2010

Examining Elementary Teachers' Perceptions of the Impact of High-Stakes Testing on Classroom Teaching Practices: A Mixed Methods Study

Latonya Borden-Hudson
University of Alabama at Birmingham

EXAMINING ELEMENTARY TEACHERS' PERCEPTIONS OF THE IMPACT OF HIGH-STAKES TESTING ON CLASSROOM TEACHING PRACTICES: A MIXED METHODS STUDY

by

LATONYA BORDEN-HUDSON

NATALIYA IVANKOVA, COMMITTEE CO-CHAIRPERSON
LINDA SEARBY, COMMITTEE CO-CHAIRPERSON
MARGARET RICE
WILLIAM BOYD ROGAN
GEORGE THEODORE

A DISSERTATION

Submitted to the graduate faculty of The University of Alabama and The University of Alabama at Birmingham, in partial fulfillment of the requirements for the degree of Doctor of Philosophy

BIRMINGHAM, ALABAMA

2010

EXAMINING ELEMENTARY TEACHERS' PERCEPTIONS OF THE IMPACT OF HIGH-STAKES TESTING ON CLASSROOM TEACHING PRACTICES: A MIXED METHODS STUDY

LATONYA BORDEN-HUDSON

EDUCATIONAL LEADERSHIP

ABSTRACT

The No Child Left Behind Act of 2001 (NCLB) requires schools to be held accountable for academic performance. It is believed the pressure of accountability will lead teachers to narrow the curriculum by engaging students in test preparation activities. The purpose of this two-phase, explanatory mixed methods study was to examine elementary teachers' perceptions of the impact of the Stanford Achievement Test 10 (SAT-10) and the Alabama Reading and Math Test (ARMT) on classroom teaching practices from a sample of third-grade, fourth-grade, and fifth-grade teachers in three large school systems in Alabama.

The purpose of the first, quantitative phase of the study, was to reveal teachers' perceptions of the impact of high-stakes testing on curriculum and instructional approaches, the amount of time spent on critical thinking skills, the amount of time spent on test preparation activities, and the perceived impact of state tests on students and teachers by surveying 123 third-grade through fifth-grade teachers in three large Alabama school systems. In the second, qualitative phase of this study, purposeful sampling strategy and maximal variation sampling strategy were employed to interview nine teachers who responded to the survey in the first, quantitative phase of the study to explore the results from the statistical tests in more depth.

Findings suggested urban teachers spent more time on critical-thinking skills than rural and suburban teachers, and low-socioeconomic, rural teachers experienced more stress caused by high-stakes testing than their geographical counterparts. All teachers independent of socioeconomic status or school geographical location reported they increased their focus on reading and math, which were the subjects assessed on high-stakes tests and de-emphasized subjects not tested such as social studies and science. Finally, most teachers reported they decreased the teaching of critical thinking skills due to the SAT-10 but increased the teaching of critical thinking skills due to the ARMT. Due to the lack of research regarding high-stakes testing in Alabama elementary schools, there was a need for teachers to discuss the specific impact of testing on classroom teaching practices because they work directly with students and are cognizant of the challenges that teachers face.

Keywords: high-stakes testing, curriculum, instruction, critical thinking, mixed-methods research

DEDICATION

This dissertation is dedicated to my very special parents Paul and Brenda Borden for the numerous sacrifices they made so that I could have a good education and a better quality of life. They have always taught me "I can do all things through Christ Jesus who hath strengthened me" and have gone above and beyond the call of duty to support me in accomplishing all of my goals. This dissertation is also dedicated to my husband, Ashley Hudson and to my sister, Dr. Temeka Borden for their relentless support and consistent encouragement. Finally, this dissertation is dedicated to my son, Ashton Hudson who served as the inspiration for me to complete the Ph.D.

ACKNOWLEDGEMENTS

First, I acknowledge my Lord and Savior, Jesus Christ for giving me the stamina to endure through this process. I know “through Christ Jesus I can do all things through him who hath strengthened me.”

I am especially thankful for the guidance and patience provided by my committee chairpersons, Drs. Nataliya Ivankova and Linda Searby. I appreciate their advice and the expertise they shared with me to enhance the quality of this study. Furthermore, I am grateful for the time, assistance, and encouragement given to me by the dissertation committee members: Drs. Margaret Rice, William Boyd Rogan, and George Theodore. Finally, I thank Mr. Carl Brezausek for the numerous hours he spent providing me with statistical consultation.

Without Jesus Christ, the dissertation committee members, and Mr. Carl Brezausek, I would not have been able to accomplish this great educational goal. I will forever be grateful for their encouragement and steadfast assistance.

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

CHAPTER

TABLE OF CONTENTS (Continued)

Page

CHAPTER

TABLE OF CONTENTS (Continued)

Page

LIST OF TABLES

Tables *Page*

LIST OF TABLES (Continued)

Table *Page*

LIST OF FIGURES

Figure *Page*

LIST OF ABBREVIATIONS

ADAW	Alabama Direct Assessment of Writing
ARMT	Alabama Reading and Math Test
ASA	Alabama Science Assessment
AYP	Adequate Yearly Progress
ELL	English Language Learner
FCAT	Florida Comprehensive Assessment Test
ITBS	Iowa Test of Basic Skills
MCAS	Massachusetts Comprehensive Assessment System
MLO	Maryland Learning Outcomes
MSPAP	Maryland School Performance Assessment Program
NCLB	No Child Left Behind Act
OLSAT	Otis-Lennon School Ability Test
SAT-10	Stanford Achievement Test 10
SOL	Virginia Standards of Learning Assessment
TAAS	Texas Assessment of Academic Skills
TAKS	Texas Assessment of Knowledge and Skills

CHAPTER 1

INTRODUCTION

Statement of the Problem

Because of the adequate yearly progress (AYP) mandated by the No Child Left Behind Act of 2001 (NCLB), there has been an increased focus on assessment (Center for Public Education, 2006). Administrators and teachers are being held accountable for funding that is tied to school performance based on students' test scores (Gardner, 2002). High-stakes tests can have a major impact on school systems and on teachers (O'Neill, 2003). Bonuses for teachers, school funding, and the control of a school or a school system can be impacted by standardized test scores (O'Neill). Hoffman, Assaf, and Paris (2001) argued, "Accountability through testing, for students, teachers, and administrators, is the key leverage point for policy makers seeking to promote educational reform" (p. 482). Some believe this pressure of accountability will lead teachers to narrow the curriculum by engaging students in a test preparation curriculum (Grant, 2004; Lane & Clement, 2002; Marchant, 2004).

Assessment is a procedure for determining the progress of the academic knowledge and skills of students (Wolf, 2007). Assessments are administered for diagnostic purposes and for accountability purposes (Wolf). The distinctive component of test for accountability purposes is that a sanction is administered if the results of the test are below an acceptable standard (Wolf). Before high-stakes tests became a part of accountability, teachers were only responsible for distributing test materials and administering the

test in a uniform fashion (Paris, 2000). The role of teachers has changed because high-stakes testing became a tool to change the curriculum and to change instruction (Paris, 2000).

High-stakes tests come in different forms; the most common form of high-stakes tests is standardized achievement tests (Altshuler & Schmautz, 2006). Many standardized achievement tests are used to make comparisons about student performance among students in various grade levels (Altshuler & Schmautz). For the purposes of this study high-stakes tests will include any standardized or state-mandated test to which sanctions are attached.

Although high-stakes testing has existed for hundreds of years, current high-stakes testing in the United States is under much debate (Cizek, 2001). The debate about high-stakes tests is frequently mentioned in the newspapers, on television, in education, business, legal, and political worlds as well as in research communities (Sloane & Kelly, 2003). Implications of the debate could be that tests are being used as an indicator of increased achievement as opposed to using it for formative, summative, and diagnostic purposes as it had been used in the past (Orlich, 2004; Sloane & Kelley). A second implication of the debate could be that some argue test scores can possibly provide inaccurate measures of student achievement (Marchant, 2004; Shepard, 2002; Smyth, 2008).

Perceptions of Opponents and Proponents of Testing

With regards to high-stakes testing there are opponents and proponents. Opponents believe high-stakes testing will lead teachers to change instruction from lifelong learning to teaching to the test which can affect the validity of the test (Smyth, 2008).

Additionally, there is very little debate that tests may impact the curriculum (Lipman, 2004; Madaus & Clarke, 2001; McNeil, 2000; Smyth; Vogler, 2005, 2006). Smyth (2008) believes high-stakes testing can narrow the curriculum by "teaching to the test," which can stifle teacher creativity, innovative instruction, and the use of a variety of teaching strategies for diverse students. Moreover, teaching to the test will result in teachers avoiding teaching higher-order thinking skills (Darling-Hammond, 2004). In the elementary school setting, narrowing of the curriculum has resulted in the reduction or dismissal of subjects that are not tested (Vogler & Virtue, 2007). Marchant (2004) asserted teachers will teach only content that is tested and will avoid content that is not tested. Gulek (2003) also believes when high stakes are associated with test results, teachers may be more likely to emphasize the objectives that will be covered on the test. Consequently, Grant (2004) believes test preparation will make teaching boring and unmindful of real-world problems and issues.

Proponents, specifically a majority of the American public, believe students will be motivated to work harder when rewards and sanctions are attached to tests (Kornhaber & Orfield, 2001). Americans also believe the tests administered to students are fair and are accurate at determining whether students should be promoted or should receive a high school diploma (Fuduka, 2007; Gardner, 2002; Kornhaber & Orfield). Those in favor of high-stakes testing argue that the tests are free from bias and are useful in depicting student achievement (Gardner). Americans also think high-stakes testing should be used to hold schools accountable for student learning (Kornhaber & Orfield).

In addition to proponents' views that high-stakes testing are useful, those in favor of high-stakes testing suggest tests will make parents and the public cognizant of the per-

formance of the students and of the school system (Kornhaber & Orfield). Proponents of high-stakes testing also believe testing mandated by the government allows control over local school systems, and the testing program serves as a means to ensure educational agencies are effectively doing their job (Egley & Jones, 2004). High-stakes testing is viewed by many proponents as a way to motivate students to pay attention and devote more effort at meeting the demands of the school system (Egley & Jones; Kornhaber & Orfield). Additionally, proponents believe the tests can be used as a tool to improve classroom practices (Gardner, 2002). For instance, some researchers have identified positive effects of high-stakes testing on teaching practices (Cizek, 2001; Jones, Jones, & Hargrove, 2003; Parke, Lane, & Stone, 2006). Specifically, some teachers have reported using a more student-centered approach rather than a teacher-centered approach to instruction (Jones et al., 2003). Others report they are implementing more teacher-centered approaches, and they believe this is an appropriate teaching method which has improved the quality of their teaching (Jones et al., 2003).

Impact of High-Stakes Testing on Students

Although the American public believes high-stakes tests will motivate students to learn, research suggests these tests do not encourage students to work harder (Kornhaber & Orfield, 2001; Paris 2000). Horn (2003) argued high-stakes testing can have a potentially negative impact on students. For example, Paris believes performance goals will lead to some students being reluctant to participate or try their best on assessments because they fret that their test results will consequently lead to them being labeled as having low ability. Second, Sloane and Kelly (2003) stated that labeling a child as low ability

can negatively affect his confidence in his ability to learn. Moreover, students who spend an excessive amount of time on high-stakes tests may possibly have problems applying the content they are learning to real-life situations (Cankoy & Tut, 2005).

Evidence suggests when teachers feel pressure to increase test scores, they in turn put pressure on students to increase achievement (Paris, 2000). Pressure from teachers to increase test scores may increase student anxiety about scores which may lead to a decline in student motivation and a decline in respect for teachers (Paris). However, a study conducted by Mulvenon, Stegman, and Ritter (2005) disaffirms these beliefs about students experiencing anxiety. The findings suggested the concerns regarding standardized testing were mainly misrepresented because most principals, counselors, parents, and students valued standardized testing and did not perceive an increase in stress or anxiety due to testing. In contrast, teachers who participated in this study had strong concerns about standardized testing and reported the greatest amount of anxiety due to testing.

Many believe NCLB is needed to ensure students of all populations are receiving an adequate education, but Jones et al. (2003) believe high-stakes testing of minorities, English Language Learners (ELLs), students with low socioeconomic backgrounds, and students with special needs make these special populations at risk of failure. The achievement gap on standardized tests among the races is a challenge that Americans and the educational systems are facing today (Fukuda, 2007). Student achievement for African American students has been a consistent problem in the United States (Haynes, 2008). In the past, minorities have not scored as well on high-stakes tests as whites (Haynes; Jones et al., 2003). This may be attributable to the fact that minorities have generally not had as many resources and in the past were denied equal educational opportuni-

ties (Jones et al.). Horn (2003) suggested minority students as well as special needs students are the groups mainly impacted by high-stakes tests.

Tests are used to determine a student's level of knowledge and are also used to compare what a student knows to other students his or her age. If a child's language is different from his peers then inaccurate conclusions can be drawn (Jones et al., 2003). Specifically, if a student does not understand English and obtains a low test score on a word problem in math, the score can be inaccurately interpreted concluding the student is below grade level in math (Jones et al.). The majority of the states in the United States are using standardized test scores to prove students are progressing as mandated by NCLB (Menken, 2008). Tests of an ELL's knowledge in a specific area of content will likely be greatly impacted by the ELL's English language proficiency (Garcia & Menken, 2006). Specifically, there are language and cultural complexities of test items that may lead to errors when determining high-stakes decisions (Abedi & Dietal, 2004; Solano-Flores & Trumball, 2003). Unfortunately, ELLs are being included in the administration of these tests which were never intended for ELLs (Menken).

Savage (2003) contended although standardized tests are well-written tests constructed by experts, they are designed to spread students along a continuum. High-stakes tests generally are multiple-choice assessments in which students fill in a bubble or a square (Grant, 2004). Hence, test items that every student is able to answer will not spread out students' scores. Standardized tests are administered once each school year and focus on content correlated to socioeconomic status and provide inaccurate data for holding schools and students accountable (Savage, 2003). Some researchers contend standardized testing has biases related to socioeconomic status (Gollnick & Chinn, 2002;

Smyth, 2008). Students from affluent families, schools, and/or school districts can afford tutorial resources or expensive materials to prepare for high-stakes tests (Smyth). However, lower-performing schools do not have funds to purchase such materials which leave minority and low socioeconomic students behind (Smyth).

Some opponents of high-stakes testing are also concerned about special education students. Jameson and Huefner (2006) argued it is almost impossible for schools with special-needs populations to comply with the mandates of NCLB because NCLB does not adequately fund the demand for highly qualified teachers. NCLB allows schools to exclude two percent of the student population from the state assessment; however, magnet and charter schools specialize in educating students with various exceptionalities (Smyth, 2008). Consequently, 100% of the school's population has special needs thereby making it irrelevant to report 98% of the school population's test results (Smyth, 2008). In cases such as these, AYP as mandated under NCLB will never be met by these schools leaving behind students with special needs. According to the Alabama School Journal (2008), 97.8% of the elementary schools in Alabama made AYP for the 2007-2008 school year. However, 16 of the 20 elementary schools statewide did not make AYP only because of the reading scores of special education students (Alabama School Journal, 2008).

Impact of High-Stakes Testing on School Geographical Locations

Hursh (2005) believes teachers, particularly in urban school districts, are under pressure to raise test scores which compel them to teach skills and knowledge that will be tested rather than more complex components of subjects. Hursh contended the pressure to

raise test scores will force weak students out of school before taking the required exam. Although the students in urban areas have the greatest need for support, students in urban schools are taught by teachers who are not well qualified and have little or no experience. Also, the teacher-to-student ratio in inner city classes is higher than in suburban and rural areas (Gollnick & Chinn, 2002). Because family incomes in rural areas are generally lower than in other school geographical locations, the challenges that low-socioeconomic students face are also significant for students in rural areas (Gollnick & Chinn). Additionally, in rural areas there are inadequate resources such as technology, advanced placement courses, and specialized courses (Gollnick & Chinn). Jimerson (2005) believes rural schools are disadvantaged by NCLB due to the school choice plan because if a student has to travel from a rural area to a school that is top-rated it could result in a long commute for students. In contrast to rural and urban schools, wealthy suburban schools are more likely to have qualified teachers, advanced-placement courses, numerous extracurricular activities, and adequate technology (Gollnick & Chinn). Although not all suburban schools have the resources to prepare students, they are more likely to have more educational resources than urban and rural schools (Gollnick & Chinn).

Impact of High-Stakes Testing in Alabama

Currently, the Stanford Achievement Test (SAT-10) is administered to students in Alabama in grades three through eight (Alabama State Department of Education, 2010). The purpose of the test is to compare individual and group performance with the performance of the norming group, to report relative strengths and weaknesses of individuals and groups, and to provide data to study changes in performance over time. Stu-

dents in grades three through eight are administered the reading and math portion of the SAT-10. The results provide Alabama educators, parents, and the public a comparison of the performance of Alabama's students, schools, school systems, and state to the related performance of the nation. Additionally, the Alabama Reading and Mathematics Test (ARMT) is administered to assess students' mastery of state content standards in reading and mathematics, to report individual and group performance, to report relative strengths and weaknesses of individuals and groups, and to provide data to study changes in performance over time. The ARMT is a criterion-referenced test. It consists of selected items from the SAT-10 which matches the Alabama state content standards in reading and math. Additional test items were developed to be included so that all content standards were fully covered. It is the combination of SAT-10 items and newly developed items that is known as the ARMT. The results are used for accountability for grades three through eight in meeting one of the requirements of NCLB.

Although critics believe high-stakes testing can have negative effects on classroom teaching practices, some researchers have identified positive effects of high-stakes testing on teaching practices (Cizek, 2001; Jones et al., 2003). Specifically in Alabama, according to the Alabama Association of School Boards (AASB) Magazine (2007), Alabama moved up in ranking of state progress from fifth to 22 under NCLB, which holds elementary schools in Alabama accountable for the scores on the ARMT for grades three through eight. Also, according to an article in the Alabama Education News (2008), 83% of all Alabama schools including elementary schools and high schools made AYP for the 2008-2009 school year. A total of 1,367 Alabama public schools were evaluated for the 2008-2009 AYP status based on 2007-2008 data from state-mandated tests. This includes

857 Title I schools which are high-poverty schools that receive federal funding. A total of 137 schools were identified as needing school improvement. Moreover, in 2007, 89 Title I schools were identified for school improvement. For the 2008 school year, only 79 Title I schools were identified as needing school improvement, which was an 18% decrease from the previous year indicating improved test scores.

The problem that needs further exploration is how the elementary curriculum is impacted by the pressure on teachers to have to increase scores on the SAT-10 and the ARMT although some researchers (Gollnick & Chinn, 2002; Smyth, 2008) believe standardized testing is biased toward particular groups. If standardized testing is biased toward groups such as minorities, low-socioeconomic students, and special needs students, then schools which serve a large number of diverse students will have a more difficult task of making AYP as opposed to other schools which serve less diverse students. The approach of having to measure all schools in the same way regardless of the fact schools serve different populations may impact how teachers teach or present the curriculum to students.

The Center for Public Education (2006) argued there are numerous amounts of literature written about the negative impact high-stakes testing has on instruction, but emerging research shows that high-stakes testing can be beneficial if certain conditions exist. Research of assessment is “overbalanced” by testimonials, essays, anecdotal reports, and even protests written in some educational publications (Center for Public Education). Moreover, Cimbricz (2002) argued there are a limited number of studies available on the impact of state-mandated testing: “studies that provide a richer, more in-depth understanding of the relationship between state-mandated testing and teaching in actual

school settings not only point toward important directions for continued research in this area, but are greatly needed" (pp. 15-16). There have been studies that examined the impact of particular state-mandated tests such as the Texas Assessment of Academic Skills (TAAS), the Virginia Standards of Learning Assessment (SOL), the Maryland School Performance Assessment Program (MSPAP), and the Florida Comprehensive Assessment Test (FCAT). However, there is a need for research from the teachers' perspectives of the specific impact of state mandated testing in Alabama which consists of the SAT-10 and the ARMT, because these tests may impact the curriculum and teaching practices differently from tests administered in different states. The teachers' perspectives are paramount because they work directly with the students and are the ones who decide specifically how the objectives of the curriculum and content will be taught.

Purpose of the Study

The purpose of this two-phase, explanatory mixed methods study was to examine elementary teachers' perceptions of the impact of the SAT-10 and the ARMT on classroom teaching practices from a sample of third, fourth, and fifth grade teachers in three large school systems in Alabama that serve rural, urban, or suburban communities. The purpose of the first, quantitative phase of the study, was to reveal teachers' perceptions of the impact of high-stakes testing on curriculum and instructional approaches, the amount of time spent on critical thinking skills, the amount of time spent on test preparation activities with administrators and in the classroom, and the perceived impact of state tests on students and on teachers by surveying 123 third through fifth grade teachers in three large Alabama school systems. In the second, qualitative phase of this study, purposeful

sampling strategy and maximum variation sampling strategy were employed to interview nine of the 123 third through fifth grade teachers who responded to the survey in the first, quantitative phase of the study to explore the results from the statistical tests in more depth. The explanatory sequential design was used to provide a better understanding of the research problem because the qualitative data extended and elaborated on the initial quantitative results (Creswell, 2008; Creswell & Plano Clark, 2007).

Research Questions

The study was guided by a mixed methods research question, quantitative research questions for Phase I, a qualitative research question for Phase II, and qualitative sub-questions. These research questions are presented in the following paragraphs.

Mixed Methods Research Question

The following mixed methods research question was used for this study:

1. What is the relationship between the school geographical location in which teachers serve, the socioeconomic status of the students they teach, and teachers' perceptions of the impact of high-stakes testing on classroom teaching practices and classroom instruction in grades three to five?

The qualitative findings from the interviews help explain the quantitative survey results in more depth (Creswell, Plano-Clark, Gutmann, & Hanson, 2003).

Quantitative Research Questions for Phase I

The following quantitative research questions were used for this study:

1. Are there differences in mean score on teacher curriculum approaches when controlled for by school geographical location and by socioeconomic status of the students taught?

