dean, Graduate School

Date

ACKNOWLEDGMENTS

Despite my passion and respect for mathematics, through my doctoral journey, I have come to appreciate Albert Einstein's view that, "Not everything that counts can be counted and not everything that can be counted counts." As a result, I would like to thank those who truly count in my life. First and foremost, I would like to acknowledge my family for tremendous patience and support during this process, specifically, my parents, Franklin and Kathy Elmore, who continually sacrifice for the benefit of their children and consistently believe in my ability to achieve any academic goal I set for myself. Additionally, the inspiration of my grandparents is also unquestionable, Granny who shared her love of trigonometry with me and Gramps who lived to call me Dr. Dee. To Curtis, my husband, as well as Abbie and Katie Bug, my daughters, I am thankful that you shared this experience with me. I appreciate your willingness to compromise to make this dream a reality.

I would also like to recognize my committee, specifically committee chair, Dr. David Stout, as well as the Doctoral Support Center Staff, especially Dr. Quincy Brown and Randi Gingerich, for scholarly guidance and insight. Nonetheless, this project was also facilitated by my supportive principal, Mr. Crane, thank you. Finally, gratitude is offered to my many friends, including Tonie, Jackie, Sandy, Stephanie, Skye, Bea, Debbie, and Beverly who all offered encouragement when degree completion seemed unattainable. On a personal and a professional level, I am blessed.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	viii
ABSTRACT.....	ix
CHAPTER I. INTRODUCTION	1
A. Significance of Study.....	5
B. Research Questions	7
C. Hypotheses.....	8
D. Definition of Terms.....	8
E. Limitations	11
F. Summary	11
CHAPTER II. REVIEW OF LITERATURE	12
A. Historical Antecedents	12
B. School Accountability Program in Alabama	13
C. Impact of Student Motivation on Standardized Tests	16
D. Impact of School Accountability on Student Performance.....	21
E. Impact of School Accountability on Curriculum and Teaching Practices	23
F. Validity and Reliability of Standardized Tests	24
G. Impact of Incentives on Student Achievement on Standardized Tests	25
H. Consequences for Failing to Achieve	26
I. Summary	27
CHAPTER III. METHODOLOGY	29
A. Research Questions	30
B. Hypotheses.....	31
C. Research Design.....	32
D. Participants.....	32
E. Protection of Human Subjects	33
F. Data Collection	34
G. Data Analysis	36
H. Limitations of the Study.....	36
I. Summary	37
CHAPTER IV. RESULTS	38
A. Data Analysis for Research Question 1	40
B. Data Analysis for Research Question 2	42
C. Data Analysis for Research Question 3	43

D. Data Analysis for Research Question 4	44
E. Data Analysis for Research Question 5	45
F. Summary	48
CHAPTER V. CONCLUSIONS.....	49
A. Conclusions and Interpretations for Research Question 1	49
B. Conclusions and Interpretations for Research Question 2.....	50
C. Conclusions and Interpretations for Research Question 3.....	51
D. Conclusions and Interpretations for Research Question 4.....	51
E. Conclusions and Interpretations for Research Question 5.....	52
F. Implications for Student Testing Programs	53
G. Recommendations for Future Research	54
H. Summary	54
REFERENCES	56
APPENDIXES	60
A. Request for Institution Permission to Conduct Research: An Analysis of the Effects of High- and Low-Stakes Testing on Student Achievement	61
B. Institutional Permission to Conduct Research: An Analysis of the Effects of High- and Low-Stakes Testing on Student Achievement	63
C. Letter from the Institutional Review Board for Human Subjects Granting Permission to Conduct Research: An Analysis of the Effects of High- and Low-Stakes Testing on Student Achievement	65

LIST OF TABLES

1.	Descriptive Statistics for the Student Sample	40
2.	Disaggregated Data Analysis for 8th Grade Mid-Year Course Averages	46
3.	Disaggregated Data Analysis for 8th Grade End-of-Course Averages	46
4.	Disaggregated Data Analysis for ARMT Scores	47
5.	Disaggregated Data Analysis for Algebra I End-of-Course Averages	47

LIST OF FIGURES

1.	Scatter plot of eighth grade mathematics mid-year course averages and eighth grade mathematics ARMT scale scores.....	41
2.	Scatter plot of eighth grade mathematics end-of-course averages and eighth grade mathematics ARMT scale scores.....	42
3.	Scatter plot of AHSGE mathematics test scores and Algebra I end-of-course averages.....	44
4.	Scatter plot of ARMT mathematics test scores and AHSGE mathematics test scores.....	45

ABSTRACT

AN ANALYSIS OF THE EFFECTS OF HIGH- AND LOW-STAKES TESTING ON STUDENT ACHIEVEMENT

Frankie Eubanks Mathis

With increased emphasis on accountability, the use of low-stakes test data to make high-stakes decisions about program effectiveness is on the rise. In order to make valid inferences about what students know and can do, it is crucial to understand the consequences of low and high stakes in testing contexts. As a result, with a sample comprised of 49 eleventh grade students enrolled in Saraland High School, this study indicates that relationships between student performance on state-mandated testing programs in Alabama and grades earned in corresponding mathematics courses exist. Through regression analyses, it was determined that grades earned in mathematics classes are predictors of scores earned on the state-mandated tests. Additionally, in this study, the sample data revealed that student achievement did not increase as the individual consequences associated with the test increased. Rather, student achievement on the state-mandated tests was consistent with the overall mathematics achievement levels demonstrated in the classroom.

CHAPTER I

INTRODUCTION

Robert Schaeffer, the Public Education Director of FairTest, the National Center for Fair and Open Testing, remarks, “Believing we can improve schooling with more tests is like believing you can make yourself grow taller by measuring your height” (Assessment Network, n.d., para. 3). The need for testing as a form of accountability in public schools is rarely disputed; rather, the controversy surrounds the frequency, number, and type of assessments administered to students. How did standardized test scores emerge as a dominating measure of school effectiveness in the United States?

In 1965, motivation for educational reform nobly stemmed from President Lyndon Johnson’s war against poverty; as a result, the Elementary and Secondary Education Act was authorized by Congress. The Act allocated funding for poor schools serving children who lived in low income communities. In addition, the Act prohibited the establishment of a national curriculum. Since 1965, the Act was revamped and reauthorized numerous times. These revisions resulted in a number of mandates designed to improve education for all, with the latest legislation resulting in the No Child Left Behind Act (NCLB) of 2001 which was last re-authorized in 2002.

Currently, in compliance with the NCLB, reading and mathematics test outcomes are used to evaluate the success of public schools in the United States. In accordance with this legislation, many states have implemented policies that result in students’ graduation being determined by their performance on an examination, placing high stakes on the examination for the students. In other words, a student’s academic career may culminate with the score from one high-stakes assessment, resulting in the student’s either earning or not earning a high school

diploma based on a single test score. As a character on the television show *Boston Public* exclaims, “Education ceases to be learning when the 3 R's are read, remember, and regurgitate” (Students Against Testing, n.d., para. 33). However, with this emphasis on testing in the United States, an unintentional back-to-basics movement is evolving, one which is based on an educational climate dominated by student assessments.

If the aim of educational initiatives is to improve the quality of education for all, why is there controversy surrounding school legislative reform? One problem that may not be obvious is that educational initiatives are often sponsored by members of Congress who may or may not have professional experience in the field of education. With compulsory school attendance laws in all states in the union, all children are required to enroll in and attend school. The fact that all citizens attended school at one time empowers many to assume they have expertise and experience valuable to the development of public school mandates, regardless of their professional background in the field of education. Although collaboration of stakeholders is essential to the development of effective initiatives, it is also important to recognize the necessity of developing relevant plans which are beneficial to all children. Specifically, the need is for mandates which are properly funded and are designed to improve the educational system rather than merely increasing bureaucracy.

In addition to being driven historically by government-mandated initiatives such as the Elementary and Secondary Education Act which eventually gave birth to NCLB, the United States' educational system often pedagogically swings on a pendulum between back-to-basics movements and progressive methods of instruction. Regardless of whether one is discussing educational mandates or pedagogy, there is educational vernacular built on acronyms, clichés, and campaigns that dominate the discussion. In the 21st century, the campaign dictating

educational initiatives is based on the term *school accountability*. Under NCLB, school accountability drives the curriculum, and the foundation of accountability systems is standardized testing. Dr. Paul Houston asserts,

Only on ‘Who Wants to be a Millionaire?’ can people rise to the top by rote memorization and answers to multiple-choice questions. The FINAL ANSWER to improving education is more than memorizing facts for a multiple-choice test. Children today need critical thinking skills, creativity, perseverance, and integrity - qualities not measured on a standardized test. (Students Against Testing, n.d., para. 22)

As a result of the emphasis on accountability and testing, a behaviorist model of stimulus-response dominates schools today. Additionally, with preparation for standardized testing emerging as the primary focus of instruction, teachers express concern that standardized testing restricts opportunities for students to participate in hands-on, discovery learning activities.

In the state of Alabama, the accountability system used to determine if schools are achieving Adequate Yearly Progress as required by NCLB is primarily based on student performance on the Alabama Reading and Mathematics Test (ARMT) and the Alabama High School Graduation Exam (AHSGE). In Alabama, the ARMT is administered in public elementary and middle schools while the AHSGE is administered in public high schools. Although there are additional criteria in the state accountability system, the ARMT and the AHSGE are the pillars of the system, providing results on academic achievement of students with the focus on the areas of mathematics and reading (Alabama Department of Education, 2011).

In 1995, the Alabama State Legislature passed the Education Accountability Law. Based on this law, high school students are required to complete 24 credits of study, including four

years of mathematics, science, social studies, and English respectively. Until changes were made in 2009, prospective graduates were required to pass all five subject-area tests on the AHSGE which included reading, language, mathematics, science, and social studies. However, in 2009, Alabama's First Choice Program was implemented and included a credit-based diploma endorsement. Under the credit-based endorsement, students who had completed all of their coursework and passed at least three sections of the AHSGE — reading and math as well as language, science, or social science — were allowed to graduate (Alabama Department of Education, 2009).

Since the results of the ARMT and AHSGE generate publicity and press, these test results hold high stakes for schools. In contrast, they produce varying stakes for students. For example, the AHSGE is administered throughout a student's high school career, typically beginning with a pre-test which is administered in the 10th grade. Despite multiple opportunities to achieve proficiency standards on the AHSGE, a student who is unable to pass the mathematics or reading portion of the AHSGE will not receive a high school diploma (Alabama Department of Education, 2009).

On the other hand, there are no individual consequences for elementary or middle school students who fail to meet ARMT proficiency standards. Alabama is not alone in this accountability practice. For example, in Virginia, the Standards of Learning tests

are high-stakes [*sic*] for the schools themselves at the elementary and middle school level, but less so for the individual students. SOL [Standards of Learning] testing results have huge implications for the school, and the results are part of the School Performance Report Cards. At this level, the pass rate is the determining factor of whether or not a school attains or retains its accreditation status. It forms the basis for assessment of

school, teacher, and student performance, and the results are widely publicized in the media. Standardized test scores are also a factor in determining local real estate values as homebuyers seek to purchase homes in neighborhoods with high achieving students.

(U.S. Department of State, n.d., para. 4)

Similar to the stakes resulting from the Standards of Learning tests in Virginia, the ARMT holds high stakes for teachers and administrators. Conversely, with a high school diploma at stake, the AHSGE is a test that generates high stakes for students, teachers, and administrators.

Significance of Study

With the reputation of educators and funding being based on standardized test results, one must question student achievement on a test that has no direct impact on promotion or retention and thus offers little or no extrinsic incentive for students to excel on the test. Do low-stakes test scores truly reflect the academic knowledge of student test takers? Is it a fair practice to base the evaluation of a school's academic program on low-stakes test results?