2. Are there differences in mean score on teacher instructional approaches when controlled for by school geographical location and by socioeconomic status of the students taught?

3. Are there differences in mean score on class time spent on critical thinking skills when controlled for by school geographical location and by socioeconomic status of the students taught?

4. Are there differences in mean score on the amount of time spent on school-wide test preparation activities with administrators when controlled for by school geographical location and by socioeconomic status of the students taught?

5. Are there differences in mean score on the amount of class time spent on classroom test preparation activities when controlled for by school geographical location and by socioeconomic status of the students taught?

6. Are there differences in mean score on the perceived impact of state tests on students and teachers when controlled for by school geographical location and by socioeconomic status of the students taught?

Qualitative Central Research Question for Phase II

The central research question that guided the second, qualitative phase is as follows:

1. In what ways does the combination of school geographical location, the socioeconomic status of students taught, and high-stakes testing influence the classroom instruction of elementary teachers in grades three through five?

The five sub-questions to the qualitative central research question were the following:

1. What influence do teachers of grades three through five perceive high-stakes testing has on the curriculum?

2. What influence do teachers of grades three through five perceive high-stakes testing has on instructional practices?

3. What influence do teachers of grades three through five perceive high-stakes testing has on critical thinking skills?

4. What influence do teachers of grades three through five perceive high-stakes testing has on test preparation activities?

5. What specific influence do teachers of grades three through five perceive high-stakes testing has on students and teachers?

Theoretical Framework

The perceptual theory (Combs & Snygg, 1959) and the self-efficacy component of Albert Bandura's Social Cognitive Theory (1986) served as the theoretical framework

for this study. These two theories helped position the study as well as explain the study results.

Perceptual Theory

According to Combs, Richards, and Richards (1976), a theory is an organization of data or a way of examining the information to make it meaningful. Facts alone have very little value, but when the facts are placed in a framework they become easy to understand or deal with problems. Studies of individuals in psychology are conducted from two frames of reference which are the external frame and the internal frame. The external frame of reference consists of observing what a person does. The internal frame of reference consists of one's perceptions and meanings and seeking to understand how things appear to an individual and what his personal experiences are.

"All behavior is a result of an individual's personal meanings or perceptions" (Combs et al., 1976, p. 16). The perceptual approach seeks to understand the behavior of the individual from his own point of view (Combs et al.). Behavior is a response to situations that occur to an individual. Individuals do not behave to the facts as other individuals see them, but instead behave according to the facts as they see them (Combs & Snygg, 1959). Additionally, individuals behave according to one's attitudes, goals, beliefs, or purposes (Combs et al.). The factors that determine the behavior of individuals are the experiences of the individual at the time of behavior.

The experiences of an individual are referred to as perceptions and the complete field of the perceptions is called the perceptual field. Perceptions do not have an impact on behavior in an isolated fashion, but each perception is meshed in a complex group of

other perceptions called the perceptual field (Combs et al.). Combs et al. concludes that "The perceptual field is the entire universe, including himself, as it is experienced by an individual at the instant of action" (Combs et al., 1976, p. 22).

Combs and Snygg (1959) stated perceptions must be accessible to individuals because behavior is related to our perceptual field, and effective behavior occurs from a wide field of perceptions. The society in which we live requires flexible individuals who possess a broad range of perceptions. In order to function in society, individuals must be able to adapt to situations.

Self-Efficacy Theory

According to Bandura (1986), "Perceived self-efficacy is defined as people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performance" (p. 391). Bandura (1994) contended that beliefs about one's self-efficacy influence how one behaves and motivates himself. Teacher efficacy also influences teachers' enthusiasm, persistence, commitment, and instructional strategies in addition to students' motivation, achievement, and beliefs about self-efficacy (Tschannen-Moran & Hoy, 2001). Teachers who are knowledgeable in the subject matter they teach and who have a high sense of efficacy about their teaching capabilities can motivate struggling learners and enhance their cognitive development (Ashton, 1985; Gibson & Dembo, 1984). School practices that include "lock-step sequences of instruction," ability grouping, and competitive practice can contribute to the failure of many students (Bandura, 1986, p. 417). When this occurs students will rank themselves among their peers, and reputations with regards to ranking are not changed easily (Bandura,

1986). A diversified classroom consists of individualized instruction, and students compare their progress to themselves as opposed to their peers (Bandura, 1986).

Individuals who believe they will succeed increase resilient self-efficacy (Bandura, 1994). Effective teachers increase beliefs about efficacy to deal with challenges that may occur. Emotional and physiological states include anxiety, fatigue, and stress that influence beliefs about efficacy. Bandura concluded people's beliefs about their efficacy influence what they do as a group, the amount of effort they place into a task, their endurance when their efforts do not yield quick results, and their probability of success.

The self-efficacy theory and the perceptual theory of behavior were used as the theoretical framework to guide the study regarding teachers' perceptions of high-stakes testing on classroom teaching practices. The researcher sought to understand teachers' perceptions of the impact of high-stakes testing on classroom teaching practices through the use of surveys and follow-up face-to-face interviews. This study will add to the literature bases regarding the perceptual theory, the self-efficacy theory, and high-stakes testing in the elementary school setting.

Significance of the Study

Due to the lack of research regarding specific state-mandated testing in Alabama, which is the combination of the SAT-10 and the ARMT, there is a need to explore the specific impact of these particular tests on classroom teaching practices in the state of Alabama. The participants in this study were provided an opportunity to share their specific experiences of high-stakes testing. This study added to the literature on high-stakes test-

ing by addressing the impact of the combination of the SAT-10 and the ARMT on elementary teaching practices in Alabama.

Smyth (2008) believes high-stakes testing can evoke negative side effects such as loss of teacher decision making power. Also, Flores and Cooke (2003) argued limiting a teacher's decision making power can stifle innovation when seeking to meet students' needs and can lead to feelings of frustration among teachers and their de-professionalization. This study gave teachers a voice to speak about the impact of testing on classroom instruction which may improve teacher's decision making power and creativity in the teaching profession.

Policymakers make decisions about how schools should be held accountable without consulting those who are affected by the use of high-stakes testing (Assaf, 2006). Because teachers work directly with students, they are fully cognizant of the reality of schools and the challenges that teachers face. Consequently, school and district administrators as well as policymakers will profit from this study because they will gain a deeper insight into teachers' perceptions of how high-stakes testing impacts classroom instruction which in turn impacts students.

Smyth (2008) concluded the impact of high-stakes testing mandated by NCLB has led to minority students, students from low socioeconomic backgrounds, students with special needs, and ELLs not receiving equal educational opportunities. Most importantly students from disadvantaged backgrounds will profit from the study because the findings from this study could lead to a fair and balanced approach to assessing student learning and schools which may lead to more equal educational opportunities for these

particular groups of students. It is anticipated that the results of this study could be used to influence educational decisions regarding testing in Alabama and nationwide.

Assumptions of the Study

The study was based on the following assumptions:

1. The teachers answered the survey accurately.

2. The teachers answered the survey honestly.

3. Teachers volunteered to participate in the study.

4. The analyses of the quantitative data were unbiased.

5. The perceptions of the participants in the qualitative phase of the study are a reality and represent their truth space.

Limitations of the Study

The following limitations will apply to this study:

1. Teachers from only three school systems in Alabama participated in the study which might not be representative of the entire teacher population of Alabama.

2. The results of the study are generalizable to only the three school systems that participated in the study.

3. Convenience sampling was employed as opposed to random sampling in the quantitative phase of the study, which limits the generalizability of the results to the entire population.

4. A small quantitative sample was obtained thereby limiting the generalizability of the results

5. The sample population of urban teachers in this study consisted of only 22 of the 123 participants (17.9%).

6. Due to the interpretive nature of qualitative research, the findings of the qualitative phase of the study may not be representative of the entire population; however, the results are transferable from one elementary school setting to another elementary school setting with similar contexts.

7. Due to the subjective nature of qualitative research, the researcher might have introduced her biases in the interpretation of the results of this study.

8. The researcher attempted to obtain three participants from each of the school geographical locations: rural, urban, and suburban. However, the researcher was only able to obtain two urban teachers to participate in the second, qualitative phase of the study. Consequently, the researcher conducted four interviews with rural teachers to compensate for the lack of a third urban participant. A fourth rural teacher was selected because the urban and rural teachers had similar teaching contexts. Specifically, these teachers both taught students of low-socioeconomic status.

Definitions of Terms

Alabama Reading and Math Test (ARMT): The ARMT is a criterion-referenced test consisting of selected items from the SAT-10 which matches the Alabama state content standards in reading and math. Additional test items were developed to be included so that all content standards were fully covered. It is the combination of SAT-10 items and newly developed items that is known as the ARMT.

Criterion-referenced Test: "how well one does on a test based on the meeting of criteria or mastering a standard" (Marchant, 2004, p. 3).

High-stakes Tests: "testing to which consequences is attached" (Cizek, 2001, p. 19). For the purposes of this study, high-stakes testing will refer to the administration of the SAT-10 and the ARMT.

Mixed Methods Designs: "Mixed methods designs are procedures for collecting, analyzing, and linking both quantitative and qualitative data in a single study or in a multliphase series of studies" (Creswell, 2005, p. 53).

No Child Left Behind: a 2002 federal law that requires yearly testing in grades three through eight in reading and in math (Crocker, 2003)

Norm-referenced: "how well an individual does on the test is based on a comparison to a large group of test takers" (Marchant, 2004, p. 2)

Qualitative research: "a type of educational research in which the researcher relies on the views of participants, asks broad, general questions, collects data consisting largely of words (or text) from participants, describes and analyzes these words for themes, and conducts the inquiry in a subjective, biased manner" (Creswell, 2005, p. 39)

Quantitative research: "a type of educational research in which the researcher decides what to study, asks specific, narrow questions, collects numeric (numbered) data from participants, analyzes these numbers using statistics, and conducts inquiry in an unbiased, objective manner" (Creswell, 2005, p. 39)

Stanford Achievement Test 10 (SAT-10): A standardized test used to compare individual and group performance with the performance of the norming group, to report

relative strengths and weaknesses of individuals and groups, and to provide data to study changes in performance over time (Alabama State Department of Education, 2010)

Stratified Sampling: to "divide the population on some specific characteristic" (Creswell, 2005, p. 148)

Title I: Title I of the Elementary and Secondary Education Act of 1965 provides financial aid to schools with a large percentage of low-incomes students to ensure all students meet state academic standards (U.S. Department of Education, 2010).

Organization of the Study

This study is organized into six chapters. Chapter 1 introduces the study, a statement of the problem, the purpose of the study, research questions, a theoretical framework, and the significance of the study. Also, the first chapter includes limitations of the study, assumptions of the study, and definitions of relative terms, an organization of the study, and a summary. Chapter 2 is the literature review which begins with the theoretical framework to guide the study which is the perceptual theory and the self-efficacy theory. The literature review also consists of an overview of the accountability movement, perceptions of high-stakes testing, the impact of high-stakes testing on the curriculum, the impact of high-stakes testing on teaching practices, the impact of high-stakes testing on students, and high-stakes testing in Alabama for elementary students. Chapter 3 describes the research designs, sampling procedures, data collection and analyses, legitimization, ethical considerations, and a summary. Chapter 4 reports the results of the quantitative data. Chapter 5 reports the results of the qualitative data. Chapter 6 provides a discussion

of the findings from the quantitative and qualitative study phases and resulting meta-inferences, conclusions, recommendations for future research, and implications.

Summary

The increased emphasis on accountability has been under much debate (Cizek, 2001). Researchers report special populations of students such as minorities, low SES students, and students with special needs are negatively impacted by high-stakes testing leaving behind the students NCLB intended to help (Gollnick & Chinn, 2002; Haynes, 2008; Huefner, 2006; Hursh, 2005; Jones et al., 2003; Savage, 2003; Smyth, 2008). Some teachers report negative effects of high-stakes testing on teaching practices while others report high-stakes testing has positively impacted teaching (Cankoy & Tut, 2005; Jones et al.; Paris & Urdan, 2000). Although there is an overabundance of literature written about the negative impact high-stakes testing has on instruction, emerging research shows that high-stakes testing can be beneficial if certain conditions exist (Center for Public Education, 2006. There have been studies that examined the impact of particular state-mandated tests such as the TAAS, the SOL, the MSPAP, and the FCAT. However, there was a need for research from the teachers' perspective of the specific impact of state mandated testing in Alabama which consists of the SAT-10 and the ARMT. The purpose of this two-phase, explanatory mixed methods study was to examine teachers' perceptions of the impact of the SAT-10 and the ARMT on classroom teaching practices of third, fourth, and fifth grade teachers in Alabama.

CHAPTER 2

LITERATURE REVIEW

This chapter includes a review of the literature related to the impact of high-stakes testing on classroom teaching practices. The literature review begins with the theoretical framework to guide the study which consists of the perceptual theory and the self-efficacy theory. The literature review also consists of an overview of the accountability movement, perceptions of high-stakes testing, the impact of high-stakes testing on the curriculum, the impact of high-stakes testing on teaching practices, the impact of high-stakes testing on students, the impact of high-stakes testing in various school geographical locations, and high-stakes testing in Alabama for elementary students. A map of the conceptual framework for this study is presented in Appendix A.

Although numerous articles have been written about the effects of high-stakes testing on instruction, most of the literature is opinion as opposed to empirical research (Center for Public Education, 2006). Consequently, this chapter will reflect the literature thereby documenting the opinions that have emerged and reporting the very few empirical studies that have been conducted. This study was conducted to add much needed empirical research to the literature on high-stakes testing and its influence on elementary teaching practices.

Theoretical Framework

Perceptual Theory

According to Combs et al. (1976), a theory is an organization of data or a way of examining the information to make it meaningful. Facts alone have very little value, but when the facts are placed in a framework they become easy to understand or deal with problems. Studies of individuals in psychology are studied from two frames of reference which are the external frame and the internal frame. The external frame of reference consists of observing what a person does. The internal frame of reference consists of one's perceptions and meanings and seeking to understand how things appear to an individual and what his personal experiences are.

"All behavior is a result of an individual's personal meanings or perceptions" (Combs et al., 1976, p. 16). The perceptual approach seeks to understand the behavior of the individual from his own point of view (Combs et al). Behavior is a response to situations that occur to an individual. Individuals do not behave to the facts as other individuals see them, but instead behave according to the facts as they see them (Combs & Snygg, 1959). Additionally, individuals behave according to one's attitudes, goals, beliefs, or purposes (Combs et al.). The factors that determine the behavior of individuals are the experiences of the individual at the time of behavior.

The experiences of an individual are referred to as perceptions, and the complete field of the perceptions is called the perceptual field (Combs et al., 1976). Perceptions do not have an impact on behavior in an isolated fashion, but each perception is meshed in a complex group of other perceptions called the perceptual field (Combs et al.). "The per-

ceptual field is the entire universe, including himself, as it is experienced by an individual at the instant of action" (Combs et al., p. 22).

A field has four properties, which include stability, fluidity, direction, and intensity (Combs et al., 1976). In regards to stability, the perceptual field is not organized, but the organization of the field gives it a sense of stability. Each individual needs a stable, predictable, and organized field to live successfully. Second, the perceptual field is always changing and can be difficult to study at times. However, the fluidity of the field makes change in behavior feasible and capable of adapting to changing conditions in an effort to live and obtain satisfaction. Fluidity allows for "learning, reasoning, remembering, forgetting, and creativity" (Combs et al., 1976, p. 27). Moreover, the perceptual field constantly has direction. The perceptions are not groups of irrelevant stimuli. Instead, the field of an individual is organized according to his need and the activity by which an individual seeks to satisfy the need. What the individual perceives is related to the influence of the needs of the individual. Lastly, the intensity of occurrences experienced by the individual will be related to the differentiation and levels of cognizance. The perceptual field does encompass the entire universe in which individuals are aware, but individuals are not aware of all parts with similar degrees of understanding at a given moment.

Communication is the process of gaining understanding of another individual's perceptual field, and this can occur only when common characteristics already exist (Combs et al., 1976). Individuals who have comparable experiences are likely to have similar characteristics in their phenomenal fields. Consequently, they will show common trends in their behaviors. Individuals feel more comfortable with others whose phenomenal fields are common with those individuals. Because individuals with common cha-

racteristics in their phenomenal fields behave alike, they can more feasibly determine what the other individual will do and how one will react to his or her behavior. The overlap in phenomenal fields makes communication possible.

Combs and Snygg (1959) stated, in order to function in society, individuals must be able to adapt to situations. Otherwise, individuals will become at risk for inappropriate behavior. Perceptions must be accessible to individuals because behavior is related to our perceptual field, and effective behavior occurs from a wide field of perceptions. Hence, restrictions placed on the perceptual field will affect an individual's capability to deal with life. Accordingly, an individual will feel threatened when he perceives himself as not being able to deal with situations. Individuals possibly experience threat when there is "inconsistency between self and the experiences of the external world, or as a result of inconsistencies between two aspects of self" (Combs & Snygg, 1959, p. 180). Individuals may also feel threat when the world around them rapidly changes. The society in which we live requires flexible individuals who possess a broad range of perceptions.

The adequate self is "a self capable of dealing effectively and efficiently with the exigencies of life, both now and in the future" (Combs et al., 1976, p. 56). To obtain self-adequacy each individual must maintain the organization in which he exists and increase his adequacy of which he is cognizant. Each person seeks to not just the maintenance of self, but the development of an adequate self (Combs et al.).

Self-Efficacy Theory

Bandura (1986) states that "perceived self-efficacy is defined as people's judgments of their capabilities to organize and execute courses of action required to attain

designated types of performance" (p. 391). Bandura (1994) contended that beliefs about one's self-efficacy determine how one behaves and motivates himself. The self-efficacy concept is applicable to the field of education.

Bandura (1986) contended that the school setting is an agent for cultivating the cognitive self-efficacy. Teacher efficacy influences teachers' enthusiasm, persistence, commitment, and instructional strategies in addition to students' motivation, achievement, and beliefs about self-efficacy (Tschannen-Moran & Hoy, 2001). Teachers who are knowledgeable in the subject matter they teach and who have a high sense of efficacy about their teaching capabilities can motivate struggling learners and enhance their cognitive development (Ashton, 1985; Gibson & Dembo, 1984). School practices that include "lock-step sequences of instruction," ability grouping, and competitive practice can contribute to the failure of many students (Bandura, 1986, p. 417). When this occurs, students will rank themselves among their peers, and reputations with regards to ranking are not changed easily (Bandura). A diversified classroom consists of individualized instruction, and students compare their progress to themselves as opposed to their peers (Bandura).

Bandura concluded people's beliefs about their efficacy influence what they do as a group, the amount of effort they place into a task, their endurance when their efforts do not yield quick results, and their probability of success. Individuals who believe they will succeed increase resilient self-efficacy (Bandura, 1994). Effective teachers increase beliefs about efficacy to deal with challenges that may occur. Emotional and physiological states include anxiety, fatigue, and stress that influence beliefs about efficacy.

Perceptions of teachers were examined in a study conducted by Berger (2006) in which the self-efficacy component of the Social Cognitive Theory served as the theoretical framework for the study. The purpose of Berger's study was to determine the relationship between teacher stress and high-stakes testing and whether the perceived stress among teachers differed between urban and rural teachers in the state of Virginia. The survey administered by Berger was completed by 150 teachers from a stratified random sample of 219 teachers with a return rate of 68% to determine teachers' stress. A *t* test for independent samples concluded rural teachers had higher stress scores than urban teachers. The findings suggested rural teachers scored higher stress scores than urban teachers for the variables of personal stress, frustration with student effort, and teacher morale. Berger determined the demands on the teachers to increase high-stakes test scores increased their stress which also negatively influenced classroom discipline, self-esteem, and their emotional development.

The perceptual theory and the self-efficacy theory were used as the theoretical framework to guide this study regarding teachers' perceptions of high-stakes testing on classroom teaching practices. The researcher sought to understand teachers' perceptions of the impact of high-stakes testing on classroom teaching practices through the use of surveys and through the use of face-to-face interviews. This study will add to the literature bases regarding the perceptual theory, the self-efficacy theory, and high-stakes testing in the elementary setting.

Overview of the Accountability Movement

The publication of *A Nation at Risk* report called for frequent testing, particularly at the high school level and advocated that testing was needed to improve the public educational system (Hogan, 1983). According to the report, higher test scores would produce better workers and a better economy (Savage, 2003). Consequently, this report along with other publications and the political climate which favored an increase of testing in public schools became the main focus of educational reform in the 1980s (Vogler & Virtue, 2007). However, by the late 1980s and 1990s, the public called for more than a competency test of high school students; this led to a new era in public education known as the standards movement which held schools accountable for students' test scores (Vogler & Virtue). Consequently, in 2002, NCLB was enacted which required schools to be held accountable for academic performance (U.S. Department of Education, 2010).

The enactment of NCLB requires states to administer math and reading tests in grades three through eight and once during high school (Lane, 2004; O'Neill, 2003). Under the law, states must delineate how they plan to close achievement gaps and ensure all students achieve academic success (U.S. Department of Education, 2010). In addition, states are required to make AYP to make certain schools are held accountable for student achievement of tests mandated by the state (Orlich, 2004; U.S. Department of Education).

Second, NCLB requires states to determine AYP for schools and school districts within the guidelines set by Title I (U.S. Department of Education, 2010). Each state determines the minimum level of improvement, measured in terms of student performance that schools and school districts must meet within a specified amount of time (U.S. Department of Education). Each state sets a starting point based either on the performance

of the lowest-achieving schools in the state or of the lowest-achieving demographic group, depending on which is higher (U.S. Department of Education).

Next, the state sets a level for student performance schools must achieve after two years to continue to show AYP (U.S. Department of Education, 2010). Performance levels must increase once every 3 years. By the year 2014, all students must perform at proficient levels on the state-mandated test in math, language arts, and reading (U.S. Department of Education).

However, if schools do not make AYP after 5 years, modifications in which the way the school is operated will be made (Orlich, 2004; U.S. Department of Education, 2010). According to the U.S. Department of Education, the following consequences will occur for failing schools:

1. A school that does not make AYP for 2 consecutive years will be identified as needing improvement. A 2-year plan will be developed by school officials. The local education agency is required to ensure the school receives necessary assistance to develop and implement the improvement plan. Students will be given the option to transfer to another public school within the district that is making AYP.

2. Schools that do not make AYP for three consecutive years will remain in the school-improvement status. The school district will still be required to offer public school choice to students. Additionally, students from low-socioeconomic backgrounds will be eligible for additional services such as remedial classes or tutoring services from a state-approved agency.

3. Schools that do not make AYP for four consecutive years will continue to offer public school choice to low-income students, and schools will receive corrective actions

by the school district such as replacing certain faculty members or changing the curriculum.

4. Schools that do not make AYP for 5 years will undergo restructuring by the school district. Restructuring may include replacing the faculty, turning the school into a charter school, or turning the school over to the state or to a company that has demonstrated effectiveness.

School officials and students from numerous states have concerns about the testing requirements outlined in NCLB (Zimmerman & Dibenedetto, 2008). One specific concern is that by the end of the 2013-2014 school year, all students are expected to score either at or above grade level, and if this does not occur, sanctions may be implemented (Zimmerman & Dibenedetto). One of the debates about the implementation of NCLB is that the results of the tests required by NCLB do not show growth of individual students from year to year. For example, decisions about whether a school makes AYP are determined by whether third graders at a school performed better than third graders from the cohort group from a previous year (Zimmerman & Dibenedetto). Consequently, concerns about this model have been raised because it does not give credit to schools for academic improvements if they do not meet AYP goals (Zimmerman & Dibenedetto). A change in the demographics of the students from different socioeconomic backgrounds can skew the scores which make the use of standardized tests an inappropriate tool to measure the effectiveness of schools (Savage, 2003; Zimmerman & Dibenedetto). Savage contended high-stakes tests are detrimental to accountability and to improvement because these tests can provide inaccurate data that can distort educational practices and give the public inaccurate information regarding how effectively schools are performing. In agreement

with Savage's statement, Altshuler and Schmautz (2006) believe the purpose of standardized tests is to analyze students' strengths and weaknesses as opposed to using it for accountability purposes. Using test scores to compare schools, to determine a school's future, to measure the effectiveness of teachers, and to determine a student's future with regard to promotion to the next grade level or whether a high school diploma will be received, "is deep, direct, and personal" (Fukuda, 2007, p. 431).

Financial Rewards and Consequences for High-Stakes Test Scores

NCLB requires states to provide achievement awards to schools that are successful at closing the achievement gaps (U.S. Department of Education, 2010). Specifically, some schools are awarded monetary rewards (Parke, Lane, & Stone, 2006). States must also recognize schools that have made the greatest amount of gains in closing the achievement gap (U.S. Department of Education). Additionally, Title I funds may be used by states to financially reward teachers who serve in schools that earn academic achievement awards (U.S. Department of Education).

For example, Golden Apple awards are the financial rewards for schools that perform well on high-stakes tests (Paris, 2000). Georgia, Michigan, California and other states have created policies to reward schools that perform well on high-stakes tests (Paris). Paris suggested when teachers are paid to increase test scores or are threatened with consequences if scores decrease, they will consequently engage in teaching methods that are geared only toward raising test scores.

Moreover, opponents of merit pay believe when teachers are paid to increase test scores, urban, rural, and poor schools who serve large numbers of low-socioeconomic

students and minority students are at a disadvantage because they are at risk for low test scores (Paris, 2000). Some state departments are cognizant of the problems associated with merit pay. Therefore, state departments have rewarded schools for increasing test scores each year. However, educators are frustrated with this practice as well because some cannot improve test scores yearly. In addition, merit pay may encourage teachers to work only in schools with students of high-socioeconomic backgrounds rather than work with at-risk students, or they may choose to change to grade levels not impacted by test scores (Paris).

Financial incentives will not necessarily result in improved teaching but will likely result in changing what is taught and who is taught (Savage, 2003). Savage believes teachers will focus on the place where the least amount of work will yield the greatest amount of return resulting in teachers spending the majority of their time teaching students in the middle and leaving out students at the top and at the bottom. Additionally, teachers who are interested in high evaluations and financial incentives will choose to teach in schools with high-socioeconomic students (Savage). Furthermore, when teachers receive monetary rewards to improve students' test scores; they will feel pressure which may shift their focus on increasing students' test scores as opposed to increasing learning (Paris & Urdan, 2000).

Unethical Practices

As the consequences for test scores increase, such as the consequences outlined by NCLB, the temptation for cheating may increase as well (Paris, 2000). According to Paris and Urdan (2000) participation in unethical test practices is a critical issue. Unethi-

cal practices include using test items from tests administered during previous years to prepare students for the test, and pointing out or erasing incorrect answers during the administration of high-stakes tests (Paris & Urdan). In addition, it has also been reported that students who score low on high-stakes tests were expelled (Paris & McEvoy, 2000). There have also been reports that answers have been changed or marked by staff members (Greene, Forster, & Winters, 2005; Paris & McEvoy). Furthermore, some schools have failed students or exempted other students from taking the test (Paris).

Numerous reports of unethical test preparation have been mentioned in most states that administer high-stakes testing (Jones et al., 2003). For instance, in Columbus, Ohio a principal questioned a teacher why her students' scores were low because the students did not perform as well compared to previous years. After the teacher questioned the students about the drop in their scores, they mentioned to her that in previous years the answers were given to them and this year they were not (Mathews & Argetsinger, 2000). Another example of unethical test practice occurred in South Carolina when a teacher was terminated from her position for violating test security by telling the students test questions and answers prior to the administration of the test (Canner, 1992). Moreover, in Virginia parents protested when students were guided toward correct answers on the test and received help in revising answers to essay questions (Mathews & Argetsinger). Furthermore, during the 1999-2000 school year in the New York city public school system, 52 school officials were charged with cheating on a high-stakes test (Popham, 2001). Additionally, in 2001, a Maryland school principal resigned after an allegation that fifth grade students were guided to marking the correct answer and received help rewriting responses on an essay (Popham). There have also been reports of certain students

being told that they could stay at home on testing day (Barksdale-Ladd & Thomas, 2000). In addition to the states previously mentioned, there have been many reports of cheating in states such as Texas and Florida (Greene et al., 2005). Consequently, school officials who engaged in breaking the rules during tests have had their teaching certificates revoked or have lost their teaching jobs. These reports of unethical test practices support Paris' belief that high-stakes testing used for accountability and for political reasons can lead to unethical practices and can have negative impacts on education.

Reports of cheating on high-stakes tests served as a catalyst to the creation of computer programs designed to count the number of erasures on exams (Jones et al., 2003). The number of erasure marks on tests for a particular homeroom class or a school is compared to the number of erasure marks for another homeroom class or a school. If there are too many erasures, it is a signal that someone has breached test security (Jones et al.). Schools are also taking precautions against unethical test practices such as allowing monitors to be present during test administration, carefully monitoring the distribution of test booklets, and not allowing teachers to see the items on the test until students receive them (Jones et al.). However, there are still concerns regarding what is ethical, such as whether it is acceptable for a teacher to redirect a student who is marking more than one answer or to redirect a student who is filling in the answers to the questions in the wrong area (Jones et al.).

Perceptions of High-Stakes Testing from Opponents and Proponents

The majority of the public as well as some teachers support high-stakes testing. However, teachers are concerned about the impact high-stakes testing has on teaching

and student learning (Center for Public Education, 2006). Critics believe high-stakes testing will lead teachers to change instruction from lifelong learning to teaching to the test, which can affect the validity of the test (Smyth, 2008). Specifically, critics believe test scores can possibly provide inaccurate measures of student achievement (Shepard, 2002). A test must measure what it intends to measure for it to be valid (Marchant, 2004). Marchant believes if judgments will be made concerning the quality of teachers, school systems, or states, the results of the given test must be valid. There may be excellent teachers who teach valuable skills to students, but if the skills they teach are not assessed on the test, students' scores will not increase (Paris & Urdan, 2000). This may occur when the curriculum and tests are not aligned (Paris & Urdan). Paris (2000) mentioned the overlap between a test and the curriculum is a critical component of validity of the test being used. "If a standardized achievement test is to be a valid measure of student learning, the quality of instruction of a teacher, or the effectiveness of the educational system of a school, district, or state, that test must match the curriculum being taught" (Marchant, 2004, p. 5).

According to Marchant (2004), selection bias is a threat to validity. He states that "Selection bias may occur when the samples being compared are not randomly assigned to different treatment groups, such that the sample in the treatment groups are qualitatively different from each other in ways that can impact the results" (p. 5). Marchant mentioned that students are not randomly assigned to school systems, to schools, and to teachers. Students may be overrepresented or underrepresented in samples. Hence, there are differences among students which do not relate to the instructional quality provided by schools (Marchant). In agreement with Marchant, Horn (2003) argued test scores con-

sist of important information, but these scores do not give us all of the information needed to make critical decisions. During daily instruction, students are normally encouraged to use dictionaries, check answers, reread, and seek help which is not permissible during the administration of high-stakes tests (Paris, 2000). Additionally, high-stakes tests require students to read short passages and answer multiple-choice questions which are inconsistent to a balanced literacy approach (Paris). Hence, Paris contended because high-stakes tests are inconsistent with students' normal instruction, they are unfair and not authentic. This type of test undermines learning and teaching because they are inappropriate for assessing certain activities students' experience (Grant, 2004; Popham, 2001). Students who are taught in engaging schools create projects, answer real-world problems, apply what they have learned to new situations, and engage in synthesis and evaluation methods (Grant). Instruction such as this is not very well assessed by multiple-choice tests (Grant).

Sloan and Kelley (2003) argued high-stakes tests are administered during the latter part of the school year. Hence, teachers rarely receive guidance on how to use test results to enhance instruction (Zimmerman & Dibenedetto, 2008). Only seldom do these tests provide diagnostic information for students and teachers in a timely manner (Sloane & Kelly, 2003). Sloane and Kelly argued the purpose of the test is to determine what students have learned or can do. However, the test provides little or no information to increase student achievement (Marchant, 2004; Sloane & Kelly). Sloane and Kelly proposed the purpose of formative assessment should be to help students learn rather than compare students to each other.

In contrast to Sloane and Kelly's (2004) statement that high-stakes tests provide little or no information to increase student achievement, the findings of the study conducted by Greene et al. (2004) suggest an opposing view to the importance of high-stakes testing. The study conducted by Greene et al. examined if standardized test results are distorted when high-stakes are attached to the exam. The researchers located states and school systems that administer high-stakes and low-stakes tests. The data from test results from seven school districts in Florida and Virginia were used. Moreover, the data from test results in the following school systems were used: Chicago, Illinois; Boston, Massachusetts; Toledo, Ohio; Fairfield, Ohio; Blue Valley, Kansas; Columbia, Missouri; and Fountain Fort Carson, Colorado. Because averages had to be computed and because different schools reported results in different ways, the test scores were standardized into *z*-scores to make comparisons between the test scores possible. A correlation was computed between high and low-stakes test results for the score level and for the yearly gain in scores. The researchers found that the test scores in all of the states studied generally correlate between the high-stakes and low-stakes assessments. Greene et al. concluded that high-stakes tests do not misrepresent the information about the academic performance level of students. The researchers proposed that if high-stakes testing is used only for the purpose of determining whether or not students perform at a certain academic level, the high-stakes tests seem to be a reliable tool. Additionally, Greene and colleagues suggested Florida's high-stakes testing program provides credible information about student achievement regardless of the high stakes placed on schools.

On the opposite side of the debate, proponents, specifically a majority of the American public, believe students will be motivated to work harder when rewards and

sanctions are attached to tests (Kornhaber & Orfield, 2001). Americans also believe the tests administered to students are fair and are accurate at determining whether students should be promoted or should receive a high school diploma (Fuduka, 2007; Gardner, 2002; Kornhaber & Orfield). Those in favor of high-stakes testing argue that the tests are free from bias and are useful in depicting student achievement (Gardner). Americans also think high-stakes testing should be used to hold schools accountable for student learning (Kornhaber & Orfield). In addition to proponents' views that high-stakes testing are useful, those in favor of high-stakes testing suggest tests will make parents and the public cognizant of the performance of the students and of the school system (Kornhaber & Orfield). Proponents of high-stakes testing also believe testing mandated by the government allows control over local school systems, and the testing program serves as a means to ensure educational agencies are effectively doing their job (Egley & Jones, 2004). High-stakes testing is viewed by many proponents as a way to motivate students to pay attention and devote more effort at meeting the demands of the school system (Egley & Jones; Kornhaber & Orfield). Additionally, proponents believe the tests can be used as a tool to improve classroom practices (Gardner).

In alignment with the beliefs of proponents of high-stakes testing, Cizek (2001) believes high-stakes testing has led to an increase of student learning despite critics' arguments that an increase in test scores may be a result of high-stakes tests but not necessarily a reflection of student learning. Cizek believes increased student achievement is a result of the implementation of high-stakes testing and that over time high-stakes tests have become "(a) highly reliable, (b) free from bias, (c) relevant and age appropriate, (d) higher order, (e) tightly related to important, public goals, (f) time and cost efficient, and

(g) yielding remarkably consistent decisions" (Cizek, 2001, p. 25). Although Cizek has positive beliefs about testing, a Rand Report determined that it is not evident if test scores are a reflection of improvements in student learning or a reflection of score inflation which may be caused by narrow test preparation (Stecher & Hamilton, 2002). This narrowing of the curriculum through the use of test preparation is discussed in more detailed in the following sections.

Impact of High-Stakes Testing on the Curriculum

There is very little debate that tests may impact the curriculum (Lipman, 2004; Madaus & Clarke, 2001; McNeil, 2000; Smyth, 2008; Vogler, 2005, 2006). Smyth believes high-stakes testing can narrow the curriculum by "teaching to the test" which can stifle teacher creativity, innovative instruction, and the use of a variety of teaching strategies for diverse students. Moreover, teaching to the test will result in teachers avoiding teaching higher-order thinking skills (Darling-Hammond, 2004). Additionally, opponents of high-stakes testing believe the pressure of accountability will lead teachers to narrow the curriculum by engaging students in a test preparation curriculum (Grant, 2004; Lane & Clement, 2002; Marchant, 2004). Marchant asserted teachers will teach only content that is tested and will avoid content that is not tested. Gulek (2003) also believes when high-stakes are associated with test results, teachers may be more likely to emphasize the objectives that will be covered on the test. Consequently, Grant believes test preparation will make teaching boring and unmindful of real-world problems and issues. Contrary to current ongoing teaching practices, Gulek believes educators should not focus only on instruction that is limited to the objectives of the test. Focusing solely on test objectives

may increase test scores; however, this narrow focus may not address issues in the curriculum necessary for preparing students for the world of work beyond school (Gulek).

Various studies have been conducted to determine the impact of testing on the curriculum. For instance, Au (2007) analyzed 49 qualitative studies which examined the influence of testing on the curriculum. Meta-synthesis was used to synthesize a group of qualitative studies. The findings of the meta-synthesis revealed testing has a narrowing effect on the curriculum influencing teachers to use more lecture-based, teacher-centered instruction. However, other findings suggest high-stakes testing caused an expansion of the curriculum and influenced student-centered instruction.

In another study, Costigan (2002) interviewed six first-year teachers of grades three through five who taught in New York City. The teachers were interviewed at the end of their first semester of teaching and a second time in the middle of their second semester of teaching. Costigan's findings suggest testing became the main concern among teachers, and they viewed testing had negative impacts on students, the curriculum, and classroom teaching practices. The teachers in this study also noted they were unprepared to deal with the volume of testing, and they also noted a loss of power for teachers. Costigan also reported these teachers were not able to balance between a testing curriculum and best practice and that there is a need for deeper dialogue between teachers and how to teach in a testing environment.

Similarly, Crocco and Costigan (2007) interviewed teachers in New York City. Crocco and Costigan's study focused on reporting narratives about teachers' work. During a 5-year period, the researchers conducted over 200 interviews with novice teachers who all had no more than 5 years of experience. Additionally, interviews with focus

groups were conducted. Crocco and Costigan concluded as a result of the accountability movement in New York, teachers reported testing narrowed the curriculum, and they perceived they have lost control over the teaching practice. As a consequence of losing control over the profession, teachers reported this was a driving force for leaving the teaching profession.

Comparatively, a study conducted by Lipman (2004) corroborated the beliefs of opponents' concerns regarding the narrowing of the curriculum. Lipman conducted a qualitative study in which she observed teachers from four elementary schools in Chicago. Based on data from observations, Lipman reported test preparation was used with students such as engaging in practice tests, filling in bubbles on scantron sheets, becoming familiar with the format of the tests and types of questions posed on the tests, and learning test strategies for eliminating incorrect answers. Lipman also observed that in one school, teachers ceased from teaching social studies and focused on reading and math the second semester. Lipman argued higher-socioeconomic students in Chicago received a more challenging curriculum while low-socioeconomic and minority students were engaged in memorization of facts and test-taking techniques.

In the elementary school setting, narrowing of the curriculum has resulted in the reduction or dismissal of subjects that are not tested (Vogler & Virtue, 2007). Some researchers (Grant, 2004; Lipman, 2004; Rock, 2006; VanFossen, 2005) contended the focus on reading, language arts, math, and science by NCLB has led to a reduction of social studies in numerous classrooms. Numerous elementary teachers do not teach social studies because it is not tested (Rock; VanFossen). Few states require testing in the area of social studies, but because few states test this subject, it is perceived to be less important

than other subjects that are tested, and it receives far less funding than subjects such as English, math, and science (Vogler & Virtue). Due to the decreased emphasis on social studies, Vogler (2006) believes social studies teachers have difficulty teaching beyond just factual information and covering higher-level, critical thinking due to limited time. Some researchers believe teachers will increase their use of teacher-centered practices such as lectures and reliance on textbooks and will move away from student-centered approaches such as role plays and cooperative learning and teach only factual knowledge (Gayler, 2005; Jones et al., 2003; Vogler; Vogler & Virtue). On the opposite side of the debate, Gerwin (2004) believes high-stakes testing has not impacted teaching practices. Grant contended social studies have always mostly been comprised of recitations and regurgitation of facts, thereby not having an impact on instruction.

Stakeholders have also become concerned about the future of science in the classroom, the many efforts to improve the science curriculum, the teaching of science, and student learning in this subject (Goldston, 2005). The efforts to provide an inquiry-based meaningful curriculum are in contrast to the emphasis being placed on reading, writing, and math. Consequently, as with social studies, time for science is being reduced in the elementary grades or is not included in the curriculum at all (Goldston; Grant, 2004). School officials are omitting science instruction while allowing literacy to dominate the curriculum resulting in a curriculum that is “fractured, unconnected to context, and out of balance” (Goldston, p. 186).

High-stakes testing is also having an impact on math. Cankoy and Tut (2005) conducted a study to determine if the effects of an instructional approach which focused on a high-stakes standardized test affected mathematical performance of 1,006 fourth

grade students in 28 North Cyprus schools. A total of 28 preservice elementary teachers were trained to use observation sheets to code teachers' instructional activities in the fourth grade classrooms. After the analyses of data via ANOVA, multivariate of analysis of variance (MANOVA), and chi-square procedures, the researchers determined students who spent more time on test-taking strategies performed better in routine mathematics items than the students who spent less time on test-taking strategies. However, those who spent more time on test-taking strategies did not perform better on math items that required critical thinking.

Teachers report the pressure of testing is impacting their instruction in reading as well (Hoffman, Assaf, & Paris, 2001; Pennington, 2004). For example, teachers are using low-level, skill-based instruction as opposed to integrated, meaningful instruction to increase students' test scores (Pennington). Instead of focusing on literacy needs, they are focusing more on the objectives that will be tested due to concerns about ensuring students pass standardized tests (Flores & Clark, 2003). The study conducted by Assaf (2006) supports this belief. Assaf conducted a qualitative ethnographic study and observed a teacher in the state of Texas during reading instruction. Marsha, the teacher, taught in a school that served a large Hispanic community. Students in Texas take the TAAS, which is a high-stakes test. When the scores on the TAAS dropped at her school, Assaf observed the teacher changing her reading instruction from rich and authentic experiences to test-focused instruction focusing on mastery of low-level test skills which aligns with Lipman's study that reported that high-stakes tests affect teachers' instruction (Assaf). Additionally, McNeil and Valenzuela (2001) have investigated the TAAS, and triangulated data from numerous sources over a period of multiple years. Analysis of the

data suggested the implementation of the TAAS changed what teachers taught and how students learned. For instance, during reading, writing, and math teachers placed more emphasis on test preparation activities than on engaging students in an intellectually challenging curriculum.

To address the concerns regarding the narrowing of the curriculum, Wolf (2007) posited to ensure the curriculum is not narrowed and that the amount of material that will be tested should be expanded. Based on a study conducted by Roediger and Karpicke, Wolf believes students who know they will be tested on a particular concept are more likely to commit it to long-term memory due to preparation or the reinforcement nature of the process of testing. Wolf contended when students participate in standardized tests, they learn to listen and follow directions carefully. They learn they will be tested in life and prepare themselves to do their best. In addition, they learn skills to focus on, and face a challenge. Hence, Wolf views testing as a positive influence on student learning which is in contrast to the overwhelming amount of negative views of researchers (Assaf, 2006; Cankoy & Tut, 2005; Goldston, 2005; Grant, 2004; Hoffman et al., 2001; Lipman, 2004; McNeil & Valenzuela, 2001; Pennington, 2004).

Impact of High-Stakes Testing on Teaching Practices

Some teachers and researchers report that high-stakes testing has negatively impacted teaching practices (Cankoy & Tut, 2005; Jones et al., 2003; Paris & Urdan, 2000). Critics believe high-stakes testing decreases teacher creativity and the enjoyment students get from learning (Gardner, 2002). Some believe high-stakes tests focus on lower-level skills and knowledge rather than higher-level skills (Jones et al.; Paris, 2000). Additional-

ly, opponents of high-stakes testing claim teachers are forced to focus mainly on teacher-centered instruction as opposed to student-centered instruction (Jones et al.).