The results of the study indicate the relationships between students' grades earned in relevant high-stakes mathematics coursework with respect to students' mathematics achievement on the eighth grade ARMT, a low-stakes test for students, and the students' performance on their first administration of the mathematics portion of the AHSGE, a high-stakes test for students. Since test results often provide the basis for school accountability measures, awareness of student achievement levels on low- and high-stakes tests with respect to student ability demonstrated in the classroom may provide beneficial information to policymakers at the federal, state, and local levels. Additionally, the results of this study indicate the cultural implications of student testing in the United States. Overall, the purpose of the study was to

determine if student achievement was impacted when the stakes of the test were raised for the student.

On a local level, course grades, ARMT scores, and AHSGE results are used to determine if students need to be scheduled into remediation classes or recommended for tutoring programs. If relationships exist between course grades earned and state tests, then using the statistical inferences, administration could determine prior to testing whether or not a student is prepared for the test. Therefore, the results of this study are useful when considering placement of students in academic intervention programs.

This study also shows the consequences for schools that result from basing student achievement conclusions using low-stakes test results. The impact of the study reaches beyond the area of mathematics education to other tested subject areas. In fact, public school system administrators and teachers throughout the United States who are facing the dilemma of achieving accountability goals based on low-stakes standardized test scores may find value in this research. Nonetheless, the information yielded from this study may be useful when policy makers at federal and state levels consider the justice or injustice of basing school accountability practices on low-stakes test results.

As 2014 approaches, NCLB (2002) requires 100% of students to score proficient in math and reading on state-mandated accountability tests. Concerning this requirement, an elementary teacher who was interviewed by Clarke et al. (2003) states,

To think that every child is going to be able to perform at the same level at the same time could only be dreamed up by someone who has no idea what children are, because it's totally unrealistic. That's not human...Not all adults are the same. Why should all ten-year-olds be the same? (p. 85)

Ultimately, the goal of 100% proficiency, especially on low-stakes tests, is proving to be a daunting if not impossible task in many schools across the country.

Research Questions

The following research questions form the basis that guided the researcher throughout the study:

1. Is there a relationship in students' high-stakes eighth grade mathematics mid-year course averages and their corresponding low-stakes scores on the mathematics portion of the ARMT?
2. Is there a relationship in students' high-stakes eighth grade mathematics end-of-course averages and their corresponding scores on the mathematics portion of the low-stakes ARMT?
3. Considering that end-of-course averages result in stakes equal to the graduation exam and that the AHSGE mathematics test is based on mathematics content taught in the eighth grade mathematics courses as well as Algebra I, another question of interest arises. Are students' high school Algebra I end-of-course mathematics grades related to students' scores on the mathematics portion of the AHSGE?
4. Is there a relationship in student achievement levels on the mathematics portions of the low-stakes ARMT and the high-stakes AHSGE?
5. In addition to the stakes associated with each test, do other demographic factors including gender, socioeconomic status, and ethnic group contribute to differences in the student test results?

In summary, the questions of interest ultimately concern the impact of the low and high stakes associated with each test in regard to student achievement. Therefore, the focus of the research questions pertains to inferences that can be made concerning student achievement

levels, particularly when the stakes are raised for the students. In other words, does student achievement increase as the individual consequences associated with each test increase?

Hypotheses

Prior to statistical analysis to determine the quantitative support for the research questions posed, the following hypotheses were conjectured:

1. A statistically significant relationship between students' high-stakes eighth grade mid-year mathematics course averages and their low-stakes eighth grade ARMT mathematics test scores does not exist.
2. A statistically significant relationship between students' high-stakes eighth grade end-of-course mathematics averages and their low-stakes eighth grade ARMT mathematics test scores does not exist.
3. A statistically significant relationship between students' high-stakes high school Algebra I end-of-course scores and their high-stakes AHSGE mathematics test results exists.
4. A statistically significant relationship between students' mathematics achievement levels on the high-stakes AHSGE and their low-stakes ARMT does not exist.
5. When sample data is disaggregated by demographic factors including gender, socioeconomic status, and ethnic group, statistically significant differences in the group means do not exist.

Definition of Terms

The following key terms and definitions used throughout this study are operational definitions within the context of this study; these may be subject to alternative interpretations.

Alabama High School Graduation Exam (AHSGE). For the purpose of this study, the AHSGE is defined as a high-stakes exam that results in serious consequences for students who fail to meet proficiency standards. Specifically, the AHSGE is a multiple choice test administered over five consecutive school days with one subject area tested per day. Students who have completed required course work may begin taking the exam in ninth grade; every student may attempt to pass any subject area subtest up to five times (Alabama Department of Education, 2003). Currently, a plan to replace the AHSGE with end-of-course tests has been proposed; however, with budget cuts, the plan to develop and implement end-of-course tests has been postponed.

Alabama Reading and Mathematics Test (ARMT). Because students' grades are not impacted and student promotion is not based on whether or not a student meets proficiency standards, the ARMT is defined as a low-stakes exam for the purpose of this study. The ARMT is a criterion-referenced test that consists of items from the Stanford Achievement Test (Stanford 10) which match the reading and mathematics Alabama state content standards. In addition, supplementary test items were developed and included so that all state content standards will be encompassed on the test. Together, the Stanford 10 items and the supplementary items are known as the ARMT. The ARMT is administered during a testing window determined by the state during the spring of each year (Madison County Schools, n.d.).

Annual Measureable Objective. The annual requirements for the percentage of students required to meet proficiency is determined by each state. In Alabama, the ultimate goal as required by NCLB is that all students reach Level III (proficient) or higher by 2013-2014 (Alabama Department of Education, 2010).

Adequate Yearly Progress. State accountability in Alabama is based on NCLB requirements. The term *Adequate Yearly Progress* is used to specify whether a school or system has met all of the annual accountability goals as determined by the percentage of students scoring at the level of proficient or higher on the ARMT or AHSGE, attendance rate, drop-out rate, and the percentage of students who participate in mandated assessments (Alabama Department of Education, 2010).

End-of-course average. End-of-course averages were calculated by averaging the student's four quarter grades. However, for students who completed Algebra I as a two-year course, the yearly averages for the two courses were combined to produce one numeric value to represent the overall Algebra I end of course average.

High-stakes test. A high-stakes test is defined as a test in which there are serious consequences for failing to meet proficiency standards. The consequences may apply to the school system, the school, or the student. For the purpose of this study, the AHSGE is a high-stakes exam that results in serious consequences for students, schools, and school systems failing to meet proficiency standards.

Low-stakes test. A low-stakes test is defined as a test in which there is little or no consequence associated with failing to achieve proficiency standards. Although the ARMT and the AHSGE produce the same high stakes for schools, the ARMT has little to no consequence for students. Therefore, the ARMT is defined as a low-stakes exam for the purpose of this study.

Mid-year course average. Mid-year course averages were calculated by averaging the student's first and second quarter grades.

No Child Left Behind Act (NCLB). Under NCLB, all students must score at the proficient level on state tests by the 2013-2014 school year. States and schools are held accountable for the

mandated requirements. All districts and schools receiving Title I funds must meet state Adequate Yearly Progress goals for their total student populations and for demographic subgroups including ethnic/racial groups, economically disadvantaged students, limited English proficient students, and students with disabilities. If a school fails to meet Adequate Yearly Progress goals for two or more consecutive years, it is classified as in need of improvement, and school leaders may face consequences (GreatSchools, n.d.).

Limitations

Given the sample used in the study was comprised of students from a suburban city school system in Southeast Alabama, the ability to generalize the findings to a larger more diverse population is limited. The overall demographic homogeneity of the sample furthermore restricts extrapolations from this study. Aside from limitations resulting from the selection of the sample, the fact that the ARMT was administered during the students' eighth grade year while the AHSGE was not administered until students' ninth or 10th grade year presents a maturation concern. Although both tests focus on similar content objectives, the ARMT and AHSGE raw scores are not based on the same ratio scale, thus complicating the data analysis process. Another potential threat to validity is teacher effect. Teacher effect was not controlled for in terms of impact on student course grades; however, the anticipated impact in this study is minimal since students are randomly assigned to one of two possible teachers.

Summary

In this chapter, the purpose and significance of the study is outlined. Additionally, research questions and hypotheses are stated. Terms relevant to the study are defined, and limitations to the study are identified.

CHAPTER II

REVIEW OF LITERATURE

The literature review for this study begins with an examination of the cultural emergence of testing and the nature of accountability in public schools in the United States. Specifically, school accountability practices in the state of Alabama are examined. The impact of these practices with respect to student motivation and performance as well as the influence on curriculum and teaching practices is also discussed.

Historical Antecedents

Throughout the 20th century and to the current day, the educational system in the United States has been consistently driven by themes varying from child-centered to society-centered. With the launching of Sputnik in 1957, there was intense criticism of education in the United States and a commitment by the federal government to promote a quality education. In the late 1970s and early 1980s, the perception that schools needed to focus on a core curriculum, the skills that citizens needed to be productive, was prevalent. This notion gave birth to the back-to-basics education movement. An outgrowth of the back-to-basics movement was outcome-based educational reform whose foundation was built on state accountability standards mandated in the NCLB Act of 2001. The key policy components of this public education reform are standards, tests, and accountability. Clarke et al. (2003) assert, “The standards outline the expectations held for all students, the test provides a way to judge student performance against these standards, and the accountability component provides an incentive – in the form of stakes attached to the test results” (p. 5). With respect to this standards-based movement, Lattimore (2001) alleges, “testing has become an attractive option for policymakers both because it has the potential to affect the behavior of educators in the educational systems and because it is often viewed by the

public as a way to guarantee a basic level of quality education” (p. 57). As a result, testing programs emerge as the impetus by which student achievement is measured.

Since 2001, with standardized testing programs serving as the foundation of the accountability measures, the standards-based NCLB dominates educational initiatives by emphasizing accountability for public schools across the United States. Although NCLB requires that schools achieve Adequate Yearly Progress and attain the ultimate goal of all children performing on the proficient level by school year 2013-2014, the yearly testing program and accountability guidelines were defined by each state and then submitted to the United States Department of Education for approval (NCLB, 2002).

Currently, “assessment results are used for ‘high stakes’ purposes such as grade promotion, certification, and high school graduation as well as holding schools accountable to improve instruction and student learning” (Ryan, Ryan, Arbuthnot, & Samuels, 2007, p. 5). Consequently, the emphasis and controversy surrounding standardized testing has been mounting for years. In fact, Goodling (1997), Chair of the House Education Committee, declares, “If more testing were the answer to the problems in our schools, testing would have solved them a long time ago” (p. A21). Goodling’s statement summarizes the frustration that many share with regard to the testing debate.

School Accountability Program in Alabama

In the state of Alabama, the accountability program is composed of the following criterion-referenced tests: the ARMT, the AHSGE, and the Alabama Science Assessment. For the purpose of this study, the Alabama Science Assessment is not discussed. To assess students’ reading and mathematics ability, each spring, all children in grades three to eight except those with an IQ below 55 are administered the ARMT. For students with learning disabilities,

accommodations for testing such as small group settings or read-alouds are utilized if specified in the student's individual educational plan (Alabama Department of Education, 2011).

Because the success of the school's academic program is determined by the students' ability to meet proficiency criteria, the ARMT assessment produces high stakes for schools. For the purpose of this study, since there are no individual consequences for students who score poorly, the ARMT is defined as a low-stakes test for students. On the ARMT, student proficiency is reported on the following four-point scale: "Level I – Does Not Meet Academic Content Standards, Level II – Partially Meets Academic Content Standards, Level III – Meets Academic Content Standards, and Level IV – Exceeds Academic Content Standards" (Alabama Department of Education, 2011, p. 9). A similar scale is used to measure student proficiency on the AHSGE. However, students who pass the AHSGE only receive a notification that they passed the test. Unlike with the ARMT, high school students are not notified as to whether they scored at Level III or Level IV proficiency on the AHSGE. In contrast, students who do not pass at least one subtest of the AHSGE receive a scaled score and a detailed mastery report by objective for each failed portion of the exam. Although these students are not formally notified of their current level of proficiency, an interval scale which classifies the student's proficiency as Level I or II is available (Alabama Department of Education, 2010).

In 2004, when Alabama's accountability plan was implemented, the state established yearly goals or annual measureable objectives for student proficiency on state-mandated tests. Students' performance on these tests determines whether or not the school meets the annual measureable objectives, which in turn factors into whether or not the school achieves Adequate Yearly Progress. Judson (2007) asserts, "If schools continually fail to meet the requirements of their state's accountability guidelines, a school can be subject to abrupt intervention from the

state's department of education" (pp. 15-16). For schools that consistently fail to meet Adequate Yearly Progress or the other measures which are used to evaluate the academic growth of students, the consequences may be simple like "requiring a school to submit an improvement plan or might be quite drastic, such as replacement of school administrators or removing the responsibility of the school's governance from the school district school board" (Judson, 2007, p. 16). In the instances in which reconstitution is ordered, all faculty, from the administration to the custodial staff, can be replaced, making the stakes for the school high. In Alabama, despite the fact that there are no individual consequences for elementary and middle school students who fail to achieve proficiency on the ARMT, there are consequences for schools failing to meet Adequate Yearly Progress goals. For high school students, a diploma is at stake for seniors who do not attain proficiency on the reading, mathematics, and one additional subtest on the AHSGE prior to their graduation date.

Overall, the question that this researcher desires to answer concerns student achievement on tests that hold high stakes for schools but not for students. Specifically, do students perform to their maximum ability if there are no personal consequences? If not, do low-stakes tests truly reflect the success of a school's academic program? Are accountability programs that have been developed under NCLB guidelines effective in measuring the success of a school? Do accountability programs not only measure but also promote student achievement? Hanushek and Raymond (2005) discovered that the introduction of accountability systems into a state's educational program tends to lead to larger

achievement growth than would have occurred without accountability. The analysis, however, indicates that just reporting the results has minimal impact on student

performance and that the force of accountability comes from attaching consequences such as monetary awards or takeover threats to school performance. (p. 298)

Therefore, there is evidence to support the benefits of accountability initiatives.

As previously mentioned, the AHSGE holds high stakes for both schools and students. In Alabama, a high school's Adequate Yearly Progress status is based on the ability of students in the junior class to meet assessment criteria in the areas of reading and mathematics as well as language arts, science, and history. At the same time, students also face high consequences because students who fail to achieve proficiency on three of the five sections of the AHSGE, two of which must be reading and mathematics, before graduation will not receive a high school diploma (Alabama Department of Education, 2011). Papay, Murnane, and Willett (2010) raise concern with the establishment of a cut-off score for graduation exam testing because with the establishment of a cut-off score, "states divide a continuous performance measure into dichotomous categories. Thus, students with scores near the cut-off either pass or fail despite having essentially equal skills" (p. 5). On the other side of the debate, Ryan et al. (2007) assert, "The 'high-stakes' nature of these tests is intended to motivate students to perform to high standards, teachers to teach better, and parents and local communities to make efforts to improve the quality of local schools" (p. 5). As a result, both sides of the debate present valid points to consider when examining the value of high-stakes testing programs.

Impact of Student Motivation on Standardized Tests

Why are some driven to perform, regardless of personal consequences, while others are impossible to motivate despite what may be at stake? How does student motivation impact achievement on standardized tests? Barry, Horst, Finney, Brown, and Kopp (2009) proclaim, "Understanding students' test-taking motivation is extremely important, especially in low-stakes

contexts where it is likely to vary, because it affects whether the test scores truly reflect the knowledge, skills, and abilities of the test takers” (p. 5). If the stakes associated with a test impact student motivation, then are low-stakes test scores an accurate measure of student achievement? To answer this question, one must define test taking motivation: “the extent to which an examinee gives his or her ‘best effort’ to the test, with the goal being to accurately represent what one knows and can do in the content area covered by the test” (Wise & DeMars, 2005, p. 2). With the objective being to encourage students to perform to their optimal abilities when testing, it is beneficial to identify factors that motivate students to strive to do their best. Barry et al. state,

Test-taking motivation is likely to be high when examinees complete tests for which there are direct personal consequences associated with their scores (e.g., achievement tests, admissions tests, placement tests). These situations are termed high stakes.

Alternatively, test-taking motivation is likely to be more variable when examinees complete tests for which there are little to no personal consequences associated with their scores. Situations in which this is the case are termed low-stakes. (p. 4)

With scores on standardized tests being used to evaluate schools across the country, high- and low-stakes accountability programs are being examined. For example, Clarke et al. (2003) conducted interviews in three states and report,

State test results were one of several pieces of information used to determine school accreditation in Kansas, but had no official stakes for students. In Michigan, school accreditation was determined by student participation in, and performance on, the state test, and students could receive an endorsed diploma and were eligible for college tuition credit if they scored above a certain level on the eleventh-grade test. In Massachusetts,

school ratings were based on the percentage of students in different performance categories on the state test, and students – starting with the class of 2003 – had to pass the tenth-grade test in order to graduate from high school. Thus, as one moves from Kansas to Michigan to Massachusetts, the stakes for educators remains fairly constant, but the stakes for students increase dramatically. (p. 73)

While the stakes for students vary from state to state, do the stakes associated with a test impact student achievement? In other words, is there a statistically significant difference in student achievement on standardized tests resulting from personal consequences such as retention for students?

Research indicates that student motivation is higher when stakes are raised for the students. In fact, a study designed to investigate the effects of material incentives on intelligence test performance found that “incentives increased IQ scores by an average of 0.64 SD, with larger effects for individuals with lower baseline IQ scores” (Duckworth, Quinn, Lynam, Loeber, & Stouthamer-Loeber, 2011, p. 7716). Duckworth et al. further claim, “findings suggest that, under low-stakes research conditions, some individuals try harder than others, and, in this context, test motivation can act as a third-variable confound that inflates estimates of the predictive validity of intelligence for life outcomes” (p. 7716). Additionally, previously conducted studies indicate that self-reported motivation was higher when there were individual incentives for the participant, i.e. self-reported motivation was higher for job applicants than for incumbents.

In another study, led by O’Neil, Abedi, Lee, Miyoshi, and Mastergeorge (2004), eighth grade students who were paid for correctly answering questions reported higher motivation than those eighth grade students not paid for correct answers although a subsequent study using a

sample of 12th grade students conducted by the same researcher did not indicate an impact of monetary incentives on achievement or motivation. However, Duckworth et al. (2011) further hypothesize that “because ratings of motivation are typically self-reported post hoc, it is possible that they reflect how well test takers think they performed as opposed to how hard they tried” (p. 7716). Duckworth et al. conclude,

test motivation is higher and less variable among participants who are above-average in measured IQ. These findings imply that earning a high IQ score requires high intelligence in addition to high motivation. Lower IQ scores, however, might result from either lower intelligence or lack of motivation. (p. 7718)

Duckworth et al. concede, “test motivation may be a serious confound in studies including participants who are below-average in IQ and who lack external incentives to perform at their maximal potential” (p. 7718). However, in the state of Alabama and numerous states across the nation, despite the IQ of students being tested, the success of a school is based on results which have no consequences for students. To diminish the likelihood that test motivation will act as a third-variable confound, experimenters should provide substantial performance-contingent incentives or design the outcomes in a manner in which they directly affect test takers, such as intelligence tests used for employment or admissions decisions (Duckworth et al., 2011).

When examining the low levels of proficiency demonstrated by students in the United States on national and international assessments, one reason often cited for poor performance is based on the premise that there are no consequences or stakes attached to performance on the tests, and as a result, students are not motivated to exert their best effort. In other words, assessment consequences, whether high- or low-stakes, are not considered to influence how students perform in a testing situation (Heubert & Hauer, 1999). Ryan et al. (2007) challenge

that notion and claim, “not only knowledge but individuals’ personal beliefs and goals influence performance,” contradicting the assumption that test taking is a similar experience for all students (p. 5).

As previously mentioned, studies have been conducted which focus on the effects of monetary incentives for student performance. O’Neil, Sugrue, and Baker (1995) assert, at least for Grade 8 participants, student effort can be increased by financial rewards offered at the time of test taking, and that such effort can result in an increase in NAEP [National Assessment of Educational Progress] math test scores. Thus, from a policy perspective, scores from low-stakes tests may not represent what the student knows. (p. 135)

If students do not put forth their best effort, the scores on these tests may underestimate actual levels of proficiency, possibly resulting in dire implications for educational institutions (Wise & DeMars, 2005). Accordingly, scores obtained on low-stakes tests may not express what students know but rather “what students will demonstrate with minimal effort” (O’Neil et al., 1995, p. 135).

Barry et al. (2009) express concern that “one may assume that students fail to provide effort in low-stakes contexts and use this assumption as an excuse to ignore low test scores that indicate program ineffectiveness or to advocate ending low-stakes testing programs altogether” (pp. 5-6). At some point, the question begins to take the form of the chicken or the egg scenario because one may assume that low motivation results in low test scores while overlooking the possibility that the low test scores result from low achievement and vice versa.

To further complicate the issue of student motivation during testing, Barry et al. (2009) determine that “in a low-stakes setting some students may have higher motivation for tests that

are not especially difficult but may have much lower motivation for tests that are especially cognitive demanding” (p. 11). Accordingly, the complexity of student motivation may vary depending on the cognitive demand of the test. However, NCLB legislation is based on a singular-dimensional view of test motivation based on the assumption that high-stakes tests motivate students to the same degree (Clarke et al., 2003). Additionally, according to this argument, motivation does not vary based on demographic, socioeconomic, content area, or other social-psychological processes (Ryan et al., 2007). This assumption relies on the existence of a model test taker, not an actual test taker. A model test taker is one who is aware of personal knowledge and whose goal is to maximize test performance (Budescu & Bar-Hillel, 1993). In reality, test takers are students who may be unsure about what they know and whether they would like to do well or not. Furthermore, an actual test taker’s goal may not be to maximize performance; instead, the student may simply want to avoid looking bad or doing poorly (Budescu & Bar-Hillel, 1993).

Impact of School Accountability on Student Performance

As previously mentioned, school accountability practices can have a positive impact on student achievement. However, a number of studies report a negative impact resulting from accountability practices. For example, students and teachers often express concern with the number of state-mandated tests. Clarke et al. (2003) write,

Overtesting was a dominant theme in the Kansas interviews and was viewed as reducing students’ motivation to do well on state tests. This apathy was perceived as being due, in part also, to the lack of consequences for students from the test results. (p. 76)

Aside from overtesting, there are additional concerns that prohibit students and teachers from supporting testing programs. For instance, from the Kansas interviews, Clarke et al. determine

that “one problem with getting students to ‘buy into’ the state test was the delay in getting back the test results, with one-fifth of these educators noting that the results came back too late to be useful” (p. 76). Given the time required to submit, score, and return the tests, a delay in receipt and distribution of test scores is not unique to Kansas.

After conducting interviews with educators in states that administer low-, medium-, and high-stakes tests, Clarke et al. (2003) acknowledge,

The study findings showed a distinction between stakes and consequences. Specifically while mandated rewards and sanctions may be directed at one level or group in the system, their impact can extend in unexpected and undesirable directions. The most striking example in this study was a consistently greater impact on both students and educators at the elementary level, regardless of the stakes attached to the test results. (p. 95)

With low-stakes tests generating high anxiety at the elementary level, the impact of the accountability programs extend beyond the apparent stakes attached to the test.

Moreover, Ryan et al. (2007) propose the following:

students’ beliefs about if they want to do well on a test (i.e. goals, value), whether they can do well on a test (i.e. self-concept, self-efficacy), and how they feel during a test (i.e. worry or emotionality) are factors influencing math test performance. (p. 10)

Lau and Roeser (2002) argue that scores from a test engagement questionnaire assessing cognitive test-taking strategies, test mood, effort, and energy were predicative of science test scores regardless of students’ science aptitude, demographic characteristics, and other motivational constructs.

Impact of School Accountability on Curriculum and Teaching Practices

Without question, curriculum and teaching practices impact student achievement, and given that the central purpose of NCLB is to promote school accountability, an examination of the impact that these accountability practices have on curriculum and teaching is warranted.