For instance, Hoffman and Paris (2001) conducted a study to determine teachers' perceptions of the impact of the TAAS on teaching practices. TAAS is a criterion-referenced assessment that focuses mainly on reading, writing, and math. Students in grades three through eight take the assessment each year in the spring. Teachers were asked to respond to a survey that consisted of 113 questions regarding participants' demographic information, the attitudes of participants regarding their attitudes about the TAAS, the administration and test preparation practices, how scores are used, the impact of TAAS on students, and the overall perceptions about TAAS testing. The responses of the 200 teachers who returned the survey were entered in a data file for analyses. Composite scores were computed after combining items from the different sections on the questionnaire. The composite scores were reported using means as well as standard deviations. The individual items were reported using percentages and categories of responses. Finally, the qualitative data from comments on the last section of the survey were analyzed by developing themes among the responses.

The results from the teachers who administer the TAAS reported they plan the objectives they will teach for the year around the areas that students will be assessed on TAAS (Hoffman et al., 2001). Although reformers considered teachers' planning of objectives around the TAAS as a positive consequence, teachers regarded this outcome as a negative impact on teaching practices. Moreover, teachers questioned the validity of the test specifically for ESL students and minority students. These students comprise the majority of students served in the Texas public school system. Teachers also reported stu-

dents were aggressive, irritable, or anxious during testing, and some students developed stomachaches and headaches while taking the TAAS. Half of the teachers who responded to the survey did not believe that the increased TAAS scores were the results of increased student learning. However, teachers believed the increased test scores were a result of teaching to the test. In addition, 85% of the respondents believed some of the best teachers are leaving the teaching profession because of the TAAS. Finally, some teachers expressed that they changed grade levels or taught in a specialization area to escape from the pressure of TAAS. The study regarding teacher perceptions of TAAS confirmed the negative views teachers have about testing in the state of Texas and also confirmed the beliefs of critics.

In a study conducted by Pringle and Martin (2005), the impact of high-stakes testing on elementary science teachers in Florida was examined. The researchers surveyed 38 teachers using open-ended and closed-ended questions. Item by item analysis was conducted for questions that pertained to teachers' background information, and for inductive analysis was conducted for open-ended questions. The findings suggested teachers were concerned about finding test preparation materials, the unfamiliarity with the test, the focal shift of the curriculum, the success of the students on the test, and the amount of time needed to prepare students for the test. The researchers noted that many of the teachers' concerns were a result of their lack of knowledge with science and science standards.

Another group of researchers examined the impact of the MSPAP and the MLO including principal, teacher, and student beliefs, classroom teaching practices, and student learning in writing and in reading (Parke et al., 2006). A sample of 59 elementary

and 31 middle schools in the state of Maryland participated. A total of 86 principals, 505 reading and writing teachers in grades two, three, four, five, seven, and eight, and 5,047 students in grades four, five, seven, and eight completed questionnaires. Additionally, 44 of the 90 schools collected classroom instruction and assessment materials. Mean scores, confirmatory factor analyses, analysis of variance, and growth model analyses were used to analyze the data from the questionnaires. Classroom instruction and test materials were collected from participating teachers, and student performance on the MSPAP within the past five years was obtained. The results of the study indicated principals and teachers were supportive of the MSPAP, and performance gains in the areas of reading and writing occurred in schools that used reform-oriented instruction. The positive effects of the MSPAP may be attributed to the fact that the MSPAP is performance based and teachers were included in the development of the assessment. The test was not new to school officials, was reliable and valid, and high stakes for the school were present but not for students. These factors reduced anxiety for students and teachers.

The findings of the study regarding the MSPAP (Parke et al., 2006) aligned with Cizek's (2001) belief that high-stakes testing leads to an increased intimacy with the discipline of teaching. For example, Cizek believes once it has been decided that a test will be mandated, usually a panel is formed that consists primarily of educators who possess familiarity with the content of the objectives to be tested and ages of the children to be tested. The panel studies pertinent documentation such as legislation and curriculum guides. During the panel discussions, the content, assessment techniques, developmental issues as well as other relevant information are discussed. Consequently, educators increase their knowledge about the discipline of teaching.

Some researchers have identified positive effects of high-stakes testing on teaching practices (Cizek, 2001; Jones et al., 2003; Parke et al., 2006). Some teachers have reported using a more student-centered approach rather than a teacher-centered approach to teaching (Jones et al.). Others report they are implementing more teacher-centered approaches, and they believe this is an appropriate teaching method which has improved the quality of their teaching (Jones et al.). Jones et al. contended the type of high-stakes test may influence instructional practices. Particularly essays, portfolios, and tests that show students' work with open-ended questions frequently influence instructional practices in a positive, student-centered way than multiple-choice questions focusing on skills and factual knowledge.

Impact of High-Stakes Testing on Students

Although the American public believes high-stakes tests will motivate students to learn, opponents believe these tests do not encourage students to work harder (Kornhaber & Orfield, 2001). Horn (2003) argued high-stakes testing can have a potentially negative impact on students. For example, Paris (2000) believes performance goals will lead to some students being reluctant to participate or try their best on assessments because they fret that their test results will consequently lead to them being labeled as having low ability. Sloane and Kelly (2003) stated labeling a child as low ability can negatively affect his confidence in his ability to learn. Paris asserted having performance goals will lead to some students experiencing anxiety. Moreover, students who spend an excessive amount of time on high-stakes test may possibly have problems applying the content they are learning to real-life situations (Cankoy & Tut, 2005).

Many believe standardized testing increases anxiety in the education community, and negatively impacts student achievement (Mulvenon et al., 2005; Paris, 2000). Opponents of high-stakes testing also believe teachers across the country are under different amounts of stress (Jones et al.). The amount of stress for students to perform well on high-stakes tests has a positive impact on some teachers but a negative impact on others (Jones et al.). There are some teachers in schools with high-stakes tests who feel they are working under an undesirable work environment as a result of the stress of high-stakes tests (Jones et al.). Evidence suggests when teachers feel pressure to increase test scores, they in turn put pressure on students to increase achievement (Paris). Pressure from teachers to increase test scores may increase student anxiety about scores which may lead to a decline in student motivation and a decline in respect for teachers (Paris).

An example of students experiencing anxiety is documented in a study conducted by Tripplett and Barksdale (2005). Tripplett and Barksdale studied elementary students' perceptions of high-stakes testing via drawings and written responses to questions posed by the researchers. Drawings and writings were collected from 225 third through sixth graders. The students were from five schools, which consisted of a diverse group of students from different races and socioeconomic statuses. Moreover, half of the students included in the study were from a rural community in a mid-Atlantic state, and the other half of the students were from an urban community in a southern state. The day after high-stakes testing, the students drew pictures and wrote about the high-stakes testing experience. The researchers categorized the drawings. The most common category was emotions. The most frequent emotions were "nervous" and "angry." Many of the drawings had sad or angry facial expressions. Drawings with smiles did not exist, and teachers

were generally excluded from the drawings. The findings suggested there was an overwhelming amount of negativity which supported previous research reports of students experiencing anxiety during high-stakes testing.

In contrast to the study conducted by Tripplett and Barksdale (2005), the study conducted by Mulvenon et al. (2005) disaffirms these beliefs about students experiencing anxiety. Mulvenon et al. assessed teacher, principal, counselor, student, and parent perceptions about standardized testing. The following participants responded to the survey: 251 fifth grade students and their parents, 8 counselors, 141 teachers, and 7 principals returned surveys. Multiple regression analyses and analysis of variance were used to analyze the data from the questionnaires. The findings suggest the concerns regarding standardized testing are mainly misrepresented because most principals, counselors, parents, and students valued standardized testing and did not perceive an increase in stress or anxiety due to testing. In contrast, teachers had strong concerns about standardized testing and reported the greatest amount of anxiety due to testing.

Another impact of high-stakes testing on students is the fact that certain states are now using high-stakes tests to make promotion decisions (Greene et al., 2004; Johnson & Johnson, 2006; Marchant, 2004). Advocates for retention argue employing negative consequences will encourage students to work harder and will encourage teachers to focus more on the needs of low-performing students (Hong & Youngs, 2008). Roderick and Nagaoka (2005) argued if students have not mastered certain skills, it would be more beneficial for the students to repeat the same grade. In alignment with this belief, a number of school districts are using tests to decide whether students will be promoted or retained (Gardner, 2002; Kornhaber & Orfield, 2001; Marchant). In Baltimore, more than

20,000 elementary and middle school students were retained in their current grade level when they did not meet the requirement of the Terra Nova national achievement test (Marchant, 2004). In Florida, the FCAT is administered. Third grade students are required to pass the exam before being promoted to fourth grade or before being eligible for graduation (Greene et al., 2004). In Chicago, students in grades three, six, and eight are required to pass the ITBS before being promoted to the next grade level (Greene et al.). Additionally, the results of the MCAS, which is administered in Boston to students in grades three, four, five, six, seven, eight, and ten is also used to determine whether students are promoted or are eligible for graduation (Greene et al; Horn 2003).

Some researchers believe high-stakes testing will negatively impact students of low-socioeconomic status and students of color (Beers, 2005; Gollnick & Chin, 2002; Smyth, 2008). Savage (2003) contended although standardized tests are well-written tests constructed by experts, they are designed to spread students along a continuum. High-stakes tests generally are multiple-choice assessments in which students fill in a bubble or a square (Grant, 2004). Hence, test items that every student is able to answer will not spread out students' scores. Standardized tests are administered once each school year and focus on content correlated to socioeconomic status and provide inaccurate data for holding schools and students accountable (Savage). Some researchers contend standardized testing has biases related to socioeconomic status (Gollnick & Chinn; Smyth). Students from affluent families, schools, and/or school districts can afford tutorial resources or expensive materials to prepare for high-stakes tests (Smyth). However, lower-performing schools do not have funds to purchase such materials which leave minority and low socioeconomic students behind (Smyth). Beers contended students of poverty

most likely attend schools that lack resources such as equipment, textbooks, Internet access, and highly qualified teachers. Also, Hursh (2005) argued AYP often discriminates against schools serving students living in poverty and students of color. Because there is a correlation among a student's family income and test scores, a school's score is more likely to reflect its students' average family income as opposed to reflecting teaching or the curriculum (Hursh). Because affluent students will be more likely to rely on cultural capital to pass exams, the students who are disadvantaged will receive additional drilling of specific skills which will cause these students to fall further behind because of lower expectations (Hursh). The National Center for Fair and Open Testing (2007) argued a great percentage of the best teachers will leave low-performing schools to go to higher-performing schools, which leaves behind the students with the greatest needs.

Some believe special education students will be positively impacted by high-stakes testing whereas others believe these students will be negatively impacted (Cizek, 2001; Huefner, 2006; Smyth, 2008; Thompson & Thurlow, 2001). Cizek argued school officials are increasing their attention to students with special needs due to federal mandates that require high-stakes testing. Students with special needs have also reported that due to accountability testing, teachers increased their attention toward them (Cizek). Cizek believes the attention given to special needs students is one of the many positive consequences of high-stakes testing on classroom teaching practices. Similarly, Thompson and Thurlow reported when students with disabilities are administered high-stakes tests, it gives them the opportunity to obtain the regular high school diploma, parents are more knowledgeable of the standards and tests, and special education teachers are more involved in instruction that helps students to meet the standards. In contrast, some oppo-

nents of high-stakes testing are concerned about special education students. Jameson and Huefner (2006) argued it is almost impossible for schools with special-needs populations to comply with the mandates of NCLB because NCLB does not adequately fund the demand for highly qualified teachers. NCLB allows schools to exclude two percent of the student population from the state assessment; however, magnet and charter schools specialize in educating students with various exceptionalities (Smyth). Consequently, 100 percent of the school's population has special needs thereby making it irrelevant to report 98 percent of the school population's test results (Smyth). In cases such as these, AYP as mandated under NCLB will never be met by these schools leaving behind students with special needs.

Research suggests when students with disabilities do not perform well on high-stakes tests, they receive more drill and practice of basic skills (Jones et al., 2003). Yet, they need programs of high quality as opposed to poor instruction involving drill and practice (Jones et al.). Jones et al. noted there are also negative impacts of students with disabilities taking high-stakes tests which include administering tests that are too difficult for these students increasing students' feelings of being overwhelmed, increasing paperwork, as well as presenting the risk of students with disabilities not being able to graduate because of an inability to meet standards.

Critics argue high-stakes tests are too difficult for students with special needs, but a recent study pertaining to accommodations may assist these students in increasing their achievement on high-stakes tests. In 2006, a study was conducted to determine if valid test accommodations benefit only the students with disabilities (Fletcher et al., 2006). Six suburban school systems in southeastern Texas were recruited for this study. In these

school systems, 48 schools and 113 teachers participated. A total of 91 third grade students with poor decoding skills were compared to 91 third grade students with average decoding skills. Third grade students who had difficulty with word decoding and were also identified as dyslexic and average readers from the same classroom as the students with dyslexia were randomly assigned to take the exact version of the Texas reading accountability assessment under standard and accommodated administrations of the test. Using a mixed model in which the analyses of the data were run using fixed effects and random effects, the researchers determined only students with decoding problems benefited from the accommodations which showed an increase in average performance. The results of the study demonstrated accommodations used for a specific disability can enhance performance on a high-stakes test.

One of the goals for NCLB is to decrease gaps in academic achievement on test scores between minority students and Caucasians and also between middle-class students and low-socioeconomic students (Hong & Youngs, 2008). Politicians desiring to increase the educational performance of low socioeconomic and minority students in the upper grades believe the stakes should be raised in the early grades in hopes of solving problems early on to reduce the drop-out problems in the future (Newman & Chin, 2003). Politicians believe standardized testing will ensure that low-socioeconomic students and minorities will receive a quality education (Assaf, 2006; Savage, 2003). Proponents of high-stakes testing believe NCLB mandates will enhance the education of minority and poor students, but opponents believe these practices will have a negative effect on their education (Menken, 2008).

The achievement gap on standardized tests among the races is a challenge that Americans and the educational systems are facing today (Fukuda, 2007). Student achievement for African American students has been a consistent problem in the United States (Haynes, 2008). In the past, minorities have not scored as well on high-stakes tests as Caucasians (Haynes; Jones et al., 2003). This may be attributable to the fact that minorities have generally not had as many resources and in the past were denied equal educational opportunities (Jones et al.). Horn (2003) suggested non-White students as well as special needs students are the groups mainly impacted by high-stakes tests.

Numerous teachers in the United States, specifically those working with students in low-income communities are feeling pressure to get students to pass standardized tests (Assaf, 2006). Grant (2004) argued socioeconomic background is highly correlated to the ability of students to perform well on high-stakes tests. Poor and minority students are less well prepared for school; hence, one would assume these children will be likely to have difficulty with high-stakes tests (Newman & Chin, 2003). Popham (2001) contended there are certain items on standardized tests that children of higher socioeconomic status are more likely to answer correctly than children of lower socioeconomic status because of greater accessibility to materials in the homes such as books, magazines, newspapers, and educational cable programs (Popham). Popham asserted test items linked to socioeconomic status spread out the test scores very well but are ineffective at evaluating the effectiveness of schools. Schools that serve low-socioeconomic students generally have less funding for education, and the students they serve entered school lacking experiences which lead to a successful school experience (Jones et al., 2003; Savage, 2003). Jones and colleagues posited when high poverty schools are under the pres-

sure of high-stakes testing, the students they serve are disadvantaged even more by an increase in test preparation activities. Moon, Callahan, and Tomlinson (2003) provide the belief that high-stakes testing negatively impacts students at low-socioeconomic schools are supported. Moon et al. conducted a study to determine if students' socioeconomic status affected teaching practices. A questionnaire was developed and sent to 8,044 elementary teachers across the United States. ANOVA and percentages of participants' responses were used to analyze the data. Teachers indicated they spent a significant amount of time preparing students for state-mandated tests, but teachers who taught low-socioeconomic students reported spending more time on test preparation strategies than teachers who taught high-socioeconomic students.

Arriaza (2004) believes when a community is rich, the school will include parents in active participation and decision making. However, in schools that serve low socioeconomic families, the parents who are not well educated, have low-literacy skills, and who do not speak English are not well prepared to provide assistance in learning to read or in math to their children (Newman & Chin, 2003). Consequently, the accountability reform in education has greatly challenged low-socioeconomic parents because they are expected to monitor their children's homework and reinforce the skills learned in school at home (Newman & Chin). For instance, the MacArthur Foundation Network on Successful Mid-Life Development sponsored a survey in which a total of 900 individuals were surveyed in New York (Newman & Chin). The researchers drew a sample of 100 families from the 900 individuals surveyed in New York for a qualitative sample to conduct an ethnographic study. From these 100 families, 12 families in New York City were the focus of a 6-year ethnographic study. The study conducted by Newman and Chin

found families transitioning from welfare to work at the time schools were implementing high-stakes tests. Parents of low socioeconomic students and of those who speak little English were unlikely to be able to provide academic instruction at home for their children. The parents of these children had to put the family's income before the children's education needs.

Another interesting study with regards to high-stakes testing and students of poverty is reported by Johnson and Johnson (2006). The researchers temporarily ceased from university teaching and taught third and fourth grade teachers at a low-socio economic school in Louisiana during the 2000-2001 school year. Louisiana became the first state to require fourth-grade elementary students and eighth-grade middle school students to pass a standardized test before being promoted to the next grade level. The authors wrote about their experiences in daily journals with the "voices" of the students, colleagues, school leaders, and politicians. As a result of their experiences as classroom teachers the authors reported major findings. The themes that emerged were effects of poverty on all aspects of life, and there were negative consequences of the demand for accountability in schools. Additionally, unreasonable demands were placed on teachers which stifled their creativity and enthusiasm which led to teachers leaving the teaching profession. Johnson and Johnson contended high-stakes testing negatively impacted teachers, students, particularly students of low-socioeconomic backgrounds, the curriculum, and the school.

High-stakes testing has also influenced ELLs. Tests are used to determine a student's level of knowledge and are also used to compare what a student knows to other students his or her age. The majority of the states in the United States are utilizing stan-

dardized test scores to prove students are progressing as mandated by NCLB (Menken, 2008). Tests of an ELLs knowledge in a specific area of content will likely be greatly impacted by the ELLs English language proficiency (Garcia & Menken, 2006). Specifically, there are language and cultural complexities of test items that may lead to errors when determining high-stakes decisions (Abedi & Dietal, 2004; Solano-Flores & Trumball, 2003). If a child's language is different from his peers then inaccurate conclusions can be drawn (Jones et al., 2003). For example, if a student does not understand English and obtains a low test score on a word problem in math, the score can be inaccurately interpreted concluding the student is below grade level in math (Jones et al.). ELLs are being included in the administration of these tests that were never intended for ELLs. However, states included these students in the same tests along with students whose native language is English (Menken, 2008). Students with limited English proficiency (LEP) face challenges meeting AYP. State assessments require high levels of English-language ability (Smyth, 2008). Consequently, many schools with a high percentage of LEP students cannot report AYP which in turn causes them to receive low test marks and lose state and federal funding leaving behind LEP students (Smyth, 2008). ELLs along with other groups of students who do not perform well on high-stakes tests are more at risk of dropping out of school (Jones et al.). Soon, NCLB will begin reauthorization (Menken, 2008). The Government Accountability Office (GAO) concluded states need additional support from the U. S. Department of Education to create reliable and valid tests for ELLs (Menken, 2008).

An example that ELLs are having difficulty with high-stakes testing is reported in a study conducted by Escamilla, Chavez, and Vigil (2005) which examined Colorado

teachers' perceptions of Spanish-speaking ELLs and Latino students' academic test results. The purpose of the study was to determine if there were achievement gaps "between Spanish-speaking ELL Latinos and other students at certain urban schools that were impacted by linguistic diversity" (Escamilla et al., 2005, p. 135) The researchers gathered descriptive data about the kindergarten through twelfth grade Spanish-speaking and Latino students in Colorado. Additionally, data from discussion groups with 35 teachers were analyzed by examining the emerging patterns from the discussion groups, and test data were gathered to address the purpose of the study. The findings from this study suggested teachers perceived Spanish speaking students were responsible for their low achievement scores on accountability tests. However, the test data showed Spanish-speaking Latino students in ELL programs who took the Spanish version of the state test met state standards. Consequently, the researchers challenged teachers' perceptions that Spanish-speaking Latino students were underachieving and noted the teachers in this study lacked evidence to support their claims. Many believe NCLB is needed to ensure students of all populations are receiving an adequate education, but Jones et al. (2003) believe high-stakes testing of students with low-socioeconomic backgrounds, minorities, students with special needs, and ELLs make these special populations at risk of failure.

Impact of High-Stakes Testing on School Geographical Locations

Hursh (2005) believes teachers, particularly in urban school districts, are under pressure to raise test scores which compels them to teach skills and knowledge that will be tested rather than more complex components of subjects. Hursh contended the pressure to raise test scores will force weak students out of school before taking the required

exam. Specifically, in Texas, students living in urban areas are more likely to be retained in school, especially in ninth grade which is the year before being required to take the TAAS. Also, with regards to problems in Texas Booher-Jennings (2005) collected qualitative data from an urban elementary school in Texas. At this school, teachers divided students into three groups: safe cases, suitable cases, and hopeless cases. The safe cases were the students likely to pass the test. The suitable cases were the students close to passing the test. The hopeless students were unlikely to pass the test. Consequently, resources and intervention were withdrawn from hopeless students which resulted in low socioeconomic and minority students losing opportunities to engage in higher-order thinking, analytical writing, and problem solving skills. Although the students in urban areas have the greatest need for support, students in urban schools are taught by teachers who are not well qualified and have little or no experience, while the teacher-to-student ratio in inner city classes is higher than in suburban and rural areas (Gollnick & Chinn, 2002).

Another qualitative study examined the teaching experience of a student teacher in a low-performing urban school. Lloyd (2007) observed and interviewed a student teacher during her ten week student teaching internship. The purpose of the study was to identify factors that influenced the student teacher's math instruction. The findings suggest the student teacher's teaching was greatly influenced by the mandates of the school that followed a test-centered curriculum consisting of worksheets and structured lessons. However, Lloyd concluded the student teacher made conscious decisions to plan lessons beyond school workbooks to allow students to be engaged in more activities which was in contrast to the school's teaching style and curriculum.

Because family incomes in rural areas are generally lower than in other school geographical locations, the challenges that low-socioeconomic students face are also significant for students in rural areas (Gollnick & Chinn, 2002). Specifically, in rural areas there are inadequate resources such as technology and advanced placement courses (Gollnick & Chinn). Additionally, Jimerson (2005) believes rural schools are disadvantaged by NCLB due to the school choice plan because if a student has to travel from a rural area to a school that is top-rated it could result in a long commute for students.

As previously mentioned, Berger's quantitative study regarding the relationship between teacher stress and high-stakes testing in the state of Virginia found rural teachers were more stressed than urban and suburban teachers. Similarly, one study documented rural area administrators' views about the impact of high-stakes testing in the school setting. Specifically, Egley and Jones (2004) conducted a study in Florida to determine how rural administrators were affected by the test compared to administrators in urban and suburban communities. A total of 325 administrators completed an online questionnaire, 42 of which were from rural administrators, 146 from suburban administrators, and 125 from urban administrators. The questionnaire consisted of 14 non-demographic information items, 11 likert-format type items, and 2 items requiring an answer of yes or no. Mean comparisons of the responses among rural, suburban, and urban administrators were calculated. The results of the study were as follows: rural elementary administrators spent a similar amount of time daily on instructional leadership as suburban and urban administrators. Rural administrators perceived the FCAT to have a more positive impact on their ability to increase teacher effectiveness than administrators in urban and suburban communities and found the FCAT more useful than urban administrators in aiding

them in assessing teachers' strengths and weaknesses in the areas of math, reading, and writing. Over half of the rural administrators believed FCAT had a positive impact on developmentally appropriate practices. A third reported the test had a negative impact, and a tenth of the respondents reported the test had no effect. Although more than half of rural administrators felt FCAT had a positive impact on developmentally appropriate practices, most of the administrators felt much pressure stemming from the FCAT and more so than urban administrators. However, most of the administrators noted the FCAT motivated the administrators to do a better job.

Moreover, Grant (2000) conducted two focus group interviews over a 2-year period. During the first year of the study, one focus group consisted of 7 elementary teachers and counselors, and the second focus group consisted of 12 high school teachers who taught in rural, urban, and suburban areas in New York State. During the second year of the study, five elementary teachers and eight high school teachers participated in focus group interviews. High school participants noted they felt pressure from their principals to maintain higher test scores. Although elementary teachers did not mention that they felt less pressure than their high school counterparts to increase students' test scores they did mention principals were "more likely to talk about test scores as part of a bigger picture of how students are progressing" (Grant, 2000, p. 16).

Schools in wealthy suburban schools are more likely to have qualified teachers, advanced-placement courses, numerous extracurricular activities, and adequate technology (Gollnick & Chinn, 2002). Although not all suburban schools have the resources to prepare students, they are more likely to have more educational resources than urban and

rural schools (Gollnick & Chinn). Hence, suburban students are more likely to have greater educational opportunities than rural and suburban students.

Testing in Alabama for Elementary Students

In the past, students in Alabama had to endure 2 weeks of testing (Alabama State Department of Education, 2010). However, beginning with the 2009-2010 school year, Alabama reduced testing from 2 weeks to 1 week. Students are now only tested in reading and math through the use of the SAT-10 and the ARMT. The subjects of language, science, and social science were dropped from the SAT-10 testing at the end of 2008-2009 school year. The SAT-10 and the ARMT are referred to as the high-stakes tests in the state of Alabama because they are the only two tests required for NCLB requirements. However, the Alabama Direct Assessment of Writing (ADAW) and the Alabama Science Assessment (ASA) are required for fifth grade students but not considered in the whether a school makes AYP for NCLB.

The SAT-10 is administered to students in Alabama in grades three through eight (Alabama State Department of Education, 2010). The SAT-10 is administered during the first 2 weeks of April. The Otis-Lennon School Ability Test (OLSAT) which is the companion school-ability assessment is also administered during the testing window. The purpose of the test is to compare individual and group performance with the performance of the norming group, to report relative strengths and weaknesses of individuals and groups, and to provide data to study changes in performance over time. Students in grades three through eight are administered the reading and math portion of the SAT 10. The results provide Alabama educators, parents, and the public a comparison of the per-

formance of Alabama's students, schools, school systems, and state to the performance of the nation.

The ARMT is administered to students in Alabama in grades three through eight (Alabama State Department of Education, 2010). The purpose of the ARMT is to assess students' mastery of state content standards in reading and mathematics, to report individual and group performance, to report relative strengths and weaknesses of individuals and groups, and to provide data to study changes in performance over time. The ARMT is a criterion-referenced test. It consists of selected items from the SAT-10, which matches the Alabama state content standards in reading and math. Additional test items were developed to be included so that all content standards were fully covered. It is the combination of SAT-10 items and newly developed items that is known as the ARMT. A student must take SAT-10 word study skills in grade three only, SAT-10 reading vocabulary, SAT-10 reading comprehension, and the ARMT part 2 reading subtest to get an ARMT reading score. A student must take SAT-10 math procedures, math problem solving, and the ARMT part 2 math subtest to get an ARMT math score. The performance is reported in the following achievement levels: Level I does not meet academic content standards. Level II partially meets academic content standards. Level III meets academic content standards (proficient or grade-level performance). Finally, Level IV exceeds academic content standards. The results are used for accountability for grades three through eight in meeting one of the requirements of the NCLB.

Although the SAT-10 and ARMT are used for high-stakes accountability purposes, the state of Alabama also requires fifth, seventh, and tenth grade students to take the ADAW (Alabama State Department of Education, 2010). The primary purposes of the

ADAW is to assess fifth grade students' writing performance in descriptive, narrative, and expository modes of writing and to report to the public the writing performance of students in these modes. The ADAW is administered in late February or early March and is timed. Performance on the ADAW is reported in the following achievement levels: Level I writing shows little understanding of the writing task. Level II writing reflects some understanding of the writing task but more author involvement than author control. Level III writing indicates a good understanding of the writing task and is sufficiently developed with a sense of audience, purpose, and author control. Level IV writing is thorough with a strong sense of audience and purpose and is precise, consistent, and elaborated with details that are clear and coherent.

The ASA is also required for fifth and seventh grade students (Alabama State Department of Education, 2010) but is not used for high-stakes accountability purposes. The ASA is a criterion-referenced test that consists of 66 multiple-choice questions. The ASA is aligned to the Alabama state content standards in science. The primary purposes of the ASA are to assess students' mastery of state content standards in science, to report individual and group performance, to report strengths and weaknesses of individuals and groups, and to provide data to study changes in performance over time. Fifth grade students are assessed on Alabama state content standards. Six of the standards are in physical science. Three standards are in life science, and two standards are in earth and space science. The performance on the ASA is reported in the following achievement levels: Level I signifies the student does not meet academic content standards. Level II signifies the student partially meets academic content standards. Level III signifies the student meets academic content standards (proficient or grade-level performance), and Level IV signifies the student exceeds academic content standards.

As mentioned by the Center of Public of Education (2008), emerging research shows testing can be beneficial. According to the Alabama School Journal (2008), 97.8% of the elementary schools in Alabama made AYP for the 2007-2008 school year. However, 16 of the 20 elementary schools statewide did not make AYP only because of the reading scores of special education students (Alabama School Journal, 2008). Additionally, according to the AASB Magazine (2007), Alabama moved up in ranking of state progress from fifth to 22nd under NCLB, which holds elementary schools in Alabama accountable for the scores on the ARMT for grades three through eight.

Also, an article in the Alabama Education News (2008) stated 83% of all Alabama schools including elementary schools and high schools made AYP for the 2008-2009 school year. A total of 1,367 Alabama public schools were evaluated for the 2008-2009 AYP status based on 2007-2008 data from state-mandated tests. This includes 857 Title I schools, which are high-poverty schools that receive federal funding. A total of 137 schools were identified as needing school improvement. In 2007, 89 Title I schools were identified for school improvement whereas for the 2008 school year, only 79 Title I schools were identified as needing school improvement which was an 18% decrease since the previous year indicating improved test scores.

Summary

The perceptual theory and the self-efficacy theory served as the theoretical framework for this study. The perceptual approach seeks to understand the behavior of the individual from his own point of view because “all behavior is a result of an individual’s personal meanings or perceptions” (Combs et al., 1976, p. 16). Combs and Snygg

(1959) stated, in order to function in society, individuals must be able to adapt to situations. Otherwise, individuals will become at risk for inappropriate behavior. Perceptions must be accessible to individuals because behavior is related to our perceptual field, and effective behavior occurs from a wide field of perceptions. Bandura (1986) stated that "perceived self-efficacy is defined as people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performance" (p. 391). Bandura (1994) contended that beliefs about one's self-efficacy determine how one acts, feels, thinks, and motivates himself.

There are opponents and proponents of high-stakes testing. Opponents believe special populations of students such as minorities, low-socioeconomic students, and students with special needs are also negatively impacted by high-stakes testing leaving behind the students NCLB intended to help (Gollnick & Chinn, 2002; Haynes, 2008; Huefner, 2006; Hursh, 2005; Jones et al., 2003; Savage, 2003; Smyth, 2008). Additionally, some teachers report negative effects of high-stakes testing on teaching practices while others report high-stakes testing has positively impacted teaching (Cankoy & Tut, 2005; Jones et al.; Paris & Urdan, 2000). The literature on high-stakes testing consists mainly of the negative impact high-stakes testing has on instruction, but emerging research shows that high-stakes testing can be beneficial if certain conditions exist (Center for Public Education, 2006). There are no other known studies regarding the impact of the SAT-10 and ARMT testing on classroom teaching practices. This study will extend the literature regarding high-stakes testing by exploring teachers' perceptions of the impact of SAT-10 and ARMT testing on elementary classroom teaching practices in the state of Alabama.

CHAPTER 3

METHODOLOGY

The purpose of this chapter is to provide a description of the methods employed in this mixed methods study. This chapter includes a review of the characteristics of mixed methods research, the researcher's philosophical assumptions that guided this investigation, a description of the mixed methods sequential explanatory research design used in this study, and the characteristics of mixed methods designs. Next, a visual diagram of the mixed methods procedures employed in this study is displayed. Moreover, a description of the target population, legitimation, research permission, and ethical considerations are presented. Finally, specific quantitative procedures used in phase I are described in chapter 4, and specific qualitative procedures used in phase II are described in chapter 5.

Characteristics of Mixed Methods Research

This study used a mixed methods research approach to answer the posed research questions. The design entailed the use of both quantitative and qualitative methods. According to Creswell (2008), mixed methods designs are procedures used to collect, analyze, and integrate quantitative and qualitative data in a study. The research context and the research purpose determine the methods that should be employed (Punch, 1998). The mixed methods approach is used when the researcher is unsure that one type of approach will adequately address the research problem (Creswell & Plano-Clark, 2007). In most

cases, mixing quantitative and qualitative data will yield the most precise and complete picture of the research problem (Teddlie & Tashakkori, 2009). In order to show a complete picture and to gain better understanding of the possible impact of the SAT-10 and the ARMT on classroom teaching practices, the mixed methods approach was employed (Teddlie & Tashakkori). In this study, the quantitative and qualitative data were collected sequentially, and the types of data were integrated at several stages in the research process. The data were analyzed and reported based on the sequence and priority of the data collected.

Teddlie and Tashakkori (2009) stated methods should be mixed so that there are nonoverlapping weaknesses and complementary strengths. Johnson and Turner (2003) referred to the nonoverlapping weaknesses and complementary strengths as the fundamental principle of mixed methods research. Johnson and Turner stated this fundamental principle must be followed for the following reasons: "(a) to obtain convergence or corroboration of findings, (b) to eliminate or minimize key plausible alternative explanations for conclusions drawn from the research data, and (c) to elucidate the divergent aspects of a phenomenon" (p. 299). Earlier, other researchers have identified five reasons to employ mixed methods research: triangulation, development, initiation, expansion, and complementarity (Greene et al., 1989). For the purposes of this study, complementarity was used which seeks elaboration, clarification, or enhancement from one research approach to another. Complementarity allowed the results from the survey data in the first, quantitative phase of the study to be further explored through the use of face-to-face interviews based on the survey data in the second, qualitative phase of the study. This enabled the

researcher to benefit from both the quantitative and qualitative strengths of the study and to counteract any biases from the research methods (Greene et al., 1989).

Philosophical Assumptions

The study adopted a pragmatic approach to answer the research questions. Pragmatists believe the research question is more important than the research method and more important than the worldview associated with the method (Creswell & Plano-Clark, 2007; Teddlie & Tashakkori, 2009). Pragmatism focuses on "what works" using different approaches and assigning value to objective and subjective knowledge (Maxcy, 2003).

Pragmatism is usually coupled with mixed methods research because it is not committed to one single philosophy (Creswell & Plano-Clark, 2007). Researchers exercise freedom in choosing the methods that will best answer the research question rather than committing to only one way of conducting research (Creswell, 2009). The research process involves combining and mixing both quantitative and qualitative data (Creswell & Plano Clark; Teddlie & Tashakkori, 2009). Multiple research roles are embraced which include biased and unbiased perspectives (Creswell & Plano-Clark). In summary, pragmatism is problem centered, pluralistic, real-world oriented, and focuses on consequences of actions (Creswell).

As outlined by Teddlie and Tashakkori (2009) and Hatch (2002), the following six philosophical assumptions are used when discussing the researcher's philosophical views that guided the study: ontology, epistemology, axiology, generalizations, causal linkages, and deductive/inductive logic. The following sections describe how these six assumptions guided this study.

Ontology

Ontology refers to individual perceptions of the nature of reality. Multiple, constructed realities existed in this study (Hatch, 2002; Tashakkori & Teddlie, 1998). Thus, different teacher participants in this study had different perceptions that may impact their realities. Teachers who administered the SAT-10 and the ARMT may have viewed testing as a positive impact on teaching practices whereas other teachers may have viewed this negatively. Based on how teachers teach the required objectives, high-stakes testing could have a negative effect, positive effect, or no effect at all on teaching practices. Consequently, there are multiple realities (Teddlie & Tashakkori, 2009). Therefore, quantitative research, through the use of a survey, was employed to examine a variety of teachers' perceptions regarding the impact of high-stakes testing. Additionally, the qualitative approach, based on participants' responses to interview questions, was employed to further examine the perspective of certain participants. Data for the study was collected by employing strategies that adequately addressed the research questions (Creswell & Plano Clark, 2007). Instead of searching for one truth, the researcher considered the truth to be multiple perceptions of the participants in the study. Because different teachers had different perceptions, there are various truths contained within the study. Through the use of surveys and interviews, the researcher had a greater opportunity to better understand the views of all participants and to capture their perceptions of reality.

Epistemology

Epistemology refers to the relationship of the knower to the known (Hatch, 2002; Tashakkori & Teddlie, 1998). In this study, the researcher and the participants' roles are

inseparable. The level of dependence between the researcher and the participants fluctuated within each phase of the study. During the first, quantitative phase of the study, participants completed the survey without dependence upon the researcher. The researcher did not have control over the survey responses or how the participants responded to the survey. During the second, qualitative phase of the study, the researcher collected data through the use of face-to-face interviews. The interview data from the participants was transcribed and coded by themes. During the second phase of this study, the participants depended on the researcher to accurately portray their perceptions during the analysis stage. Once the data were analyzed, a summary of the report was submitted to the participants. To ensure the findings of the study were accurate, each participant of the interview received a copy of the summary of the interview by mail (Creswell, 2005). The participants were asked to determine the accuracy of the report in writing and were able to make modifications to the report if they determined it did not accurately depict their views (Creswell).

Axiology

Axiology is the role of values in inquiry (Teddlie & Tashakkori, 1998). The relationship between the researcher and the participants was objective in the first, quantitative phase because the researcher did not interact with the participants. The participants completed the survey independent from the researcher. As a result, the research in the first, quantitative phase was value-free. In contrast, in the second, qualitative phase the relationship between the researcher and the participant was subjective. The researcher interacted with the participants during face-to-face interviews. The researcher values the

perceptions of the participants, who are also peers, because they work directly with students daily. Hence, the research in the second, qualitative phase is value-bound and is reflective of the core values the participants had about high-stakes testing and its impact on classroom teaching practices. It is anticipated the perceptions' of the participants will add to the literature base regarding the impact of high-stakes testing on classroom teaching practices.

Generalizations

In a mixed methods study, the researcher shifts between different modes of generalizability (Teddlie & Tashakkori, 2009). Because the first, quantitative phase of the study used deductive logic, the findings might be generalizable to similar settings (Teddlie & Tashakkori). Because the second, qualitative phase of the study used inductive logic, the findings are transferable only to settings with similar characteristics to the settings in this study (Teddlie & Tashakkori). The qualitative findings provided an individual insight into each explored case. However, the transferability of the second, qualitative phase of the study is limited because of the interpretive nature of qualitative research (Lincoln & Guba, 1985). The study encompassed the context of rural, urban, and suburban elementary schools and was bound by the time of the academic school year for the participants in the second, qualitative phase.

Causal Linkage

The findings of the second, qualitative phase provided an in depth understanding of the results from the first, quantitative phase and the impact of the SAT-10 and the

ARMT on classroom teaching practices. Causal relationships may have existed in the study (Tashakkori & Teddlie, 1998). Teacher perceptions of the impact of high-stakes testing on classroom teaching practices may be dependent on individual philosophies or perceptions. Yet, other teacher perceptions of high-stakes testing may be influenced by the socioeconomic status of the students they taught and the community in which they taught. Survey open-ended questions and individual interview questions were used to provide further explanations to the patterns that emerged from the quantitative data.

Deductive and/or Inductive logic

Deductive and inductive logic were used to enhance the richness of the study. The study began with deductive logic which was grounded in the results of previous studies, the perceptual theory, and the self-efficacy theory (Teddlie & Tashakkori, 2009). Deductive logic was implemented in the first, quantitative phase of the study to answer the research questions through the use of a survey. In the second, qualitative phase of the study inductive logic was used to generate the answers to the survey results grounded in individual participants' views.

Mixed Methods Sequential Explanatory Research Design

To address the research questions, the study employed the explanatory sequential design which consisted of collecting and analyzing quantitative and then qualitative data in two consecutive phases within one study (Ivankova, Creswell, & Stick, 2006). The rationale for using the explanatory sequential design was to provide a better understanding of the research problem because the qualitative data extended and elaborated on the ini-

tial quantitative results (Creswell, 2009; Creswell & Plano Clark, 2007). The quantitative data provided the general picture of the problem, while the qualitative data yielded the details of this picture and explanation of the trends found in the quantitative data. Specifically, the quantitative survey results were obtained to determine the trends and relationships of the teachers' perceptions of the impact of the SAT-10 and the ARMT on elementary classroom teaching practices in rural, urban, and suburban elementary schools, and the qualitative data from interviews explained and clarified the results from the statistical analyses from the first phase of the study for better understanding of how high-stakes testing impacts classroom instruction. During the connecting stage between the quantitative and qualitative phases, participants were selected for the qualitative follow-up phase, and an interview protocol for the qualitative phase of the study was developed and grounded in the quantitative results (Ivankova et al., 2006). Participants for the follow-up phase represented both typical and extreme cases.

Because quantitative research is generally weak at exploring the reasons for relationships among variables, a qualitative approach was needed to explain factors exploring the broad relationships among the variables in the second phase of the study (Punch, 1998). In essence, the participants were able to describe their specific experiences as classroom teachers by describing their daily experiences with teaching the curriculum, planning instruction, interactions with administrators and students, and personal reflections of their experiences. The exploration of the classroom teachers' specific experiences helped provide an elaborated account of the different ways high-stakes testing impacted classroom teaching practices.

The survey used for the quantitative phase of this study was a modified version of the survey developed by Moon et al. (2003), which they used to determine if teachers in high-versus low socioeconomic schools differed in instructional practices when high-stakes tests were given to the students they teach. In this study, the modified survey was used to examine teacher perceptions of the impact of high-stakes testing on classroom teaching practices specific to Alabama and to reflect the more recent perceptions of teachers seven to eight years after the passing of NCLB as opposed to teachers' perceptions shortly after its enactment as was done in the study conducted by Moon et al. Before qualitative data through the use of face-to-face interviews were collected, one teacher who completed a survey in the first, quantitative phase of the study piloted the interview. After the pilot interview was conducted and the interview protocol was modified, nine teachers were asked to participate in follow-up interviews about their responses to the survey in an effort to delve deeper into the teachers' perceptions of high-stakes testing. Thick, rich description was included in the qualitative phase of the study which led to a better understanding of the research problem because teachers were given the opportunity to go into more depth about their answers citing specific examples of their experiences. The use of both quantitative and qualitative approaches helped to "increase the scope, depth, and power" of the study and are characteristic of the mixed methods design (Punch, 1998, p. 243).

Characteristics of Mixed Methods Designs

According to Creswell and Plano Clark (2007), weighting, timing, and integration are essential characteristics of all mixed methods designs. Specific details regarding these

characteristics (weighting, timing, and integration) are discussed in the following paragraphs.

Weighting

In a mixed methods study weighting or priority refers to the emphasis placed on the data types that are used within the study (Creswell & Plano Clark, 2007). This is characteristic of all mixed methods research designs. In the sequential explanatory design, typically more priority is placed on the quantitative phase because it occurs first in the study and most of the data collection occurs during this phase (Ivankova et al., 2006). Consequently, in this study, greater emphasis or priority was placed on the quantitative component of the study because the second, qualitative phase was based on the results from the first, quantitative phase and complemented the statistical results obtained in the first phase. Specifically, the interview questions from the second, qualitative phase of the study were grounded in the responses from the survey in the first, quantitative phase of the study.