Lattimore (2001) states, “Testing has become an attractive option for policymakers, both because it has the potential to affect the behavior of educators in the educational systems, and because it is often viewed by the public as a way to guarantee a basic level of quality education” (p. 57).

Furthermore, Lattimore argues, “High stakes testing affects both the content and sequence of instruction, and efforts to affect test scores directly increase as the testing date approaches” (p. 62). Concurring, Clarke et al. (2003) report,

As the stakes attached to the test results increased, the test seemed to become the medium through which the standards were interpreted; As the stakes increased, so too did the number of reported effects on classroom practice; As the stakes increased, interviewees reported a more negative impact on students, particularly elementary students, special populations, and students in urban districts. Taken together, these findings suggest that stakes are a powerful lever for effecting change but one whose effects are uncertain; and that a one-size-fits-all model of standards, tests, and accountability is unlikely to bring about the greatest motivation and learning for all students. (p. 94)

As a result of the impact on curriculum, one of the biggest complaints associated with testing is the concern that the test dictates the curriculum, resulting in a narrowing of curriculum.

Meredith Scriver of Advocates for Education, Wisconsin, summarizes the effect of school accountability on the curriculum and teaching practices: “Teaching to the test [is] a practice likened to memorizing an eye chart. With enough drill and rote work, even a person

with 20/150 vision can rattle off 'E-F-P-T-O-Z'. Of course this doesn't mean that person can truly see” (Students Against Testing, n.d., para. 35). Despite the inadequacies and because of the public outcry for accountability, standardized tests have emerged as the most cost-effective manner to ensure a level of quality control in education.

Validity and Reliability of Standardized Tests

With the ramifications associated with high-stakes testing, critics question the validity of the test scores. Those interviewed by Clarke et al. (2003) “had two main concerns about the validity of the test results. The first was that overtesting reduced students’ motivation to exert effort on state tests, thereby compromising the test’s ability to measure what they had learned” (p. 74). The second concern centered on whether or not schools and districts could be effectively compared based on test results because they are affected by “out-of-school factors” (Clarke et al., 2003, p. 74).

Another concern expressed by the educators interviewed by Clarke et al. (2003) pertains to the “lower-performing students’ arbitrary choice of answers” (p. 82) which also raises questions not only about the motivational power of the test but also the validity of the test results. In fact, Wise (2009) asserts, “moreover, whenever test-taking effort varies across examinees, there will be a differential biasing effect, which will introduce construct-irrelevant variance into test score data” (pp. 152-153). However, Wise further states,

Whenever the test scores have personal consequences for examinees (such as grades, diplomas, licensure, etc.), low effort is not generally considered a major validity threat.

In these situations, test performance is considered to be the responsibility of the examinee, and if the examinee chooses to not give good effort to the test, it is generally not viewed as a meaningful threat to the validity of score-based inferences. In contrast,

there are numerous measurement contexts in which the scores have important consequences for the test givers but little (or no) consequence for examinees. (p. 153)

In summary, students who do not put forth effort on low-stakes tests may introduce a detrimental construct-irrelevant variance into the data. On the other hand, the presence of high stakes for test takers minimizes this concern.

Impact of Incentives on Student Achievement on Standardized Tests

At the time of their study, Clarke et al. (2003) discovered that stakes for educators are fairly constant, but stakes for students increase or decrease as one moves from state to state. Across the country, there are a variety of accountability systems in place with educator accountability emerging as the common denominator. Stakes for students vary and often change within a state based on the age of students, with high school students having the greatest consequences. In interviews conducted in Kansas, Michigan, and Massachusetts, Clarke et al. discovered that

interviewees reported more negative than positive test-related effects on students.

Perceived negative effects included test-related stress, unfairness to special populations, and too much testing. Massachusetts interviewees were the most likely to note these negative effects, and Kansas interviewees, the least likely. (p. 73)

In Massachusetts, as a result of the implementation of the 10th grade graduation test, urban educators were apprehensive about potentially high failure rates and increased dropouts. In Michigan, although scholarship money was offered to 11th grade students who scored proficient on state tests, the educators in the large urban districts were the least likely to note that scholarship money provided an incentive for their students. In fact, data indicates that Caucasian, Asian, and affluent students are the most likely to receive the scholarships,

suggesting that Michigan's goal of increasing access to higher education through the program is not being accomplished. One reason offered as to why the scholarship incentive is not effective in motivating students in the large urban districts is based on a pessimistic mentality of the students. It is hypothesized that the students did not exert their best effort because they felt their chance of receiving the scholarship was unlikely (Clarke et al., 2003).

Consequences for Failing to Achieve

In most states, school accountability is based on the federal designation of Adequate Yearly Progress required by NCLB with consequences for schools and systems that consistently fail to achieve goals. As previously mentioned, consequences for failing schools can range from submission of an improvement plan to reconstitution of the school. In many states, the stakes are also high for students because "high school students who are unable to pass their state's accountability tests in multiple subject areas are not awarded high school diplomas" (Judson, 2007, p. 16). Based on interviews conducted, Clarke et al. (2003) reported that the effects of testing programs for high school students can result in demoralization rather than motivation of students. In fact, the majority of the interviewees commented that the tests have negatively affected students' view of education, particularly that of special education students. Despite the fact that students in Massachusetts who fail the 10th grade test can retake it up to four times before the end of high school, Clarke et al. discovered through interviews that students were frustrated and convinced themselves that they would not pass the test regardless of the number of opportunities and would instead have to drop out of school.

Prior to accepting a position as a middle school administrator, this researcher was a high school mathematics teacher in both an inner city and a suburban school. After witnessing first-hand the student disappointment of failing to pass the graduation exam, this researcher was

compelled to request permission from local school administration for the opportunity to develop a graduation exam remediation program. Students enrolled in this researcher's mathematics remediation classes were seniors who had failed the graduation exam numerous times. When these students initially entered the classroom, they lacked confidence, considered themselves dumb, and definitely hated math. Before these students could successfully learn math skills, their perception of their mathematical ability as well as their opinion of math had to be altered. Typical methods of instruction failed and did not interest these students. They needed to be provided mathematical opportunities they would enjoy. Many of the students in these remediation classes earned the credits necessary for graduation but simply could not pass the graduation exam. Throughout this researcher's teaching career, over 300 students were enrolled in these remediation classes with approximately 85% passing the mathematics portion of the graduation exam as a result of participation in the remediation program. The assistance provided in these classes was life changing for both the teacher and many of the students. Without a high school diploma, the students had limited career options. However, upon meeting the final graduation requirement and earning a diploma, there were more career opportunities for the students to pursue.

Summary

Because success in mathematics is often fundamental to future educational and economic opportunities, in this study, the investigation centers on mathematics achievement. As a result, this researcher analyzes students' grades earned in eighth grade mathematics courses and high school Algebra I courses to determine the relationship between course averages and the eighth grade ARMT and the 10th grade AHSGE mathematics test. The researcher also analyzes the students' levels of achievement on these tests, given that in Alabama, the stakes for students

increase in high school. Throughout the investigation, this researcher aspires to answer the following overall question: is there a difference in achievement levels on the mathematics portion of these tests when the stakes are raised for the students?

The purpose of this chapter is to present a review of literature which includes a discussion of the historical antecedents pertaining to the study, an overview of the school accountability program for the state of Alabama and other states, a discussion of the impact of student motivation on standardized tests, an analysis of the impact of school accountability practices on student performance as well as on curriculum and teaching practices, an examination of validity and reliability of standardized tests from the perspective of student motivation, and the impact of incentives and consequences of failing on student achievement on standardized tests. As the review of literature indicates, with the increased emphasis on accountability, the use of low-stakes data to make high-stakes decisions about program effectiveness is also on the rise. In order to make valid inferences about what students know and can do, it is crucial to understand the consequences of low and high stakes in testing contexts. Wise and DeMars's (2005) analysis of the impact of test-taking motivation on performance reveals that motivated students consistently outperform less motivated students. Although the impact of accountability programs vary, high-achieving and suburban students are most likely to be motivated, and low-achieving and at-risk students are most likely to be discouraged (Clarke et al., 2003).

CHAPTER III

METHODOLOGY

In this chapter, the procedures, including the research questions, hypotheses, research design, participants, and data collection techniques, used to determine if students' mathematics achievement levels are related to their achievement levels on state-mandated standardized tests are outlined. Also identified in this chapter are data necessary to complete the study. For example, academic data utilized in the statistical analyses are eighth grade mid-year and end-of-year mathematics course averages; high school Algebra I end-of-course averages; test scores from the mathematics portion of the ARMT administered during the students' eighth grade year; and test scores from the mathematics subtest of AHSGE typically administered during the students' 10th grade year, although a small group of ninth grade students are eligible to take the test. Students who completed Algebra I in the eighth grade are allowed to take the mathematics subtest of the AHSGE in the fall of their ninth grade year, but students who received high school credit for successfully completing Algebra I in the eighth grade were excluded from the sample.

Both the ARMT and the AHSGE have high stakes for the school, but only the AHSGE has personal consequences for the students. If necessary, students are allowed multiple opportunities to take the AHSGE. Ultimately, students who are unable to pass at least three of the five sections of the exam, two of which must be mathematics and reading, prior to graduation will not earn a diploma.

Statistically significant findings from this study could be used to predict if students are prepared to take the state-mandated standardized tests, allowing time for unprepared students to participate in intervention programs prior to test administration. Also, because ARMT test results are not received by the school until just prior to school opening in the fall, regression

equations could simulate whether or not a student scored proficient on the ARMT. Students predicted to fall below accepted proficiency levels could then be scheduled into intervention classes prior to the start of the next school year. If the sample data shows that relationships exist in the student achievement levels based on the stakes, additional studies to expand the demographic and socioeconomic diversity of students being studied will be recommended. If future studies indicate significant results, changes in assessment policy may be recommended for adoption at the federal, state, and local levels.

Research Questions

The following research questions form the basis that guides the researcher throughout the study:

1. Is there a relationship in students' high-stakes eighth grade mathematics mid-year course averages and their corresponding low-stakes scores on the mathematics portion of the ARMT?
2. Is there a relationship in students' high-stakes eighth grade mathematics end-of-course averages and their corresponding scores on the mathematics portion of the low-stakes ARMT?
3. Considering that end-of-course averages result in stakes equal to the graduation exam and that the AHSGE mathematics test is based on mathematics content taught in the eighth grade mathematics courses as well as Algebra I, another question of interest arises. Are students' high school Algebra I end-of-course mathematics grades related to students' scores on the mathematics portion of the AHSGE?
4. Is there a relationship in student achievement levels on the mathematics portions of the low-stakes ARMT and the high-stakes AHSGE?

5. In addition to the stakes associated with each test, do other demographic factors including gender, socioeconomic status, and ethnic group contribute to differences in the student test results?

In summary, the questions of interest ultimately concern the impact of the low and high stakes associated with each test in regard to student achievement. Therefore, the focus of the research questions pertains to inferences that can be made concerning student achievement levels, particularly when the stakes are raised for the students. In other words, does student achievement increase as the individual consequences associated with each test increase?

Hypotheses

The research questions previously posed led to the following hypotheses which were analyzed using appropriate statistical procedures:

1. A statistically significant relationship between students' high-stakes eighth grade, mid-year mathematics course averages and their scores on the low-stakes, mathematics portion of the ARMT does not exist.
2. A statistically significant relationship between students' high-stakes eighth grade end-of-course mathematics averages and their low-stakes, eighth grade ARMT mathematics test scores does not exist.
3. A statistically significant relationship between students' high-stakes high school Algebra I end-of-course scores and their high-stakes AHSGE mathematics test results exists.
4. A statistically significant relationship between students' mathematics achievement levels on the high-stakes AHSGE and their low-stakes ARMT does not exist.

5. When sample data is disaggregated by demographic factors including gender, socioeconomic status, and ethnic group, statistically significant differences in the group means do not exist.