Timing

Timing or implementation is the sequence in which quantitative and qualitative data are collected in the study (Creswell & Plano Clark, 2007; Creswell, Plano Clark, Gutmann, & Hanson, 2003). In this study, data were collected and analyzed in a sequential fashion. Hence, the quantitative data were collected and analyzed before the qualitative data because the second, qualitative phase built on the statistical results from the sur-

vey in the first, quantitative phase and helped understand the meaning of these results in more depth (Creswell et al., 2003).

Integration/Mixing

According to Yin (2006), the mixing and integration of approaches within a single mixed methods study may occur at different levels. Integration is essential in the following components of the study: (a) research questions, (b) units of analysis, (c) samples for study, (d) instrumentation and data collection, and (e) analytic strategies (Yin, 2006). There is an overarching mixed methods question that addresses the overall goal of the study. The research questions are divided into quantitative and qualitative so that certain methods address the specific questions at each study phase. The units of analysis, which are the responses to the survey and the interview questions, keep the study together as a single study rather than multiple ones. Consequently, to ensure that the study was a single mixed methods study, participants for the second, qualitative phase were selected from the same pool of participants who responded to the survey in the first phase. Additionally, the qualitative sample was nested within the quantitative sample. Moreover, the interview questions that were asked in the qualitative phase complemented the items from the survey in the quantitative phase to ensure there was no divergence from the study purpose which would lead to multiple studies instead of one as mentioned by Yin. Finally, it was determined that “analyses should be formulated in directly analogous fashion” (Yin, 2006, p. 45). Consequently, both the quantitative and qualitative methods addressed the same range of variables and factors as opposed to different variables to ensure that a single study is being conducted. The quantitative and qualitative phases were connected by

examining the responses of the survey in the first, quantitative phase and developing an interview protocol for the second, qualitative phase of the study based on the analysis of the survey responses. Additionally, the selection of the participants in the qualitative phase of the study was also the connecting point between the first, quantitative phase and the second, qualitative phase because the results from the second, qualitative phase further explained the results of the first, quantitative phase.

Thus, the study methods were integrated in the statement of the problem, the purpose statement, the research questions, and the data collection and data analyses phases (Creswell & Plano Clark, 2007). Furthermore, the results of the quantitative phase in addition to the results from the qualitative phase were integrated at the data interpretation stage of the study where the meaning of both quantitative and qualitative results were explained and mixed methods inferences from the whole study were generated (Hansen et al., 2005; Ivankova et al., 2006).

Visual Diagram

A visual diagram is needed to represent the complex mixed methods procedures in the study (Morse, 1991). The visual diagram allows one to view the procedures and outcomes of each phase (Ivankova et al., 2006; Morse). A visual diagram of the procedures outlined in this study is presented in Figure 1. The capital letters in the diagram indicate the weighting or priority of the quantitative (QUAN) method has more priority in the study; lower case letters indicate the qualitative (qual) method has less priority. The arrows show the sequence of the methodologies employed and the flow of the phases of

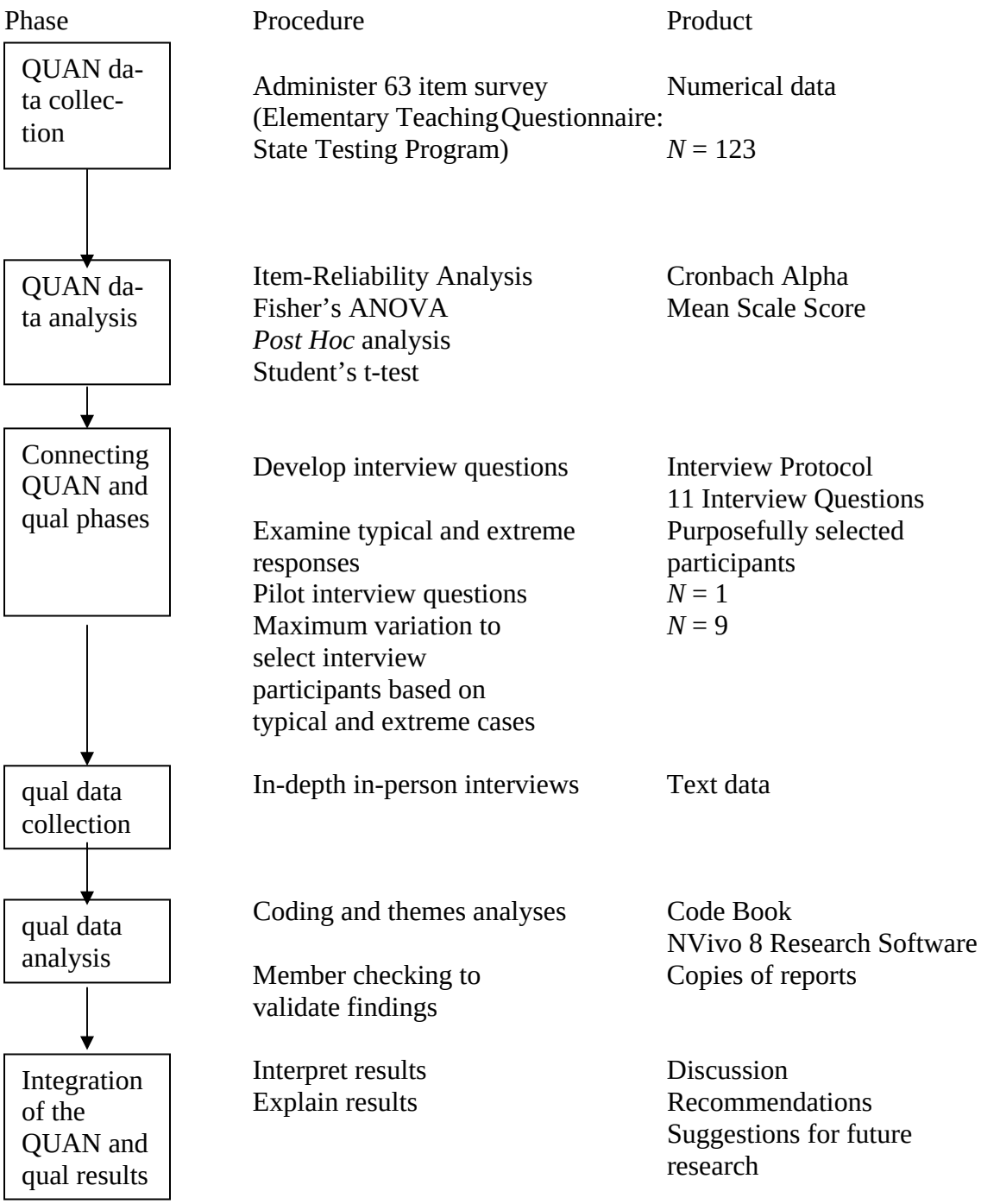


Figure 1. Visual diagram of sequential explanatory design. Using format and guidelines developed by Ivankova et al. (2006). Using mixed methods sequential explanatory design: From theory to practice. *Field Methods, 18*(1), 3-20.

data collection and analyses. The diagram also indicates the stages where connection and integration of the quantitative and qualitative methods occur.

Target Population

The target population in this study consisted of 362 elementary teachers of grades three, four, and five in three large school systems in the state of Alabama who administer the SAT-10 and the ARMT. A total of 123 surveys were completed making the return rate 33%. The three school systems were chosen because of the large number of teachers in the school systems, the diverse economic status of the students, and the school geographical location. School Systems A and B serve rural and suburban students. School System C serves urban students. These school systems serve diverse communities and giving teachers an opportunity to discuss problems specific to their community was imperative and enhanced the richness and significance of the study findings.

After the analyses of the quantitative data were completed, maximum variation sampling (Creswell & Plano Clark, 2007) was used to select extreme case and typical case participants from rural, urban, and suburban areas for face-to-face interviews in the qualitative phase. One participant who served in a suburban school piloted the interview questions. A total of nine other teachers were interviewed. Four teachers served in rural schools, three teachers served in suburban schools, and two teachers served in urban schools. The interview participants' responses to the interview protocol were used as data for the second, qualitative phase of the study.

Legitimation

The term legitimation is used for establishing the validity for mixed methods research (Onwuegbuzie & Johnson, 2006). Mixed methods research consists of "combining complementary strengths and nonoverlapping weaknesses of quantitative and qualitative research methods" (Plano Clark & Creswell, 2008, p. 294). As a result, evaluating the validity of findings in mixed methods research can become complex resulting in a problem of the integration (Plano Clark & Creswell). Hence, the term legitimation was established in an effort to develop a common language that quantitative and qualitative researchers could both use (Johnson & Onwuegbuzie, 2004; Plano Clark & Creswell; Tashakkori & Teddlie, 1998). Legitimation is a method of evaluation used to ensure the data analyses in mixed methods research are of high quality and the inferences as a result of the mixing of data are credible, trustworthy, and transferable (Creswell & Plano Clark, 2007; Johnson & Onwuegbuzie; Tashakkori & Teddlie). To increase the credibility, transferability, and trustworthiness of this study, the researcher used sequential legitimation, sample integration legitimation, inside-outside legitimation, and weakness minimization.

Sequential Legitimation

Sequential legitimation is "the extent to which one has minimized the potential problem wherein the metainferences could be affected by reversing the sequence of the quantitative and qualitative phases" (Onwuegbuzie & Johnson, 2006, p. 57). The quantitative results from the first, quantitative phase could have threatened the results from the second, qualitative phase if the answers on the survey did not represent the teachers' per-

ceptions. Additionally, the participants' answers during the interview may not have represented the views of all elementary teachers of grades three, four, and five. These threats were minimized by following the steps to ensure reliability and validity during the first, quantitative phase and to ensure credibility and trustworthiness during the second, qualitative phase of the study (Onwuegbuzie & Johnson).

Sample Integration Legitimation

Sample integration is defined as the degree to which integration of the quantitative and qualitative sampling designs generate quality meta-inferences (Onwuegbuzie & Johnson, 2006). The sequential explanatory design has a nested relationship in which the sample members selected for one phase of the study are a subset for the other phase of the study (Onwuegbuzie & Collins, 2007). Maximum variation sampling was used in the second, qualitative phase to select nine participants for interviews who completed a survey in the first, quantitative phase of the study. As a result, the quantitative and qualitative samples were integrated because in the second, qualitative phase of the study, the participants from the first, quantitative phase were asked to follow up on the results from the initial phase of the study.

Inside-Outside Legitimation

Inside-outside legitimation is "the extent to which the researcher accurately presents and appropriately utilizes the insider's view and the observer's view for purposes such as description and explanation" (Onwuegbuzie & Johnson, 2006, p. 57). This type of legitimation seeks a balance between the outsider viewpoint that took place in the

first, quantitative phase with the insider viewpoint, which took place in the second, qualitative phase. Member checking was used to ensure the researcher accurately depicted the participants' perceptions.

Weakness Minimization

Weakness minimization is employed when the weakness of a research approach is compensated by the strengths of another research approach (Plano Clark & Creswell, 2008). Specifically, the qualitative research approach was used in the second, qualitative phase to compensate for the initial survey data collection due to the fact that the quantitative data alone did not fully address the research questions because the quantitative results showed generalizations and trends. The second, qualitative phase was used to further elaborate the findings from the first, quantitative phase (Creswell & Plano Clark, 2007).

Research Permission and Ethical Considerations

Ethical considerations must be addressed prior to conducting research. The researcher was trained in conducting ethical research. The ethical considerations regarding the protection of the participants in the quantitative and qualitative phases of this study were anonymity and confidentiality which the researcher ensured were employed in both phases of the study.

The first ethical consideration consisted of obtaining permissions from the boards of education in three different school systems. Afterwards, University of Alabama at Birmingham Institutional Review Board (UAB IRB) approval was obtained (Creswell,

2008). Once UAB IRB approval was granted the researcher emailed principals and solicited voluntary participation of the teachers from the three school systems to respond to a survey and possibly a follow-up face to face interview conducted by the researcher. Consent forms informing the participants of the purpose, benefits, risks, confidentiality, and the option to withdraw from the study were given to the participants prior to collecting data. To ensure anonymity, completed surveys were given a numerical code in the first, quantitative phase.

Due to the nature of the sequential explanatory design, the survey was first submitted for review to the UAB IRB for approval because the interview questions for the second, qualitative phase of the study could not be developed until the data from the surveys in the first, quantitative phase of the study were analyzed. After the data from the surveys were analyzed, the follow-up interview questions were developed and submitted to the UAB IRB for approval prior to conducting the interviews for the qualitative phase of the study. After UAB IRB approval of the interview protocol was granted, participants were contacted and asked to participate in follow-up interviews in the second, qualitative phase of the study. Fifteen teachers were asked to participate in an interview. Nine of the teachers agreed and scheduled an interview with the researcher. Participants in the second, qualitative phase were assigned pseudonyms to ensure anonymity and to preserve confidentiality. Moreover, quantitative and qualitative data were stored in a locked metal file cabinet in the researcher's home. All confidential data will be destroyed 3 years after the conclusion of the study.

Role of the Researcher

The researcher is cognizant of the roles of elementary teachers. For 12 years, the researcher taught in various elementary schools. Two of the schools were located in a low-socioeconomic rural setting, and the other two schools were located in a high-socioeconomic suburban setting. Moreover, the researcher taught first grade for 9 years and third grade for 3 years. While teaching third grade, the researcher administered the SAT-10 and the ARMT to third graders in the high-socioeconomic, suburban schools in which she taught.

The researcher's involvement with the research participants fell along a continuum that ranged from objective to subjective. During the first, quantitative phase the researchers did not interact with the participants. The participants were selected from school systems within the three targeted school systems in Alabama. Participants names were obtained from school web sites, and then the completed surveys were assigned a numerical code to protect confidentiality. During the second, qualitative phase the researcher interacted with one participant who piloted the interview questions and nine purposefully selected participants in face-to-face interviews.

Summary

Chapter 3 described the methodology used in this study, the characteristics of mixed methods research, philosophical assumptions, the sequential explanatory design, legitimation, permission, and ethical considerations. The sequential explanatory mixed methods research design was used in this study to better answer the research questions. The use of both quantitative and qualitative approaches helped to "increase the scope,

depth, and power" of the study (Punch, 1998, p. 243). The results of the quantitative phase and the qualitative phase are reported separately in chapter 5 and chapter 6, respectively.

CHAPTER 4

PHASE I: QUANTITATIVE METHODS AND RESULTS

Chapter 4 describes the methods and results of the first, quantitative phase of the study. A description of the variables and the instrument used to test the variables are described in this chapter. Additionally, quantitative sample recruitment, survey administration, procedures for descriptive statistics, and item-reliability analysis are addressed. Finally, quantitative results for each research question are summarized and discussed.

Methods and Procedures

Variables in the Quantitative Analysis

The focus of the quantitative phase of this study was to understand the relationship between high-stakes testing and classroom teaching practices. The relationship between two independent variables and five dependent variables was analyzed. The independent variables in this study are school geographical location and the socioeconomic status of the students taught. The dependent variables are curriculum and instructional approaches, the amount of time spent on critical thinking skills, the amount of time spent on test preparation activities, the perceived impact of state tests on students, and the perceived impact of state tests on teachers. School geographical location included three levels: rural, urban, and suburban. Socioeconomic status of the students taught was calculated by the type of lunch received by the majority of the students at the school where the teacher served and included two levels: fully paid lunch and free or reduced lunch.

Measurement

Elementary Teaching Questionnaire (State Testing Program)

The questionnaire used in this study was modified from the original survey developed by Moon et al. (2003) at the University of Virginia. Permission to use the survey or a modified version of the survey was obtained from Moon et al. who developed the survey (Appendix B). The original survey consisted of 99 items, but 59 of the items were retained by the authors after principal component analysis was conducted. The original survey was used for a quantitative study to determine if teachers in high-versus low socioeconomic schools differed in instructional practices when high-stakes tests were given to the students they teach. The original questionnaire inquired about school geographical location and school poverty levels, the perceived impact of state-mandated testing on curriculum and instruction, pressure to improve test scores, the amount of time spent for test preparation, the perceived effects of standardized testing, and the teacher's perceptions of the consequences of testing (Moon et al., 2003).

The original 99 item survey developed by Moon et al. (2003) was modified and used for this study (Appendix C). Questions that did not align with the research questions for this study were deleted, and questions that were not Likert-type scale items were modified to Likert-type scale items. As a result of the deletion of items, a total of 63 items were retained from the original survey. Additionally, for the purposes of this study, the first question from the original survey which asked "In what state do you teach?" was changed to "In what type of community do you teach: rural, urban, or suburban?" Moreover, so that first year teachers could be included in the study, questions that asked how instruction had changed as a result of the state-testing program were deleted in addition to

a question that asked how the outcomes of state-level tests had changed over the past 3 years. Information about the percentages of the students receiving fully paid and free or reduced lunch at the respective schools was obtained from the Alabama State Department of Education website which is accessible to the public.

Quantitative Data Collection

Quantitative Sample Recruitment

The study target population was elementary teachers of grades three, four, and five in three large school systems in the state of Alabama who administer the SAT-10 and the ARMT. The three school systems were chosen because of the large number of teachers in the school systems, the diverse economic status of the students, and the geographical location of the school. School Systems A and B served rural and suburban areas. School System C served an urban area. Once the University of Alabama Institutional Review Board (UAB IRB) granted approval for the study (Appendices D and E), each elementary principal in all three school systems was sent an email asking permission to send the survey to the entire population of third through fifth grade teachers at their respective schools to ensure an adequate sample size was obtained. All of the principals who responded to the email agreed to give permission for the researcher to conduct research at their school sites.

Survey Administration

On January 21, 2009 the researcher sent emails to the elementary principals asking to conduct the study at their school site. Upon approval of the principal, a packet was

mailed to the teachers at the chosen schools. The packet included a recruitment letter (Appendix F) two copies of the consent form (Appendix G), and the surveys along with a self-addressed stamped envelope. One signed consent form from the participant was kept on file, and the second consent form was kept by the participant for his or her records. As the surveys were returned to the researcher's home address, the researcher assigned a different numerical code to each survey to keep track of which participants completed surveys. To increase the response rate, the three step follow-up procedure as outlined by Dillman (2008) was used. A reminder postcard was sent to the teachers one week after mailing the survey. Two weeks after the postcards had been mailed, a letter along with another survey was mailed to the non-respondents. Four weeks later the researcher emailed the nonrespondents. In June of 2009, after surveys had been collected, the data were compiled into a spreadsheet. In October of 2009, the data were entered into the SPSS computer software program version 17.0 for analysis.

Description of the Sample

A total of 123 surveys of 362 were returned yielding a return rate of 33%. Of the 123 completed surveys 57 (46.3%) were completed by suburban teachers, 44 (35.8%) were completed by rural teachers, and 22 (17.9%) were completed by urban teachers. Regarding socioeconomic status, 63 (51.2%) of the 123 surveys were completed by teachers who served in schools with the majority of the students fully paying for lunch and 60 (48.8%) of the surveys were completed by teachers who served in schools with the majority of the students receiving free and/or reduced lunch. The description of the demographic data, which includes the school geographical location, the socioeconomic status

of the students taught, gender, and years of experience of the participants is presented in Table 1.

Table 1

*Demographic Information for the Sample (*N*=123)*

Variables	Frequency	Percent
Location		
Rural	44	35.8
Suburban	57	46.3
Urban	22	17.9
Lunch		
Free/reduced	60	48.8
Fully paid	63	51.2
Gender		
Male	2	2.0
Female	121	98.0
Years of Experience		
>-4	51	41.0
5-9	30	24.0
10-14	22	18.0
15-19	11	9.0
20-24	2	2.0
25-29	2	2.0
30-34	5	4.0

Quantitative Data Analysis

The data were analyzed using the SPSS computer software program version 17.0. Descriptive statistics yielded mean scale scores and standard deviations for each of the factors. The statistical procedures to answer the research questions included ANOVA, Post Hoc tests, and Student t-tests. ANOVA is a test of significance used to test for differences among more than two sample means and is designed to be used with interval-ratio-level dependent variables (Healey, 2002). School geographical location was factorialized into three levels: urban, suburban, and rural. ANOVA was performed to test for

significant differences among the school geographical locations. Additionally, a post hoc test is a technique used to determine which pairs of means are significantly different (Healey, 2002). Post Hoc analysis was performed to identify which geographical regions was the source of the statistically significant differences. Finally, the Student t- test is used when an independent variable has only two categories (Healey, 2002). Socioeconomic status was factorialized into two levels: fully paid and free or reduced lunch and was analyzed using Student's t-test.

Factor Analysis

Factor analysis was used to test for the construct validity of the modified survey. The first step included generating a correlation matrix for all 63 items. Pearson's correlation was used for continuous data, and Spearman's Rho was used for categorical data. Using the matrix, factor analysis was conducted utilizing the principal component analysis (PCA) approach and varimax rotation. PCA is used in factor analysis to extract factors (Mertler & Vannatta, 2005). Specifically, "original variables are transformed into a new set of linear combinations by extracting the maximum variance from the data set with each component" (Mertler & Vannatta, 2005, p. 343). Varimax rotation is a procedure used to "maximize the variance of factor loadings by making high loadings higher and low ones lower for each factor" (Tabachnick & Fidell, 2007, p. 620).

A scree plot is generated in PCA. A scree plot is "a graph of the magnitude of each eigenvalue (vertical axis) plotted against their ordinal numbers (horizontal axis)" (Mertler & Vannatta, 2005, p. 344). The scree plot is similar to scree during mountain climbing. Scree is formed at the base of the mountain and drops off to level ground where

the side of the mountain ends. Like scree on a mountain, for this study the bend in the scree plot seen between factors 3 through 12 suggests the number of factors found in the instrument to be between factors 3 and 12. Based on the scree plot an initial decision was made to retain 6 factors for further analysis. After further removing items that loaded on multiple factors or failed to load on any factors, 37 items loading on six factors remained. The number of six factors was consistent with the scree plot. The loading of 37 items on six individual factors was also consistent with the research questions posed for the first, quantitative phase of the study. The item loadings on the individual factors are presented in Table 2.

Table 3 demonstrates how 37 survey items that were retained after the factor analysis are related to the research questions that were addressed in the first, quantitative phase of the study.

Item-Reliability Analysis

An item-reliability was performed on each factor using Cronbach's Alpha. The alpha values for the factors ranged from 0.65 to 0.90. An alpha of 0.70 indicates a modest reliability (Nunally, 1994). Two of the scales yielded the reliability score below 0.70, whereas the scores for four other scales exceed it. Table 4 summarizes the item reliabilities for each factor.

Table 2

Factor Loadings of Survey Items

	Item from Survey	I	II	III	IV	V	VI
8b.	Using an interdisciplinary curriculum				0.66		
8c.	Differentiating the curriculum (e.g. using tiered assignment, lessons targeted to student past achievements, accelerating pace of learning for some students)				0.64		
8d.	Doing hands-on work				0.71		
8g.	Teaching which allows for in-depth exploration where one critical exemplar of a concept or principle can be understood as a basis to generalize to other exemplars				0.63		
8i.	SAT-10 and ARMT help clarify and specify learning goals					0.66	
8k.	I teach to the SAT-10 and ARMT more than I normally would					0.68	
9a.	Use of constructed response items (short essays)						0.44
9b.	Use of multiple-choice items		0.46				
9d.	Use of performance type items (e.g., presentations, science experiments)					0.56	
9f.	Making sure the content and skills covered on the SAT-10 and ARMT are reviewed prior to the test administration		0.54				
10a.	Higher-order thinking skills						0.74
10b.	Problem-solving skills						0.74
10c.	Topics which are not assessed on the SAT-10 and the ARMT						0.66
10d.	The fine and performing arts (e.g., music, art, drama)					0.66	
11a.	Student worksheets				0.56		
11b.	Instruction for students on test-taking strategies		0.72				
11d.	Student practice in the kinds of item formats that are on the SAT-10 and the ARMT		0.78				
11f.	Instruction for students on test-taking strategies		0.87				

Table 2. (Continued)

Item from Survey	I	II	III	IV	V	VI
11g. Review/practice using state released test items		0.74				
11h. Student practice in the kinds of item formats that are on the SAT-10 and the ARMT		0.89				
11j. Instruction for students on test-taking strategies		0.62				
11l. Student practice in the kinds of item formats that are on the SAT-10 and the ARMT		0.68				
11m. Student worksheets				0.58		
12a. Reviews test scores at staff meetings			0.80			
12b. Discusses ways to improve test scores			0.88			
12c. Provides materials to improve test scores			0.75			
12d. Checks to see that teachers are emphasizing areas which showed weakness from past test results			0.67			
12e. Introduces or discusses important new instructional ideas			0.83			
13d. Students are treated as test-takers rather than learners	0.84					
13e. Students are under too much pressure to increase test scores	0.86					
13f. Students see learning as a chore because of pressure from SAT-10 and ARMT testing	0.74					
13g. Students feel bad if they do not have high test scores	0.76					
14a. Having to prepare students for the SAT-10 and ARMT impacts my approach to teaching	0.78					
14b. My current students' most recent test results impact my approach to teaching	0.76					
14c. Our school's overall test results impact my approach to teaching	0.66					
14f. Teachers in my school feel there is discrepancy between what they think should be taught and what the SAT-10 and ARMT emphasize	0.53					

Table 3

Research Question and Item Survey Numbers after Factor Analysis

Research Question	Items from Survey
Are there differences in mean score on teacher curriculum approaches when controlled for school geographical location and by by socioeconomic status of the students taught?	8b, 8c, 8d, 8e, 8g, 11a, 11m
Are there differences in mean score on teacher instructional approaches when Controlled for by school geographical location and by socioeconomic status of the students taught?	9d, 10d, 8i, 8k
Are there differences in mean score on the class time spent on critical thinking skills when controlled for by school geographical location and by socioeconomic status of the students taught?	9a, 10a, 10b, 10c
Are there differences in mean score on the amount of time spent on school-wide test preparation activities with administrators when controlled for by school geographical location and by socioeconomic status of the students taught?	12a, 12b, 12c, 12d, 12e
Are there differences in mean score on the amount of class time spent on classroom test preparation when controlled for by school geographical location and by socioeconomic status of the students taught?	9b, 9f, 11b, 11d, 11f, 11g, 11h, 11j, 11l
Are there differences in mean score on the perceived impact of state tests on students and teachers when controlled for by school geographical location and by socioeconomic status of the students taught?	13d, 13e, 13f, 13g, 14a, 14b, 14c, 14f

Table 4

Summary of Item Reliabilities for Each Scale

Factor Description	Cronbach's Alpha
Curriculum Approaches	0.73
Instructional Approaches	0.65
Time Spent on Critical Thinking Skills	0.67
Time Spent on School-Wide Test Preparation With Administrators	0.88
Time Spent on Classroom Test Preparation	0.88
Perceived Impact of State Tests on Students and Teachers	0.90

Quantitative Results

The mean scale score and standard deviation for each factor yielded the following results respectively: curriculum approaches 24.9(4.3), instructional approaches 14.1(3.0), time spent on critical thinking skills 9.2(2.8), time spent on school-wide test preparation with administrators 17.9(3.3), time spent on classroom test preparation 20.8(7.1), perceived impact of state tests on students and teachers 31.6(6.3). The summary of mean scale scores and standard deviations is presented in Table 5.

Table 5

Summary of Mean Scale Scores and Standard Deviations

Factor Description	*M*	*SD*
Curriculum Approaches	24.9	4.3
Instructional Approaches	14.1	3.0
Time Spent on Critical Thinking Skills	9.2	2.8
Time Spent on School-Wide Test Preparation With Administrators	17.9	3.3
Time Spent on Classroom Test Preparation	20.8	7.1
Perceived Impact of State Tests on Students and Teachers	31.6	6.3

The presentation of the results of the statistical analyses is organized by research questions. The research questions and the results of the statistical analysis are presented in the following paragraphs.

Research Question 1

Are there differences in mean score on teacher curriculum approaches when controlled for by school geographical location and by socioeconomic status of the students taught?

Mean factor score differences by school geographical location for curriculum approaches were tested using ANOVA. The results did not yield statistically significant differences on mean scale score with $F(2,120) = 1.83$, $p = .16$. This indicates that rural, urban, and suburban teachers had similar perceptions of the impact of high-stakes testing on their curriculum practices. Table 6 summarizes the ANOVA results.

Table 6

ANOVA Results for Curriculum Approaches and Location

Scale		Sum of squares	df	Mean square	F	Sig.
1	Between Groups	67.443	2	33.722	1.826	.165
	Within Groups	2216.036	120	18.467		
	Total	2283.480	122			

$N = 123$

Mean factor score differences by socioeconomic status for curriculum approaches were tested using a Student t-test. The results did not yield statistically significant differences in the mean scores of participants who taught in low-socioeconomic and high-socioeconomic schools ($t = 1.77$, $df = 116.60$, $p = 0.04$, one-tailed). Hence, the difference

between the values tested is statistically zero indicating the responses among teachers of low-socioeconomic and high-socioeconomic students were similar. Yet, emerging trends existed because the probabilities of these tests fell within the interval $0.05 < p < =0.10$. A larger sample size might allow these differences to emerge as statistically significant if the tests were repeated on a larger sample size. Table 7 summarizes the Student t-test results.

Table 7

Student t-test Results for Curriculum Approaches and SES

Scale	Description	M	Mean Difference	t	df	Sig. (2-tailed)
1	Equal Variances Assumed	25.63	1.36	1.76	121	.081
	Equal Variances Not Assumed	24.30	1.36	1.77	116.60	.079

$N = 123$

Research Question 2

Are there differences in mean score on teacher instructional approaches when controlled for by school geographical location and by socioeconomic status of the students taught?

Mean factor score differences by school geographical location for instructional approaches were tested using ANOVA. The results did not yield statistically significant differences on mean scale score with $F(2,120) = 1.021$ and $p = .363$. This indicates that high-stakes testing similarly affected the teachers across all there school geographical areas. Table 8 summarizes the ANOVA results.

Table 8

ANOVA Results for Instructional Approaches and School Geographical Location

Scale		Sum of squares	df	Mean square	F	Sig.
2	Between Groups	18.892	2	9.446	1.021	.363
	Within Groups	1109.759	120	9.248		
	Total	1128.650	122			

N = 123

Mean factor score differences by socioeconomic status for instructional approaches were tested using Student's t-test of independent samples. The results did not yield statistically significant differences in the mean scores of participants who taught in low-socioeconomic and high-socioeconomic schools (t = -.076, df = 121, p = 0.47, one-tailed). Hence, the difference between the values tested is statistically zero indicating the responses among teachers of low-socioeconomic and high-socioeconomic students were similar. Table 9 summarizes the Student t-test results.

Table 9

Student t-test Results for Instructional Approaches and SES

Scale	Description	M	Mean Difference	t	df	Sig. (2-tailed)
2	Equal Variances Assumed	14.11	1.36	-0.76	121	.939
	Equal Variances Not Assumed	14.16	1.36	-0.76	115.01	.940

N = 123

Research Question 3

Are there differences in mean score on the class time spent on critical thinking skills when controlled for by school geographical location and by socioeconomic status of the students taught?

Mean factor score differences by school geographical location for the amount of time spent on critical thinking skills were tested using ANOVA. The results yielded statistically significant differences on mean scale score with F(2,120) = 12.496, p = .000. This indicates that high-stakes testing differently impacted teachers in rural, suburban, and urban schools. Post Hoc Analysis was conducted to identify the differences among the three school geographical locations. Table 10 summarizes the ANOVA results.

Table 10

ANOVA Results for Time Spent on Critical Thinking Skills and Location

Scale		Sum of squares	df	Mean square	F	Sig.
3	Between Groups	165.023	2	82.512	12.496	.000
	Within Groups	792.392	120	6.603		
	Total	957.415	122			

N = 123

Post Hoc Analysis yielded the following results with mean differences reported between school locations. The mean differences between rural and suburban schools were -3.20973, p = .000. The mean differences between teacher responses in rural and urban schools were -2.54545, p = .001. The mean differences between teacher responses in suburban and urban schools were .66427, p = .439. This indicates urban teacher responses were significantly different from both suburban and rural teacher responses. The urban teacher responses were the lowest. Rural responses followed the urban responses, and suburban responses were the highest. This indicated urban teachers spent more time on critical thinking skills than rural and suburban teachers. On the opposite, suburban teachers spent less time on critical thinking skills than both rural and urban teachers. Table 11 summarizes the Post Hoc Analysis results.

Table 11

Post Hoc Analysis Results for Time Spent on Critical Thinking Skills and Location

Dep. Variable	(I) setting	(J) setting	Mean Difference (I-J)	Std. Error	Sig	95% CI Lower Bound	95% CI Upper Bound
Scale 3	1.00	2.00	-3.20973*	.64498	.000	-4.8084	-1.6111
		3.00	-2.54545*	.67099	.001	-4.2086	-.8823
	2.00	1.00	3.20973*	.64498	.000	1.6111	4.8084
		3.00	.66427	.51568	.439	-.6139	1.9424
	3.00	1.00	2.54545*	.67099	.001	.8823	4.2086
		2.00	-.66427	.51568	.439	-1.9424	.6139

N = 123 *The mean difference is significant at the 0.05 level.

Mean factor score differences by socioeconomic status for the amount of time spent on critical thinking skills were tested using Student's t-test of independent samples. The results did not yield statistically significant differences in the mean scores of participants who taught in low-socioeconomic and high-socioeconomic schools (t = -1.37, df =121, p = 0.08, one-tailed). Hence, the difference between the values tested is statistically zero indicating the responses among teachers of low-socioeconomic and high-socioeconomic students were similar. Table 12 summarizes the Student t-test results.

Table 12

Student t-test Results for Time Spent on Critical Thinking Skills and SES

Scale	Description	M	Mean Difference	t	df	Sig. (2-tailed)
3	Equal Variances Assumed	8.82	-.691	-1.37	121	.172
	Equal Variances Not Assumed	9.50	-.691	-1.37	115.01	.172

N = 123

Research Question 4

Are there differences in mean score on the amount of time spent on school-wide test preparation activities with administrators when controlled for by school geographical location and by socioeconomic status of the students taught?

Mean factor score differences by school geographical location for the amount of time spent on school-wide test preparation activities with administrators were tested using ANOVA. The results did not yield statistically significant differences on mean scale score with $F(2,120) = .609$ and $p = .546$. This indicates that high-stakes testing similarly impacted teachers' perceptions of the time spent on school-wide test preparation activities with administrators across three school geographical locations. Table 13 summarizes the ANOVA results.

Table 13

ANOVA Results for Time Spent on School-Wide Test Preparation Activities and Location

Scale		Sum of squares	df	Mean square	F	Sig.
4	Between Groups	13.383	2	6.691	.609	.546
	Within Groups	1318.243	120	10.985		
	Total	1331.626	122			

$N = 123$

Mean factor score differences by socioeconomic status for the amount of time spent on school-wide test preparation activities were tested using Student t- test. The results did not yield statistically significant differences in the mean scores of participants who taught in low-socioeconomic and high-socioeconomic schools ($t = -1.79$, df $=112.03$, $p = .04$, one-tailed). Hence, the difference between the values tested is statistically zero indicating the responses among teachers of low-socioeconomic and high-

socioeconomic students were similar. Yet, emerging trends existed because the probabilities of these tests fell within the interval $0.05 < p < =0.10$. A larger sample size might allow these differences to emerge as statistically significant if the tests were repeated on a larger sample size. Table 14 summarizes the results of the Student t-test.

Table 14

Student t-test Results for Time Spent on School-Wide Test Preparation Activities and SES

Scale	Description	M	Mean Difference	t	df	Sig. (2-tailed)
4	Equal Variances Assumed	17.35	-1.06	-1.79	121	.074
	Equal Variances Not Assumed	18.41	-1.06	-1.79	112.03	.076

N = 123

Research Question 5

Are there differences in mean score on the amount of class time spent on classroom test preparation activities when controlled for by school geographical location and by socioeconomic status of the students taught?

Mean factor score differences by school geographical locations for the amount of time spent on classroom test preparation activities and school geographical location were tested using ANOVA. The results yielded no significant differences on mean scale score with $F(2,120) = 2.100$, $p = .127$. This indicates that teachers across all three school geographical areas had similar perceptions of the time spent on classroom test preparation activities. Table 15 summarizes the ANOVA results.

Table 15

ANOVA Results for Amount of Time Spent on Classroom Test Preparation and Location

Scale		Sum of squares	df	Mean square	F	Sig.
5	Between Groups	207.523	2	103.762	2.100	.127
	Within Groups	5929.892	120	49.416		
	Total	6137.415	122			

N = 123

Mean factor score differences by socioeconomic status for the amount of time spent on classroom test preparation were tested using Student's t-test of independent samples. The results did not yield statistically significant differences in the mean scores of participants who taught in low-socioeconomic and high-socioeconomic schools (t = -0.91, df = 121, p = 0.18, one-tailed). Hence, the difference between the values tested is statistically zero indicating the responses among teachers of low-socioeconomic and high-socioeconomic students were similar. Table 16 summarizes the results.

Table 16

Student's t-test Results for Amount of Time Spent on Classroom Test Preparation and SES

Scale	Description	M	Mean Difference	t	df	Sig. (2-tailed)
5	Equal Variances Assumed	20.23	-1.16	-0.91	121	.365
	Equal Variances Not Assumed	21.40	-1.16	-0.91	115.85	.367

N = 123

Research Question 6

Are there differences in mean score on the perceived impact of state tests on students and teachers when controlled for by school geographical location and by socioeconomic status of the students taught?

Mean factor score differences by school geographical location for the perceived impact of state tests on students and teachers were tested using ANOVA. The results yielded significant differences on mean scale score with $F(2,120) = 14.8$, $p < .001$. Post Hoc Analysis was conducted to identify the differences among the three school geographical locations. Table 17 summarizes the ANOVA results.

Table 17

ANOVA Results for Perceived Impact on Students and Teachers and Location

Scale		Sum of squares	df	Mean square	F	Sig.
6	Between Groups	959.141	2	479.570	14.802	.000
	Within Groups	3887.786	120	32.398		
	Total	4846.927	122			

$N = 123$

Post Hoc Analysis was performed to identify which school geographical regions was the source of the significant differences. Post Hoc Analysis yielded the following results with mean differences reported between school locations. The mean differences between rural and suburban teacher responses were -7.14035, $p = .000$. The mean differences between rural and urban teacher responses were -7.45455, $p = .000$. The mean differences between suburban and urban teacher responses were -.31419, $p = .963$. Urban teacher responses were the lowest. Suburban teacher responses followed urban teacher responses, and rural responses were the highest. This indicates that urban teachers per-

ceived students and teachers were not as greatly impacted by high-stakes testing as opposed to suburban and rural teachers. Rural teachers perceived students and teachers were more greatly impacted by high-stakes testing than urban and suburban teachers. Table 18 summarizes the Post Hoc analysis results.

Table 18

Post Hoc Analysis Results for Perceived Impact on Students and Teachers and Location

Dep. Variable	(I) setting	(J) setting	Mean Difference (I-J)	Std. Error	Sig	95% CI Lower Bound	95% CI Upper Bound
Scale 6	1.00	2.00	-7.14035*	1.42865	.000	-10.6814	-3.5993
		3.00	-7.45455*	1.48626	.000	-11.1384	-3.7707
	2.00	1.00	7.14035*	1.42865	.000	3.5993	10.6814
		3.00	-.31419	1.14224	.963	-3.1454	2.5170
	3.00	1.00	7.45455*	1.48626	.000	3.7707	11.1384
		2.00	.31419	1.14224	.963	-2.5170	3.1454

N = 123 *The mean difference is significant at the 0.05 level.

Mean factor score differences by socioeconomic status for the perceived impact of state tests on students and teachers were tested using Student's t-test of independent samples. The results did not yield statistically significant differences in the mean scores of participants who taught in low-socioeconomic and high-socioeconomic schools (t = -1.51, df = 121, p = 0.07, one-tailed). Hence, the difference between the values tested is statistically zero indicating the responses among teachers of low-socioeconomic and high-socioeconomic students were similar. Table 19 summarizes the results.

Table 19

Student t-test Results for Perceived Impact on Students and Teachers and SES

Scale	Description	M	Mean Difference	t	df	Sig. (2-tailed)
6	Equal Variances Assumed	30.10	-1.71	-1.51	121	.133
	Equal Variances Not Assumed	31.81	-1.71	-1.51	96.93	.138

N = 123

Summary of the Quantitative Results

Tests of mean differences for all six research questions were performed using ANOVA as well as Student's t-test of independent samples. No statistically significant difference of mean scale scores were found for four of the research questions when scores were compared by school geographical location and socioeconomic status. In essence, teachers' perceptions of the impact of high-stakes testing on curriculum practices, instructional approaches, time spent on school-wide test preparation activities with administrators, and the amount of class time spent on classroom test preparation activities were similar among rural, urban, suburban, high-socioeconomic and low-socioeconomic schools. However, significant differences on mean scale scores were found for research questions 3 and 6. Urban teachers were more likely to teach critical thinking skills than suburban and rural teachers. At the same time, rural teachers perceived students and teachers were more impacted by testing than urban and suburban teachers. However, there were no differences in these perceptions among the teachers from high-socioeconomic and low-socioeconomic schools.

Summary

Chapter 4 describes the statistical procedures employed to analyze the data collected from the surveys. No statistically significant difference of mean scale scores was found for research questions 1, 2, 4, and 5 when scores were compared by school geographical location. However, statistically significant difference of mean scale scores were found for research questions 3 and 6. Post Hoc analysis was performed to identify which geographical regions was the source of the statistically significant differences. Student t-test was performed to test for differences between the teachers from high-socioeconomic and low-socioeconomic schools. No statistically significant differences were found.

The following chapter presents a discussion of the qualitative methods and findings used to elaborate on the quantitative results discussed in this chapter.

CHAPTER 5

PHASE II: QUALITATIVE METHODS AND RESULTS

This chapter describes the methods and qualitative findings from Phase II associated with the central qualitative research question "In what ways do the combination of the school geographical location, the socioeconomic status of students taught, and high-stakes testing influence the classroom instruction of elementary teachers in grades three through five?" A description of the procedures for sample recruitment, sampling procedures, characteristics of the sample, participant descriptions, procedures for data collection, methods of data analysis, verification procedures, and qualitative research findings are presented in this chapter.

Qualitative Sample Recruitment

Due to the nature of the mixed methods sequential explanatory design, the quantitative and qualitative phases were connected during the intermediate stage in which an interview protocol (Appendix H) grounded in the quantitative results was developed, and participants for the second, qualitative phase of the study were chosen (Ivankova et al., 2006). To follow up on the results from the first, quantitative phase of the study, the researcher selected a diverse group of ten teachers out of 123 respondents who completed surveys in the first, quantitative phase of the study to conduct face-to-face interviews. The first participant was interviewed to pilot test the interview protocol. The data from the pilot interview were excluded from the data analysis because the purpose of the pilot

interview was to test the relevance of the interview questions. After the interview protocol was pilot tested, additional probing questions were added to ensure the researcher captured a rich description of the participants' experiences. A total of nine teachers were selected to interview, and the data from these nine interviews were analytically aggregated into themes and subthemes that captured participants' perspectives on the influence of high-stakes testing on classroom teaching practices.

Sampling Procedure

Maximum variation sampling was used to select participants who varied on at least one of the variables tested in the quantitative phase of the study (Lincoln & Guba, 1985). The purpose of maximum variation sampling was to select participants who varied demographically with regards to the factors that would possibly influence the outcomes of the study such as the socioeconomic status of the students taught and the school geographical location. This form of sampling is a form of purposive sampling which is as effective as random sampling because the sample size for this study was less than 30 participants (Onwuegbuzie & Collins, 2007).