Research Design

The research design for this study was selected to guide the researcher in answering the previously stated research questions. The research design differs from experimental or quasi-experimental designs in that the treatment is included by selection of participants rather than manipulation of variables. This study is an ex post facto design because the researcher analyzed existing data to determine if the stakes associated with the various factors impacted student achievement. The non-experimental, parametric and non-parametric statistical analyses were used to examine student mathematics achievement in the classroom and on standardized tests to determine if the stakes associated with each factor influenced students' achievement levels.

Participants

The participants for the study are currently enrolled in the Saraland City School System. Saraland is located in the suburbs of Mobile, Alabama. As of the 2010 Census, the population of Saraland was 13,405. The Saraland City School System is comprised of three schools: one elementary, one middle, and one high. The system opened in 2008; Saraland was the first city school system to break away from the Mobile County Public School System, the largest public school system in the state of Alabama whose jurisdiction encompasses over 100 schools.

In 2006, believing that a better education could be provided to students with local control, 70% of the citizens of Saraland voted in favor to form its own city school system. As an additional show of support, the citizens at the same time approved a half-cent sales tax increase to aid in funding the new city school system. When the system opened, there was one

elementary and one middle school in Saraland. Together, the two schools serviced students in kindergarten through ninth grade. Within months of opening the system, construction of a high school began. The new state-of-the-art school opened in January 2010.

During the first four years of operation, Saraland City Schools' student population has grown to approximately 2,200 students in grades K-12. Currently, middle school renovations are planned for summer of 2012, and construction of a new elementary school is set to begin during the fall of 2012. As a result of the growth and sense of community, there is a tremendous amount of pride and support for schools in the city of Saraland.

In this research study, the sample includes all students in the 11th grade with available eighth grade ARMT and 10th grade AHSGE mathematics test scores who are currently enrolled in the 2011-2012 class at the suburban public school system in Southeast Alabama. Students enrolled in the 11th grade class who were excluded from the sample may have been absent or not enrolled in a public school in Alabama during test administration dates. Additionally, students considered potential outliers were excluded from the study. Potential outliers were defined as academically gifted students who earned credit for high school algebra while in middle school.

Protection of Human Subjects

Prior to the collection and analyses of data, written permission to conduct the study was submitted, and permission to conduct the study was granted by the superintendent of the Saraland City School System as well as the University of West Florida Institutional Review Board (see Appendix A, B, and C). Upon notification of approval, data was provided to the researcher by the high school counselor in a coded manner using unique student identifiers rather than student names in order to protect students' identities. Because student identities are

confidential and no manipulation of the variables associated with students' end-of-course averages or standardized test scores occurred, there is no risk to the human subjects in this study.

Data Collection

During the proposal phase of this study, the researcher considered the types of student demographic and academic data available and selected data appropriate for use in this study. For example, eighth grade mathematics and high school Algebra I course grades were included because of course alignment with the ARMT and AHSGE objectives. Upon analysis of state course of study objectives, the mathematics required to pass the AHSGE is taught in seventh grade mathematics, eighth grade mathematics or pre-algebra, and Algebra I (Alabama Department of Education, 2003).

To summarize the AHSGE item analysis, the content of the mathematics portion of the exam includes 100 questions with 75% of the test focused on algebraic concepts and 25% derived from pre-geometric concepts. Although the content tested on the ARMT and AHSGE is not identical, numerous objectives overlap and are assessed on both tests. With respect to the ARMT format, the ARMT consists of multiple-choice, gridded-response, and open-ended questions while the AHSGE is comprised solely of multiple choice questions.

Students' ARMT results are provided to the school and students based on four categories. The categories defined by the state are as follows: "Level I – Does Not Meet Academic Content Standards, Level II – Partially Meets Academic Content Standards, Level III – Meets Academic Content Standards, and Level IV – Exceeds Academic Content Standards" (Alabama Department of Education, 2011, p. 9). In addition to the category classification, a raw score is also provided to each student administered the ARMT.

In contrast, students' AHSGE results are reported to the school and students using pass or fail dichotomous standards. Students scoring at least a 477 on the mathematics subtest of the AHSGE are notified that they achieved a passing score (Alabama Department of Education, 2003). However, for those who pass the exam, no additional information is made available from the state to the students, school, or school system. Conversely, students who fail the AHSGE are notified they did not meet standards. Unlike those who passed the test, these students also receive a scaled-score and a detailed mastery report by objective for each failed portion of the exam. Although these students are not formally notified of their current proficiency level based on the four categories used by the state to classify student performance, the numeric intervals which classify the student's proficiency as Level I or II is available (Alabama Department of Education, 2010). Although the numeric intervals used to determine category placement based on AHSGE raw scores is different from the one used to classify ARMT scores, the state titles the four categories in the same manner.

With respect to the analysis of sample demographics, this researcher collected the following information for each student in the sample: identification number, ethnicity, gender, socioeconomic status, and special education status. Additionally, the following academic data was collected for each participant in the sample: eighth grade ARMT mathematics scores, eighth grade mid-year mathematics course averages, eighth grade end-of-course mathematics averages, 10th grade AHSGE mathematics scores, and high school end-of-course Algebra I averages.

Course grades hold high stakes for the students because grade promotion or retention is based on student achievement of 60% or higher for the end-of-course average. The ARMT is defined as a low-stakes test since there are no individual consequences related to promotion or retention based on student performance on this test. Since a high school diploma is at stake for

students failing to meet proficiency standards on the AHSGE, the AHSGE is defined as a high-stakes test.

Data Analysis

Prior to beginning the statistical analyses, data was entered into an Excel spreadsheet and uploaded into Statistical Package for the Social Sciences (SPSS) version 19. Data was analyzed via SPSS using appropriate inferential statistical methods, particularly correlations. Both Spearman rank-order and Pearson product-moment correlation coefficients were utilized in this study. Since outliers have a disproportionate effect on the correlation coefficient, the academically gifted students who received high school credit for completing Algebra I in the eighth grade were excluded from the sample. For Research Questions 1 and 2, a Pearson product-moment correlation coefficient was calculated and regression analysis performed. Because of the ordinal nature of reported AHSGE results, for Research Questions 3 and 4, a non-parametric, Spearman rank-order correlation coefficient was calculated to determine the relationship between the high-stakes, end-of-course Algebra I averages and the high-stakes AHSGE scores as well as the low-stakes ARMT scores and the high-stakes AHSGE scores. A conservative alpha level of .05 was applied to all statistical tests conducted in the study.

Limitations of the Study

Efforts were made to develop an appropriate research design in order to answer the research questions and yield valid results. However, methodological limitations are a natural part of the research process and must be acknowledged. The following limitations were factors beyond the researcher's control that may possibly impact the results of the study and subsequent conclusions.

The ability to generalize the findings to a larger, more diverse population is limited because the sample used in the study is comprised of students from a suburban city school system in Southeast Alabama. Although the sample is representative of the student population at Saraland High School, the racial or ethnic homogeneity of the sample restricts extrapolations from this study. Aside from limitations resulting from the selection of the sample, a threat to internal validity emerges in that the ARMT was administered during the students' eighth grade year while the AHSGE was not administered until students' ninth or 10th grade year, presenting a maturation concern. Although both the ARMT and the AHSGE focus on similar content objectives, the raw scores are not based on the same ratio scale, complicating the data analysis process. Another threat to internal validity is mortality because from the eighth grade year until the 10th grade year, some students who initially qualified as members of the sample no longer attend the Saraland City School System, resulting in exclusion from the analysis. Although the researcher is an administrator in the Saraland City School System, the design of the study minimizes researcher bias because the study is ex post facto, and no manipulation of variables occurred. Another potential threat to validity is teacher effect. Teacher effect was not controlled for in terms of impact on student course grades; however, the anticipated impact in this study is minimal since students are randomly assigned to one of two possible teachers.

Summary

In this chapter, a description of the participants and the methodology used to conduct the study is identified. Additionally, throughout this chapter, a discussion of the procedures used to achieve the purpose of this study is outlined. The data was collected and analyzed to determine if student achievement was impacted when the level of stakes associated with the tests increases. The findings of the statistical analyses of the research are presented in Chapter 4.

CHAPTER IV

RESULTS

The purpose of this study is to investigate the effects of high- and low-stakes testing on student achievement. The following research questions served as the catalyst for the study:

1. Is there a relationship in students' high-stakes eighth grade mathematics mid-year course averages and their corresponding low-stakes scores on the mathematics portion of the ARMT?
2. Is there a relationship in students' high-stakes eighth grade mathematics end-of-course averages and their corresponding scores on the mathematics portion of the low-stakes ARMT?
3. Considering that end-of-course averages result in stakes equal to the graduation exam and that the AHSGE mathematics test is based on mathematics content taught in eighth grade mathematics courses as well as Algebra I, another question of interest arises. Are students' high school Algebra I end-of-course mathematics grades related to the students' scores on the mathematics portion of the AHSGE?
4. Is there a relationship in student achievement levels on the mathematics portions of the low-stakes ARMT and the high-stakes AHSGE?
5. In addition to the stakes associated with each test, do other demographic factors including gender, socioeconomic status, and ethnic group contribute to differences in the student test results?

The questions of interest ultimately concern the impact of the low and high stakes associated with each test in regard to student achievement. Therefore, the focus of the research questions pertains to inferences that can be made concerning student achievement levels,

particularly when the stakes are raised for the students. In other words, does student achievement increase as the individual consequences associated with the test increase?

Using SPSS as the statistical analysis software and a 95% confidence interval ($p < .05$) as the level at which statistical significance is declared, this researcher investigated the five research questions. Student academic data was available for 75 students. Because of the effect that outliers have on the correlation coefficient, potential outliers were identified as academically gifted students who earned high school credit for algebra while in eighth grade, and these students were excluded from the study. After the researcher excluded the potential outliers, 26 students who earned high school credit for algebra while enrolled in the eighth grade, the remaining 49 students' academic data was analyzed in this study.

Before the inferential analyses were conducted, the frequencies in the sample were tabulated with respect to gender, socioeconomic status, and ethnicity. Regarding gender, the sample of 49 students was comprised of 25 females and 24 males. Eligibility for free or reduced school lunch was evaluated to determine student socioeconomic status. In this sample, 19 students qualified for free lunch, three students qualified for reduced priced lunch, and 27 students paid full price for lunch. Additionally, the sample's ethnic composition was representative of the overall school population with 78% of the sample or 38 students best identifying with the Caucasian category, 20% or 10 students best identifying with the African American category, and 2% or one student best identifying with the Other category as representative of ethnic group. After examining the frequencies, the researcher calculated descriptive analyses of the data (Table 1).

Table 1

Descriptive Statistics for the Student Sample

Variable	<i>N</i>	Minimum	Maximum	<i>M</i>	<i>SD</i>
8th Grade ARMT Scale Score	49	648	727	682.00	20.390
8th Grade Mid-Year Course Average	49	40	92	69.39	10.735
8th Grade End-of-Year Course Average	49	49	91	72.06	9.549
Algebra I End-of-Year Course Average	49	65	89	76.88	6.521

In addition to the descriptive analysis of the sample, in this chapter, the inferential analysis for each question is summarized, and the results of these analyses are discussed. Following the initial examination of the sample data, the disaggregated analysis of student achievement based on gender, socioeconomic status, and ethnicity was conducted. The results of these additional analyses are also included in this chapter.

Data Analysis for Research Question 1

The first hypothesis is that a statistically significant relationship between students' high-stakes eighth grade mid-year mathematics course averages and their low-stakes eighth grade ARMT mathematics test scores does not exist. In order to test this hypothesis, a visual examination of the probable linearity of the data was conducted using a scatter plot (Figure 1). Based on the linear nature of the data present in the scatter plot as well as the ratio nature of the data, a Pearson product-moment correlation coefficient was calculated to measure the statistical significance of the relationship between students' eighth grade mid-year mathematics course

averages and their eighth grade ARMT mathematics test scores. The results of the analysis yielded an $r = .529$, $p < .001$.

Given the significance of the relationship, a regression analysis was conducted to determine the proportion of variance in ARMT scores that can be predicted using the mid-year eighth grade mathematics course averages. The results of the regression analysis yielded a coefficient of determination, $R^2 = .280$, $F(1,47) = 18.282$, $p < .001$, which indicated that 28% of the variance within ARMT scores can be predicted by mid-year eighth grade mathematics course averages. Additionally, the regression analysis indicated $\beta = 612.256$, $t(47) = 1.005$, $p < .001$. The linearity of the data and the line of best fit are also illustrated in the scatter plot.

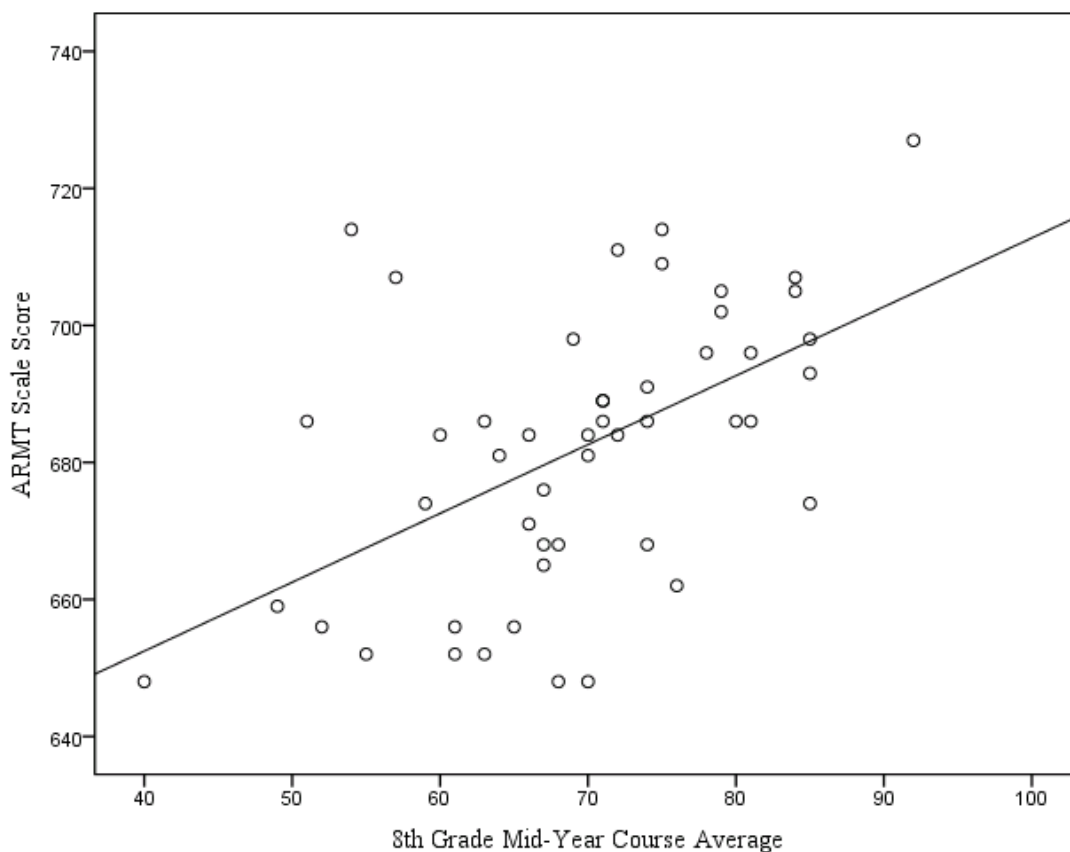


Figure 1. Scatter plot of eighth grade mathematics mid-year course averages and eighth grade mathematics ARMT scale scores.

Data Analysis for Research Question 2

The second hypothesis is that a statistically significant relationship between students' high-stakes eighth grade end-of-course mathematics averages and their low-stakes eighth grade ARMT scores does not exist. To test this hypothesis, a visual examination of the potential linearity of the data was conducted using a scatter plot (Figure 2).

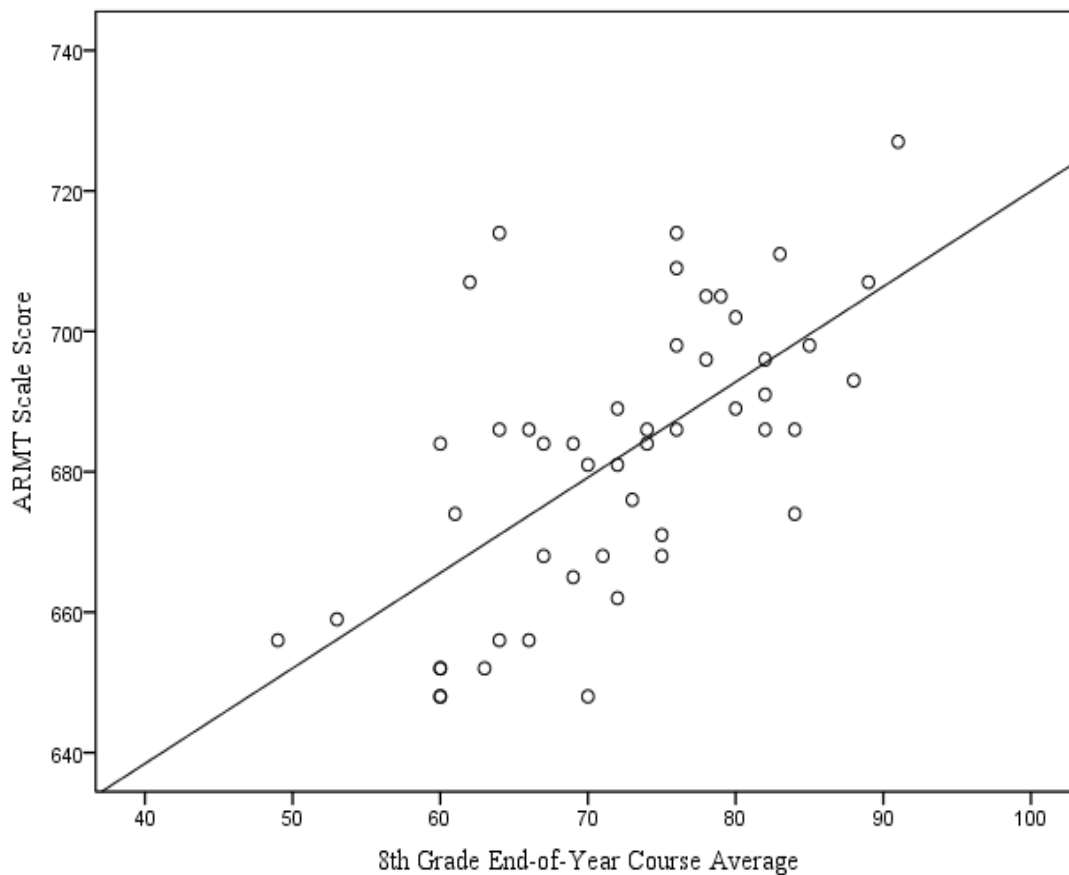


Figure 2. Scatter plot of eighth grade mathematics end-of-course averages and eighth grade mathematics ARMT scale scores.

Based on the clustering of data in a linear trend and the ratio nature of the data, a Pearson product-moment correlation coefficient was calculated to measure the statistical significance of the relationship between students' high-stakes eighth grade end-of-course mathematics averages

and their low-stakes eighth grade ARMT mathematics test scores. The results of the analysis indicated an $r = .636$, $p < .001$. Given the significance of the relationship, a regression analysis was conducted to determine the proportion of variance that can be predicted within ARMT scores by the eighth grade end-of-year mathematics course averages. The coefficient of determination was $R^2 = .405$, $F(1,47) = 31.980$, $p < .001$, which indicated that approximately 41% of the variance within ARMT scores can be predicted by end-of-year eighth grade mathematics course averages. Additionally, the regression analysis indicated $\beta = 584.087$, $t(47) = 33.448$, $p < .001$. The graphical representation of the data and the line of best fit are also illustrated in the scatter plot.

Data Analysis for Research Question 3

The third hypothesis is that a statistically significant relationship between students' high-stakes high school Algebra I end-of-course scores and their high-stakes AHSGE mathematics test results exists. To test this hypothesis, a visual examination of the potential linearity of the data was conducted using a scatter plot (Figure 3). Since the clustering of data exhibited a linear relationship and based on the categorical nature of the AHSGE results, a Spearman rank-order correlation coefficient was calculated to measure the relationship between students' high-stakes high school Algebra I end-of-course scores and high-stakes AHSGE mathematics test scores. The results of the analysis yielded an $r = .776$, $p < .001$. The relationship of the data is illustrated in a scatter plot.

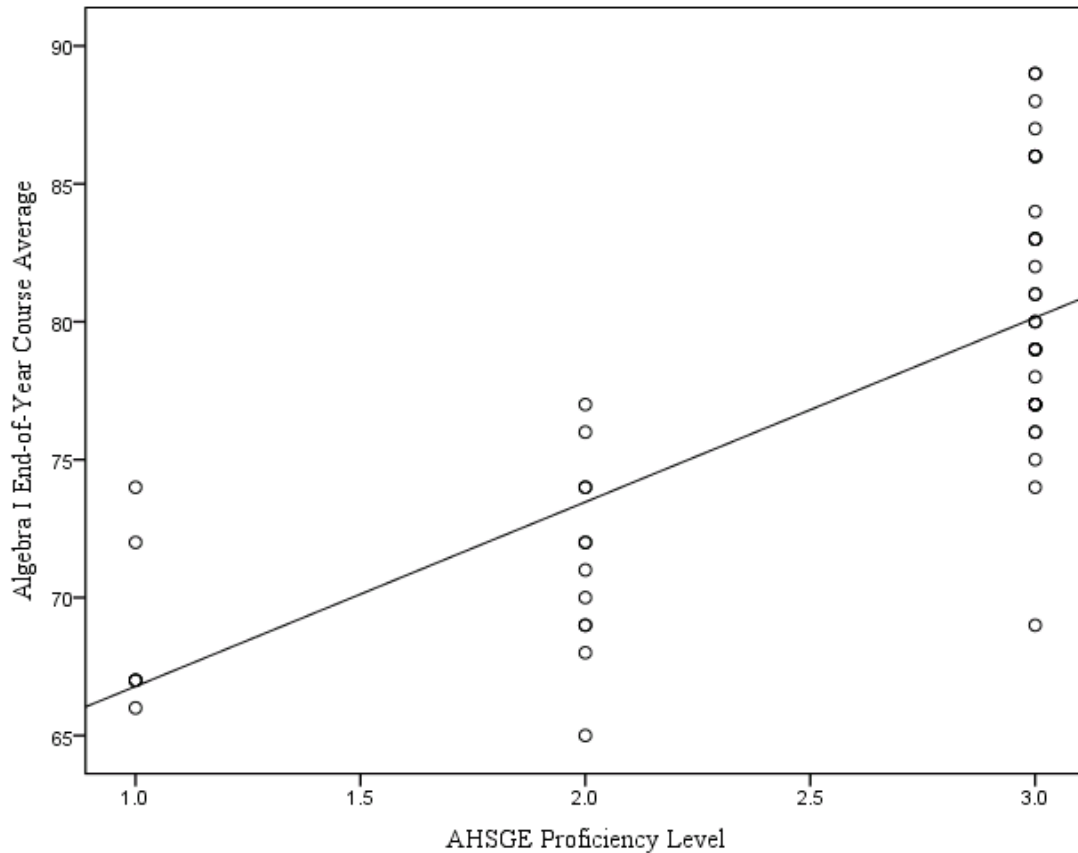


Figure 3. Scatter plot of AHSGE mathematics test scores and Algebra I end-of-course averages.

Data Analysis for Research Question 4

The fourth hypothesis is that a statistically significant relationship between student achievement levels on the high-stakes AHSGE and the low-stakes ARMT does not exist. To test this hypothesis, based on the categorical nature of the AHSGE scores and an examination of the constructed scatter plot (Figure 4), a Spearman rank-order correlation coefficient was calculated to quantify the relationship between student achievement levels on the high-stakes AHSGE and the low-stakes ARMT. The results of the analysis yielded an $r = .700$, $p < .001$.

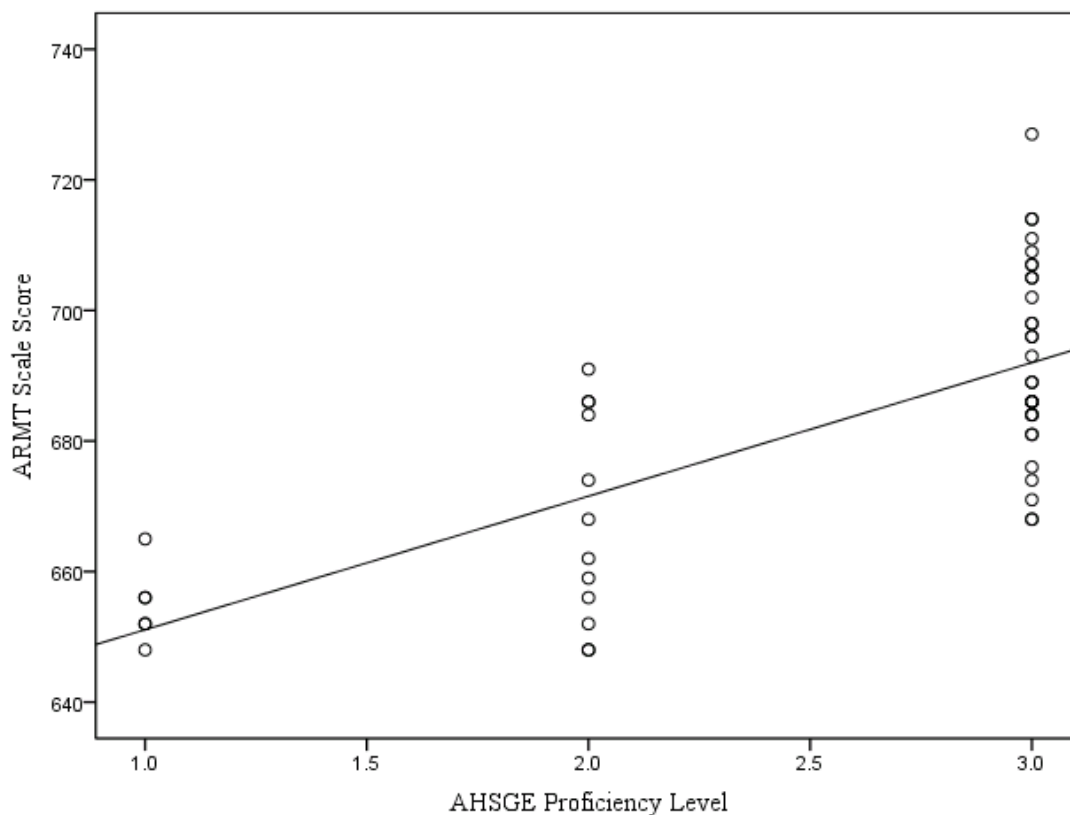


Figure 4. Scatter plot of ARMT mathematics test scores and AHSGE mathematics test scores.

Data Analysis for Research Question 5

The fifth hypothesis is statistically significant relationships do not exist when sample data is disaggregated by gender, socioeconomic status, and ethnic group. To test this hypothesis, the relationships between the academic variables and other demographic characteristics of the sample that may contribute to differences in the test scores were also examined using SPSS.

Since an analysis of variance (ANOVA) yields the same results as a *t* test when there are only two groups being compared, an ANOVA was used to test the difference in group means and report findings regardless of whether there were two or three groups in a factor (Tables 2-5).

Academic variables including eighth grade mid-year course averages, eighth grade end-of-course

averages, ARMT scores, and AHSGE mathematics scores were examined with respect to gender, ethnicity, and socioeconomic status.

Table 2

Disaggregated Data Analysis for 8th Grade Mid-Year Course Averages

Group	<i>N</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>p</i>
Gender				2.484	.122
Male	24	66.96	9.153		
Female	25	71.72	11.774		
Socioeconomic Status				.807	.452
Free Lunch	19	67.53	12.420		
Reduced Lunch	3	65.67	22.855		
Paid Lunch	27	71.11	7.653		
Ethnicity				2.399	.102
Caucasian	38	71.13	10.156		
African American	10	63.10	11.513		

Table 3

Disaggregated Data Analysis for 8th Grade End-of-Course Averages

Group	<i>N</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>p</i>
Gender				2.754	.104
Male	24	69.79	7.337		
Female	25	74.24	10.986		
Socioeconomic Status				.922	.405
Free Lunch	19	69.74	10.754		
Reduced Lunch	3	73.00	15.588		
Paid Lunch	27	73.59	7.943		
Ethnicity				3.281	.047
Caucasian	38	73.79	8.774		
African American	10	66.70	10.446		

Table 4

Disaggregated Data Analysis for ARMT Scores

Group	<i>N</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>p</i>
Gender				.077	.783
Male	24	681.17	21.487		
Female	25	682.80	19.689		
Socioeconomic Status				5.430	.008
Free Lunch	19	673.47	19.842		
Reduced Lunch	3	709.00	20.952		
Paid Lunch	27	685.00	17.737		
Ethnicity				7.258	.002
Caucasian	38	687.08	18.507		
African American	10	662.50	16.655		

Table 5

Disaggregated Data Analysis for Algebra I End-of-Course Averages

Group	<i>N</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>p</i>
Gender				3.056	.087
Male	24	75.25	5.825		
Female	25	78.44	6.880		
Socioeconomic Status				.265	.769
Free Lunch	19	76.37	6.978		
Reduced Lunch	3	79.33	8.737		
Paid Lunch	27	76.96	6.167		
Ethnicity				1.066	.353
Caucasian	38	77.58	6.408		
African American	10	74.20	6.925		

Despite the academic data being examined, the results of the ANOVA when gender was used as the grouping factor did not produce significant results. Additionally, an ANOVA was used to compare the differences in the means of the academic variables using socioeconomic

status as the grouping factor. With the exception of the analysis of the eighth grade ARMT mathematics scores, the analyses indicate no significant difference in group means based on socioeconomic status. On the other hand, there was a statistically significant difference in the group means when ethnicity was the grouping factor with the Caucasian students scoring higher than the African-American students on the eighth grade ARMT as well as the eighth grade end-of-course mathematics averages. However, upon completion of Algebra I, the statistically significant difference in the means was no longer present.

Summary

The purpose of this study is to investigate the effects of high- and low-stakes testing on student achievement. In this chapter, a summary of the results of the statistical analyses used to test the hypotheses is explained in narrative, graphic, and tabular forms. In Chapter 5, the conclusions and interpretations for the research questions are presented. In addition, a synopsis of the implications for student testing programs and recommendations for future research which stem from conclusions of this study are also discussed.

CHAPTER V

CONCLUSIONS

The purpose of this study is to investigate the effects of high- and low-stakes testing on student achievement. In this chapter, the conclusions and interpretations for the research questions are presented. Additionally, a synopsis of the implications for student testing programs and recommendations for future research that stem from this study is discussed.

Conclusions and Interpretations for Research Question 1

To answer the first research question, a statistical investigation of the relationship of students' high-stakes eighth grade mid-year mathematics course averages and their corresponding low-stakes scores on the mathematics portion of the ARMT was required. Using the sample data for these two variables, a scatter plot was constructed (Figure 1). Upon examination of the scatter plot, the clustering of the data in the scatter plot indicated the existence of a positive linear relationship between students' eighth grade mid-year mathematics course averages and their corresponding scores on the mathematics portion of the ARMT.

Accordingly, the Pearson product-moment correlation coefficient was calculated to assess the direction, nature, and strength of the linear relationship between students' mid-year eighth grade mathematics course averages and their corresponding scores on the mathematics portion of the ARMT which was administered in eighth grade. There was a strong positive correlation between the two variables. To summarize the nature and direction of the relationship, students who earn high eighth grade mathematics course averages also have high scores on the eighth grade mathematics portion of the ARMT. Since eighth grade mathematics mid-year course averages can be used to predict student mathematics performance on the eighth grade

administration of the ARMT, students predicted to fail to meet proficiency standards could be enrolled in an intense intervention program prior to the administration of the ARMT.

Conclusions and Interpretations for Research Question 2

To answer the second research question, a statistical investigation of the relationship of students' high-stakes eighth grade mathematics end-of-course averages and their corresponding scores on the mathematics portion of the low-stakes ARMT was conducted. Using the sample data for these two variables, this researcher constructed a scatter plot (Figure 2). In the scatter plot, the tightly clustered points indicated a relatively strong positive relationship between students' high-stakes eighth grade end-of-course mathematics averages and their low-stakes eighth grade ARMT scores.

The Pearson product-moment correlation coefficient was calculated to assess the direction, nature, and strength of the linear relationship between students' high-stakes eighth grade end-of-course mathematics averages and their low-stakes eighth grade ARMT scores. This calculation and the subsequent regression analysis indicated a strong positive correlation between the two variables. As a result, increases in students' high-stakes eighth grade end-of-course mathematics averages were directly correlated with increases in their low-stakes eighth grade ARMT scores. Given that student performance on the ARMT is not reported to the school until the end of the summer, school administration can use the predictive ability of the end-of-course averages to anticipate student achievement on the ARMT. With this information, students in need of additional academic support can be appropriately scheduled into intervention classes prior to the opening of a new school year.

Conclusions and Interpretations for Research Question 3

To answer the third research question, a statistical investigation of the relationship of students' high school Algebra I end-of-course mathematics grades to students' scores on the mathematics portion of the AHSGE was required. Using the sample data for these two variables, a scatter plot was constructed. The clustering of the data in the scatter plot indicated the existence of a positive relationship between students' Algebra I end-of-course scores and their AHSGE mathematics test results (Figure 3).

The Spearman rank-order correlation coefficient was calculated to assess the direction, nature, and strength of the relationship between Algebra I end-of-course scores and AHSGE mathematics test results. This calculation indicated a strong positive correlation between the two variables. As a result, increases in students' AHSGE mathematics test results are directly correlated with increases in their Algebra I end-of-course averages. Knowing that performance in Algebra I is related to achievement on the AHSGE, teachers and administrators can assess students' preparedness for the AHSGE and recommend additional academic intervention as necessary based on this information.

Conclusions and Interpretations for Research Question 4

To answer the fourth research question, a statistical investigation of the relationship of student achievement levels on the mathematics portions of ARMT and the AHSGE was performed. Using the sample data for these two variables, the researcher constructed a scatter plot. The clustering of the data in the scatter plot indicated the existence of a positive relationship between student achievement levels on the graduation exam and the ARMT (Figure 4).

The Spearman rank-order correlation coefficient was calculated to assess the direction, nature, and strength of the relationship of AHSGE and ARMT mathematics test results. This calculation revealed a strong positive correlation between the two variables which indicated that increases in AHSGE mathematics test scores coincide with increases in ARMT mathematics test scores. Regardless of the personal stakes associated with the state-mandated test, based on these analyses, student achievement was consistent on the mathematics portions of both the ARMT and the AHSGE. As a result, in this sample, student academic achievement was not impacted by the level of the personal consequences associated with the test. In the review of literature, high academic achievement of students regardless of extrinsic stakes associated with the test was evident when there was communication of expectations combined with a history of success in the school. With the school culture in Saraland City Schools promoting achievement and prior academic success of Saraland students documented on state tests, the study's statistical results are consistent with the literature review.

Conclusions and Interpretations for Research Question 5

To answer the fifth research question, following the whole group analyses examining the relationships between the academic variables, other demographic characteristics of the sample including gender, socioeconomic status, and ethnic group that may contribute to differences in the test results were also studied using appropriate statistical methods, specifically ANOVA. In addition to the stakes associated with each test, do other demographic factors including gender, socioeconomic status, and ethnic group contribute to differences in the student test results?

When the ANOVA was calculated for the academic data using gender as the factor, the analyses of the academic data indicated no statistically significant differences in the group means. With the exception of the eighth grade ARMT scores, the same was true when the data

was analyzed similarly using socioeconomic status as the grouping factor. Prior to making inferences based on this analysis, one must note that there are only three students in the reduced-lunch subgroup, with two of the three students scoring above the mean.

When the ANOVA was run using ethnicity as the grouping factor, there was a statistically significant difference in the group means for eighth graders on both the end-of-course mathematics averages and the ARMT scores. The Caucasian students' mean scores were significantly higher than the African-American students' scores. However, there was not a statistically significant difference in group means in the analysis using high school Algebra I data. Prior to supposition regarding the results of the achievement analysis based on ethnicity as the factor, one must consider the sample size of the African-American subgroup. Given there are 10 students in this subgroup, the power of this analysis is restricted, and inferences from these results are limited.

Implications for Student Testing Programs

In this study, based on the analyses of the sample data, low-stakes ARMT test scores are consistent with the high-stakes course averages as well as high-stakes AHSGE scores. Therefore, in schools similar to Saraland City Schools in which a culture of success and parental support is apparent, the low-stakes test scores reflect the academic knowledge of student test takers in a manner similar to the high-stakes measures of achievement. In contradiction with this researcher's initial supposition prior to beginning this study, this study's conclusions support that basing the evaluation of a school's academic program on low-stakes test results is a fair practice in the Saraland City School System.

Recommendations for Future Research

This study's results indicate at the eighth and ninth grade levels in schools in which there exists a culture of academic success and community support, student achievement in the classroom is positively related to student achievement on standardized tests regardless of the stakes associated with the test. To provide additional support for this argument, one may expand the study to include additional grade levels since students in grades three through eight are administered the ARMT. Additionally, to investigate the impact of stakes on student achievement further, one could expand the sample selection to include urban and rural schools both within Alabama and in other states. Studies are also recommended to include schools in which overall academic success and parental support are not optimal.

During 2010-2011, a progress monitoring testing program designed to gauge student preparation for the ARMT was purchased and implemented by the Saraland City School System. In order to assess the benefits of this formative assessment, a future research proposal could investigate the impact of the use of this tool on student achievement. Another extension of the study would include examination of student performance on other high- and low-stakes achievement tests.

Summary

With increased emphasis on accountability, the use of low-stakes test data to make high-stakes decisions about program effectiveness is on the rise. In order to make valid inferences about what students know and can do, it is crucial to understand the consequences of low and high stakes in testing contexts. With respect to student achievement and regardless of the level of consequences or rewards that were associated with each test, statistically significant relationships were found between the academic achievement variables in this study.

Additionally, through regression analyses, it was determined that grades earned in mathematics classes are predictors of scores earned on the state-mandated tests. As a result, in this study, the sample data revealed that student achievement did not increase as the individual consequences associated with the test increased. Rather, student achievement on the state-mandated tests was consistent with the overall mathematics achievement levels demonstrated in the classroom.

REFERENCES

- Alabama Department of Education. (2003). *Great expectations: A guide to Alabama's high school graduation exam*. Retrieved from http://www.alsde.edu/general/great_expectations.pdf
- Alabama Department of Education. (2009). *Alabama's first choice and supporting programs*. Retrieved from <http://www.mps.k12.al.us/Download.asp?L=2&LMID=232450&PN=DocumentUploads&DivisionID=5802&DepartmentID=5689&SubDepartmentID=&SubP=&Act=Download&T=1&I=77969>
- Alabama Department of Education. (2010). *Interpretive guide: Alabama accountability system*. Retrieved from <https://docs.alsde.edu/documents/91/2010%20Interpretive%20Guide.pdf>
- Alabama Department of Education. (2011). *Interpretive guide: Alabama accountability system*. Retrieved from <http://issuu.com/consolidatedpublishing/docs/2011aypinterpretiveguide>
- Assessment Network. (n.d.). Facts and FAQ's [Web page]. Retrieved from www.assessmentnetwork.net/facts_and_faqs.php
- Barry, C., Horst, S., Finney, S., Brown, A., & Kopp, J. (2009, October). *Modeling motivation over the course of a testing period: A mixture modeling approach*. Paper presented at the Northeastern Educational Research Association, Rocky Hill, CT.
- Budescu, D., & Bar-Hillel, M. (1993). To guess or not to guess: A decision-theoretic view of formula scoring. *Journal of Educational Measurement*, 30(4), 277-291.
doi:10.1111/j.1745-3984.1993.tb00427.x

- Clarke, M., Shore, A., Rhoades, K., Abrams, L., Miao, J., & Li, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from interviews with educators in low-, medium-, and high-stakes states*. Retrieved from ERIC database. (ED474867)
- Duckworth, A., Quinn, P., Lynam, D., Loeber, R., & Stouthamer-Loeber, M. (2011). Role of test motivation in intelligence testing. *Proceedings of the National Academy of Sciences, USA*, 108(19), 7716-7720. doi:10.1073/pnas.1018601108
- Goodling, B. (1997, August 13). More testing is no solution. *The Washington Post*. Retrieved from <http://www.washingtonpost.com>
- GreatSchools. (n.d.). No Child Left Behind (NCLB) requirements for schools [Web page]. Retrieved from www.greatschools.org/definitions/nclb/nclb.html
- Hanushek, E., & Raymond, M. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24(2), 297-327. doi: 10.1002/pam.20091
- Heubert, J., & Hauer, R. (Eds.). (1999). *High stakes: Testing for tracking, promotion, and graduation*. Retrieved from http://www.nap.edu/catalog.php?record_id=6336
- Judson, E. (2007). Retaking a high stakes mathematics test: Examination of school interventions and environments. *American Secondary Education*, 36(1), 15-30.
- Lattimore, R. (2001). The wrath of high stakes tests. *The Urban Review*, 33(1), 57-67. doi:10.1023/A:1010337013080
- Lau, S., & Roeser, R. (2002). Cognitive abilities and motivational processes in high school students' situational engagement and achievement in science. *Educational Assessment*, 8(2), 139-162. doi:10.1207/S15326977EA0802_04

Madison County Schools. (n.d.). *Alabama Reading and Mathematics Test (ARMT)*.

Retrieved from <https://www.madison.k12.al.us/departments/instruction/studentassessment/Pages/ARMT.aspx>

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).

O'Neil, H., Abedi, J., Lee, C., Miyoshi, J., & Mastergeorge, A. (2004). *Monetary incentives for low-stake tests* (Report 625). Retrieved from Center for the Study of Evaluation website: <http://www.cse.ucla.edu/products/reports/r625.pdf>

O'Neil, H., Sugrue, B., & Baker, E. (1995). Effects of motivational interventions on the national assessment of educational progress mathematics performance. *Educational Assessment*, 3(2), 135-157. doi: 10.1207/s15326977ea0302_2

Papay, J., Murnane, R., & Willett, J. (2010). The consequences of high school examinations for low-performing urban students: Evidence from Massachusetts. *Educational Evaluation and Policy Analysis*, 32(1), 5-23. doi:10.3102/0162373709352530

Ryan, K., Ryan, A., Arbuthnot, K., & Samuels, M. (2007). Students' motivation for standardized math exams. *Educational Researcher*, 36(1), 5-13. doi:10.3102/0013189X06298001

Students Against Testing. (n.d.). Test quotes [Web page]. Retrieved from www.nomoretests.com/quotes.php

U.S. Department of State. (n.d.). *Diplomacy in action*. Retrieved from www.state.gov/m/dghr/flo/c21998.htm

Wise, S. (2009). Strategies for managing the problem of unmotivated examinees in low-stakes testing programs. *The Journal of General Education*, 58(3), 152-166. doi:10.1353/jge.0.0042

Wise, S., & DeMars, C. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, 10(1), 1-17.

doi:10.1207/s15326977ea1001_1

APPENDIXES

Appendix A

Request for Institutional Permission to Conduct Research: An Analysis of the Effects of High- and Low-Stakes Testing on Student Achievement

Frankie Eubanks Mathis
670 Margiana Drive
Satsuma, AL 36572

April 2, 2012

Dr. Wayne Vickers
Superintendent
Saraland City School System
943 Highway 43 South
Saraland, AL 36571

Dear Dr. Vickers:

As you know, I am in the process of completing requirements for an Ed.D. at the University of West Florida. The research objectives for my dissertation involve a comparative analysis of Saraland City School students' ARMT and AHSGE mathematics test scores. I request that the data needed to conduct the study be accessed by individual student identification number with no names being communicated to me as the researcher.

The enclosed documentation is required by the University of West Florida to comply with federal policy regarding the protection of human research participants. On page 2 of the application, you will find a summary of the research objectives and assurances that no students will be placed at risk during this study.

I respectfully request your permission to conduct the study as outlined in the enclosed documents. Please contact me if you have any questions or concerns about the wellbeing of our students during this study.

Sincerely,

A handwritten signature in cursive script that reads "Frankie Eubanks Mathis". The signature is written in dark ink and is positioned above the printed name.

Frankie Eubanks Mathis

Appendix B

Institutional Permission to Conduct Research: An Analysis of the Effects of High- and Low-Stakes Testing on Student Achievement



SARALAND Board of Education

BOARD OF SCHOOL COMMISSIONERS

Mr. William 'Bill' Silver – President
Mrs. Renee Clarke – Vice-President
Dr. Lonnie Burnett – Board Member
Ms. Josephine Rodgers – Board Member
Rev. Gary Shockley – Board Member

943 Highway 43 S. Saraland, Alabama 36571 Phone- 251-375-5420

SUPERINTENDENT- L. Wayne Vickers, Ed.D.

April 4, 2012

To Whom It May Concern:

Mrs. Frankie Eubanks-Mathis has permission to access and utilize relevant documents pertaining to her research proposal, including ARMT and AHSGE data.

Sincerely,

Dr. L. Wayne Vickers
Superintendent

"Building a World Class Educational System for Our Children"

Appendix C

Letter from the Institutional Review Board for Human Subjects Granting Permission to Conduct Research: An Analysis of the Effects of High- and Low-Stakes Testing on Student Achievement

Ms. Frankie Eubanks
670 Margiana Drive
Satsuma, AL 36572

May 02, 2012

Dear Ms. Eubanks:

The Institutional Review Board (IRB) for Human Research Participants Protection has completed its review of your proposal titled "An Analysis of the Effects of High and Low Stakes Testing on Student Achievement," as it relates to the protection of human participants used in research, and granted approval for you to proceed with your study on 05-01-2012. As a research investigator, please be aware of the following:

- * You will immediately report to the IRB any injuries or other unanticipated problems involving risks to human participants.
- * You acknowledge and accept your responsibility for protecting the rights and welfare of human research participants and for complying with all parts of 45 CFR Part 46, the UWF IRB Policy and Procedures, and the decisions of the IRB. You may view these documents on the Research and Sponsored Programs web page at <http://www.research.uwf.edu/internal>. You acknowledge completion of the IRB ethical training requirements for researchers as attested in the IRB application.
- * You will ensure that legally effective informed consent is obtained and documented. If written consent is required, the consent form must be signed by the participant or the participant's legally authorized representative. A copy is to be given to the person signing the form and a copy kept for your file.
- * You will promptly report any proposed changes in previously approved human participant research activities to Research and Sponsored Programs. The proposed changes will not be initiated without IRB review and approval, except where necessary to eliminate apparent immediate hazards to the participants.
- * **You are responsible for reporting progress of approved research to Research and Sponsored Programs at the end of the project period 12-31-2012. If the data phase of your project continues beyond the approved end date, you must receive an extension approval from the IRB.**

Good luck in your research endeavors. If you have any questions or need assistance, please contact Research and Sponsored Programs at 850-857-6378 or irb@uwf.edu.

Sincerely,



Dr. Richard S. Podemski, Associate
Vice President for Research
And Dean of the Graduate School



Dr. Carla Thompson, Chair
IRB for the Protection of Human
Research Participants

CC: David Stout, Karen Rasmussen