The following paragraphs describe the procedure used to choose typical cases and extreme cases for follow-up qualitative analysis. The mean was chosen as the measure of central tendency to determine typical case scores because the survey contained both categorical and ordinal data (Healey, 2002). Because the purpose was to describe the data, the mean was the preferable measure of central tendency even for ordinal-level variables (Healey). The scores from the survey instrument were computed using the summed mean scores for all the participants and for their group means based on SES and location inde-

pendently using SPSS software version 17.0 (Ivankova, 2004). Additionally, the standard error of the mean for socioeconomic status and location were entered independently of each other by one standard deviate.

Once the summed mean and the one standard error of the mean were entered, the typical case scores were identified (Ivankova, 2004). Cross tabulation in SPSS was used to select nine participants whose scores clustered around the mean. Consequently, urban participants whose scores fell into the range of 105.3-110.88 were considered typical cases because these scores were clustered around the mean. Additionally, suburban and urban participants whose scores fell into the range of 118.47-121.15 were considered typical cases because these scores were clustered around the mean. Participants whose scores did not fall into the typical case range were identified as extreme cases because their scores did not cluster around the mean and were more than one standard deviate from the mean. Table 20 summarizes these results.

Table 20

Typical and Extreme Cases

Location	Summed Mean	Standard Error	Range of Scores for Typical Cases	Range of Scores for Extreme Cases
Urban	108.09	2.79	105.30 – 110.88	74.5 – 90.8
Suburban	119.81	1.34	118.47 – 121.15	82.3 – 110.2
Rural	120.45	2.31	118.47 – 121.15	79.7 – 112.8

Moreover, the socioeconomic status of the students taught was determined by the percentage of students receiving free or reduced lunch. Title I schools are defined by the federal government as schools with at least 40% of low income students who receive free or reduced lunch (U.S. Department of Education, 2010). This 40% benchmark was used

to classify schools as low-socioeconomic or high-socioeconomic. Specifically, schools that served at least 40% of the students receiving free or reduced lunch were classified as low socioeconomic, and schools that served less than 40% of students receiving free or reduced lunch were classified as high socioeconomic. A total of 40% of all of the participants in the first, quantitative phase of the study served in low-socioeconomic schools, and 22% of all of the participants in the first, quantitative phase of the study served in high-socioeconomic schools. For the qualitative phase of the study, six participants who served in low-socioeconomic schools and three participants who served in high-socioeconomic schools were selected to follow-up on the results from the first, quantitative phase of the study.

Furthermore, participants were selected from each of the school geographical locations: rural, suburban, and urban. The researcher attempted to obtain a total of three teachers from each of the school geographical locations to interview but was not able to obtain three urban teachers. Consequently, only two urban teachers were available for interviews. As a result of not being able to obtain three urban teachers, four rural teachers were interviewed instead of three because the rural teachers were similar to urban teachers based on the characteristic of the socioeconomic status of the students taught. However, three suburban teachers were available for the interviews.

Characteristics of the Qualitative Sample

A total of nine participants were selected based on responses to the survey administered in the first, quantitative phase of the study. Typical case and extreme case participants were selected. Teachers who taught in high-socioeconomic, low-socioeconomic,

rural, urban, and suburban schools were selected for face-to-face interviews. Moreover, to ensure a variety of perceptions were obtained, teachers who taught grades three, four, and five were selected to interview. The information regarding selected teacher demographics is presented in Table 21. All teachers received a unique pseudonym to protect anonymity.

Table 21

Characteristics of the Qualitative Sample

Participant	Free/Reduced Lunch Percentage	SES	Location	Grade	Case Type
Alpha	96.3%	Low	Urban	5	Typical
Kappa	96.3%	Low	Urban	5	Typical
Gamma	45.8%	Low	Rural	3	Typical
Zeta	45.7%	Low	Rural	5	Typical
Delta	64.1%	Low	Rural	3	Typical
Theta	9.3%	High	Suburban	4	Extreme
Beta	49.6%	Low	Rural	5	Extreme
Eta	13.6%	High	Suburban	3	Extreme
Iota	35.0%	High	Suburban	4	Typical

Note: Maximum Variation Sampling (N = 9)

All the participants were required to sign consent letters prior to completing the survey which was in the first, quantitative phase of the study. Participants were informed in the consent letter that they would possibly be asked to participate in an interview. Furthermore, before the interview began participants were reminded their actual names would be replaced with pseudonyms assigned by the researcher to protect anonymity. The pseudonyms were letters of the Greek alphabet.

Interview Protocol Development

Due to the nature of the sequential explanatory design, the interviews were used as a follow up to the first, quantitative phase to explain the results from this phase of the study in more depth (Teddlie & Tashakkori, 1998). The interview protocol to guide the second, qualitative phase was developed after the data collection and analysis of the first, quantitative phase had been completed. The goal of the qualitative interviews was to increase the depth of the study and to explore how the combination of the school geographical location, the socioeconomic status of students taught, and the SAT-10 and ARMT influence the classroom instruction of elementary teachers in grades three through five.

The interview protocol consisted of 11 open-ended questions that explored specifically how SAT-10 and the ARMT testing had impacted classroom teaching practices in specific subjects and how it had impacted teachers and students (Appendix H). The interview began by asking the participants to tell the researcher about themselves. This first question was posed to make the participants feel at ease and comfortable during the interview and to gain a better understanding of the participants' teaching experience (Hatch, 2002). Probing questions such as why the participant chose teaching, how many years of experience one had with teaching and testing, how long one taught at his or her current school, and what grade level was taught by the participant were asked if the participants did not discuss these demographics when asked to tell about himself or herself.

Four of the six factors in the first, quantitative phase of the study did not yield differences among mean scale scores by school geographical location or the socioeconomic status of the students taught. Thus, question 2 asked participants to describe their experiences with SAT-10 and ARMT testing. This question was asked to obtain a deeper un-

derstanding of teachers' general perceptions of the positive and/or negative experiences with testing and to determine if common themes emerged among teachers in similar socioeconomic and school geographical locations. The probing question, please describe one example of a positive experience with testing and one example of a negative experience with testing were asked if participants did not give specific answers or if clarification was needed.

The first, quantitative phase of the study did not yield differences among mean scale scores by school geographical location or the socioeconomic status of the students taught for curriculum approaches. Hence, question 3 asked the participants to describe any influences SAT-10 and ARMT testing had on the curriculum. This question was posed to determine if teachers placed more priority on the subjects that were tested as opposed to the subjects that were not tested on the high-stakes test and to determine if common themes emerged among teachers in similar socioeconomic and school geographical locations. The probing question asked what influence SAT-10 and ARMT testing had on the teaching of reading, math, social studies, and science. An additional probing question asked what influence testing had on the priority of the subjects taught.

In the first, quantitative phase mean factor score differences by school geographical location for the amount of time spent on critical thinking skills were tested using ANOVA. The analysis yielded significant results. Post Hoc analysis indicated urban teachers spent more time on critical thinking skills than rural and suburban teachers. The Student *t*-test did not yield significant results by socioeconomic status of the students taught. Hence, question 4 asked the participants to describe how SAT-10 and ARMT testing influenced the teaching of critical thinking skills. This question was posed to deter-

mine if teachers increased or decreased the amount of time spent on critical thinking skills as a result of testing. The second purpose of this question was to determine if common themes emerged among teachers in similar socioeconomic and school geographical locations.

Additionally, in the first, quantitative phase of the study mean factor score differences by school geographical location or socioeconomic status did not yield differences for instructional approaches. Question 5 asked the participants to describe how SAT-10 and ARMT testing influenced teaching practices. This question was posed to determine what specific learning activities teachers planned in their classrooms as a result of testing and to determine if common themes emerged among teachers in similar socioeconomic and school geographical locations. The probing question asked teachers specifically what types of learning activities they planned for students as a result of testing.

The first, quantitative phase of the study did not yield differences among mean scale scores influences by high-stakes testing on school geographical location or the socioeconomic status of the students taught for test preparation in the classroom and test preparation activities with administrators. Question 6 asked the participants to describe any influences SAT-10 had on the amount of time spent on test preparation in the classroom and with administrators. This question was posed to determine what specific test preparation activities teachers planned for students, what type of test preparation activities teachers were involved in with administrators, and to determine if common themes emerged among teachers in similar socioeconomic and school geographical locations. A probing question asked the participant to describe examples of the types of test preparation in which students were engaged or in which they were involved in as a teacher.

In the first, quantitative phase of the study, mean factor score differences by school geographical location for the perceived impact of state tests on students and teachers were tested using ANOVA. The results yielded significant differences on mean scale score among the teachers. Post Hoc analysis revealed urban teachers' responses were the lowest. Suburban teachers' responses followed urban teachers' responses, and rural teachers' responses were the highest. This indicates that urban teachers perceived students and teachers were not as greatly impacted by high-stakes testing as opposed to suburban and rural teachers. Rural teachers perceived students and teachers were more greatly impacted by high-stakes testing than urban and suburban teachers. Question 7 asked the participants to describe how testing influenced them as a teacher. This broad question was posed to allow teachers to discuss how testing has influenced them personally in ways they may not have described in previous answers. A second purpose for asking this question was to determine if common themes emerged among teachers in similar socioeconomic and school geographical locations.

Additionally, as a result of the findings that urban teachers were not as greatly impacted by testing as opposed to rural teachers who were more greatly affected, questions 8, 9, and 10 asked how SAT-10 and ARMT testing influenced students, their motivation for learning, and the influence of testing on specific student populations such as special education students, ELLs, and minorities. This question was posed to determine in what ways students were influenced by testing, and if a certain population of students were more influenced than others. A second purpose of this question was to determine if common themes emerged among teachers in similar socioeconomic and school geographical locations.

Question 11 asked if there was any other information the participant would like to share regarding experiences with SAT-10 and ARMT testing. This question was posed to allow the participants to share any perceptions about testing that were not covered by the interview questions. A second purpose was to determine if common themes emerged among teachers in similar socioeconomic and school geographical locations.

Pilot Testing of the Interview Protocol

A pilot test of the interview protocol was conducted by administering the interview to Lambda. Lambda was a third grade teacher who taught in a high-socioeconomic, suburban school. Lambda earned a bachelor's degree in education in addition to a master's degree. Lambda chose the teaching field because she had liked going to school. Lambda had taught for 25 years which included high school, middle school, and elementary school experience and had administered high-stakes testing since testing began.

Before the interview began, Lambda was told this was a pilot test and that the interview would be audio taped and transcribed by a professional transcriptionist. She was asked not to say her name or any other information that would link her to the study. The researcher assured Lambda that her identity would remain anonymous by assigning a pseudonym which was a Greek letter of the alphabet. The pilot interview was administered and completed in approximately 25 min.

As a result of the interview, minor changes were made to the interview protocol to ensure the researcher obtained a rich description of the teachers' experiences with high-stakes testing. Question 2 stated, "What are your experiences with SAT-10 and ARMT testing?" The probing statements "please describe a positive experience of testing, and

please describe a negative experience of testing" were added to question 2. Additionally, question 3 stated, "please describe any influence SAT-10 and ARMT testing has on the curriculum." The probe "how does it influence the priority of the subjects you teach?" was added as a probing question to question 3. Question 5 stated "how do you think SAT-10 and ARMT testing influences your teaching practices. The probe, "how does SAT-10 and ARMT testing influence the types of learning activities you plan for students?" was added to question 5. Additionally, as a result of the pilot interview, participants were asked to give specific examples of their experiences with high-stakes testing whenever responses were not detailed. Participants were informed their perceptions would possibly add to the professional literature of high-stakes testing. Participants were also informed the awareness regarding the impact of high-stakes testing could possibly result in greater understanding regarding the issues facing elementary school teaching practices in Alabama.

Qualitative Data Management

Recording Procedures and Timeframe

The participants for the second, qualitative, phase of the study were contacted in January of 2010. The researcher scheduled interviews at the participants' convenience. These interviews were conducted at the teachers' schools or at places convenient for the participants and lasted approximately 30 min to 40 min. Each interview was audio taped and transcribed by a professional transcriptionist who was hired by the researcher. The participants were informed before the interview began that a professional transcriptionist would hear the audio. Participants were advised not to reveal information such as the

school name or the name of the school principal to ensure confidentiality. All of the interviews were completed in a 2-week timeframe.

Data Analysis

After all nine of the interviews were conducted, the data were transcribed by a professional transcriptionist. Thus, the first step in the data analysis began by listening to the audiotapes of the interviews to ensure accuracy of the transcripts. Once the transcripts were checked for accuracy, the researcher read the text from all of the interviews several times to get a sense of the complete picture of the data (Creswell, 2009). As the data were read, notable quotes were highlighted and memos were recorded in the margins of the transcripts (Glaser & Strauss, 1967) to assist in identifying patterns and relationships in the data. After three re-readings of the first two transcripts, the researcher began coding by assigning labels to text passages. Next, the researcher began to identify patterns and relationships within the data by aggregating similar codes (Hatch, 2002). These data patterns and codes were verified by an outside coder (the committee co-chair person) who was an expert in mixed methods and qualitative research to ensure credibility of the codes.

After insider-outsider coding of the initial interviews was conducted by the researcher and the committee co-chair person (Hatch, 2002), a code book of codes, emergent themes and supporting quotes was developed to aggregate the data and to make comparisons among the data. Then analytic coding using the code book was conducted of the remaining interviews to develop themes and to make comparisons across the data (Bazeley, 2003; Morse & Richards, 2002). NVivo 8 qualitative software was used for da-

ta analysis. As a result of coding, making comparisons among the data, and aggregating the data, four themes emerged: testing drives the curriculum, loss of creativity in teaching, causing stress or anxiety, and promoting inequities. These themes and subthemes were used to describe participants' perceptions of the impact of high-stakes testing on classroom teaching practices in Alabama.

Qualitative Data Quality

Verification procedures were used in the qualitative phase of this study to increase the trustworthiness and credibility of the findings. Lincoln and Guba (1985) introduced the term "trustworthiness" which means the researcher can persuade the audience to believe the research findings are "worth paying attention to" and are "worth taking account of" (Lincoln & Guba, p. 290). Credibility of the researcher's interpretation of the interview data were paramount to establishing trustworthiness of the research results (Teddlie & Tashakkori, 1998). In the qualitative phase of the study, the strategies of member checking, peer debriefing, and thick, rich descriptions (Lincoln & Guba; Maxwell, 2005) were implemented to ensure the credibility and trustworthiness of the findings and are described in the following paragraphs.

Member Checking

To ensure the findings of the study were accurate, each interview participant received a copy of the summary of the interview by mail (Creswell, 2005). Participants were asked if the description of the interview was complete and if the interpretations of their answers to the interview questions were true and correct to help establish the mea-

ningfulness of the study results (Creswell, 2008; Creswell & Miller, 2000). The participants were asked to determine the accuracy of the report in writing and were able to make modifications to the report if they determined it did not accurately depict their views (Creswell, 2005). The participants were able to email or phone the researcher to discuss the modifications. The participants responded through email that the researcher portrayed their perceptions accurately. Thus, no modifications had to be made regarding the accuracy of the participants' perceptions.

Peer Debriefing

This process was used to increase the accuracy of the study (Creswell, 2009). Peer debriefing involves finding a person who will review and will ask questions about the study (Creswell, 2003). This strategy was used as the researcher discussed the progress of the study with a colleague who was an experienced teacher and was cognizant of qualitative research. This colleague was asked to offer suggestions to clarify the ideas and information in the study. Additionally, the researcher followed up on the interviews by email with the participants to ask clarifying questions as needed.

Thick, Rich Description

A clearly delineated description of the participants, the setting, and the themes generated from the study were included. The implementation of thick, rich descriptions allows the reader to transfer findings to similar settings or groups (Creswell, 2009). This gives the reader the perception that he or she actually experience or could possibly experience the events outlined by the researcher (Creswell, 2005; Creswell & Miller, 2000).

The transfer of finding to other settings or groups is termed transferability and is used in qualitative research as opposed to the term generalizabilty which used in quantitative research.

Participants

A total of nine participants based on the survey results of the first, quantitative phase were contacted by email and asked to participate in face-to-face interviews in January of 2010. The interviews were completed in a two week timeframe. The participants' descriptions include the grade level taught by the participant, the socioeconomic status of the students taught, the school geographical location in which the participant taught, educational degrees earned, reasons for choosing the teaching field, years of teaching experience, and years of experience with high-stakes testing which are included in the following section. Assigned numerical codes from the surveys completed by the participants in the first, quantitative phase were replaced with pseudonyms which consisted of letters of the Greek alphabet for the second, qualitative phase of the study to increase the feasibility of participant descriptions and the reporting of the perceptions obtained from the participants.

Participant 1: Beta

Beta was a fifth grade teacher in a low-socioeconomic, rural school. She earned a bachelor's degree in elementary education and at the time of the interview was pursuing a master's degree in elementary education. She chose the teaching field because she had

always known this was what she wanted to do. Beta had 3 years of teaching experience and had administered high-stakes tests for 3 years.

Participant 2: Gamma

Gamma was a third grade teacher in a low-socioeconomic, rural school. She earned a bachelor's and master's degree in elementary education. She chose the teaching field because she loved education and children. She also valued education and learning. She had 4 years of teaching experience and had administered high-stakes tests for 4 years.

Participant 3: Delta

Delta was a third grade teacher in a low-socioeconomic, rural school. Her first career was athletic training. She earned a master's degree in elementary education. She chose the teaching field because she enjoyed being around children. She had 7 years of teaching experience and had administered high-stakes tests for 5 years.

Participant 4: Zeta

Zeta was a fifth grade teacher in a low socioeconomic, rural school. He earned a bachelor's degree in elementary education. He chose the teaching field because after working with children in a church he discovered he had a talent teaching children. He had 13 years of teaching experience and had administered high-stakes tests for 13 years.

Participant 5: Eta

Eta was a third grade teacher in a high-socioeconomic, suburban school. She earned a bachelor's degree in elementary education and a master's degree in educational administration. She chose the teaching field because she enjoyed watching children learn and enjoyed being a part of their lives. She had 11 years of teaching experience and had administered high-stakes tests for 9 years.

Participant 6: Theta

Theta was a fourth grade teacher in a high-socioeconomic, suburban school. She earned a bachelor's degree and a master's degree in elementary education. She chose the teaching field because she had two brothers who had learning disabilities, and they hated school. Consequently, she wanted to work with children who struggled and to prove to her brothers that "not all school was bad." She had 9 years of teaching experience and had administered high-stakes tests for 9 years.

Participant 7: Iota

Iota was a fourth grade teacher in a high-socioeconomic, suburban school. She earned a bachelor's degree in elementary education. She chose the teaching field because she admired her teachers and their work. She had 5 years of teaching experience and had administered high-stakes tests for 5 years.

Participant 8: Alpha

Alpha was a fifth grade teacher in a low-socioeconomic, urban school. She earned a bachelor's degree in elementary education and a master's degree in educational administration. She chose the teaching field because she came from a family of educators. Her mother was a teacher, and she was encouraged to teach after watching her mother. She had 13 years of experience and had administered high-stakes tests for 13 years.

Participant 9: Kappa

Kappa was a third grade teacher in a low-socioeconomic, urban school. She earned a bachelor's degree in elementary education. She chose the teaching field because it was a life-long desire. She came from a family of educators and had a passion to teach. She had 4 years of teaching experience and administered high-stakes tests for 4 years.

Qualitative Research Findings

The goal of this section is to provide interpretation of the qualitative themes and subthemes that emerged from the thematic analysis of the interview data from teachers. The factors addressed in the quantitative phase of the study were curriculum approaches, instructional approaches, critical thinking skills, school-wide and classroom test preparation activities, and the influence of high-stakes testing on students and teachers. The following themes emerged from discussions with teachers that help explain the role of quantitative factors tested in the first phase: testing drives the curriculum, loss of creativity in teaching, causing stress or anxiety, and promoting inequities. The themes and related subthemes are presented in Table 22.

Theme: Testing Drives the Curriculum

"Testing drives the curriculum" was the theme which emerged through data analysis and interpretation to answer the research subquestion "What influence do teachers of grades three through five perceive high-stakes testing has on the curriculum?" The four subthemes associated with this theme were as follows: imbalances in the content areas, breadth versus depth, changes in critical thinking skills, and teaching to the test.

Table 22

Themes and Subthemes

Themes	Subthemes
Testing drives the curriculum	Imbalances in the content areas
	Breadth versus depth
	Changes in teaching of critical thinking skills
	Teaching to the test
Loss of creativity in teaching	Loss of teachable moment
	Teaching becomes robotic
	Rigid scheduling
	Loss of student motivation for learning
Causing stress or anxiety	Teacher anxiety over student performance
	Pressure on teachers from administrators for AYP scores
	Student anxiety over their performance
	Teacher burnout
Promoting inequities	Inadequate learning resources
	Inadequate parental support
	Special needs children
	ELLs

Discussions disclosed all teachers, independent of socioeconomic status and school geographical location, placed more priority on reading and math. Additionally, three of the four low-socioeconomic, rural teachers and one high-socioeconomic, suburban teacher revealed they had to cover a vast amount of information instead of going into

depth with concepts. Furthermore, most teachers, independent of socioeconomic status and school geographical location, reported they decreased the teaching of critical thinking skills due to SAT-10 testing. In contrast, two low-socioeconomic teachers who taught in a rural and an urban school noted SAT-10 testing promoted an increase in the teaching of critical thinking skills. Furthermore, with regards to ARMT testing, most teachers independent of socioeconomic status and school geographical location believed they increased the teaching of critical thinking skills, while two teachers mentioned they decreased the teaching of critical thinking. Finally, with the exception of two high-socioeconomic, suburban teachers all of the other participants perceived they were "teaching to the test" due to high-stakes testing and viewed this as a negative influence. Each subtheme is discussed in depth in the following paragraphs, respectively, along with a summary of the theme.

Imbalances in Content Areas

Teachers were asked what influence SAT-10 and ARMT testing had on the curriculum. Comments made by all of the participants, independent of the socioeconomic status or school geographical location, revealed they placed more priority on math and reading. Subjects such as social studies and science received less precedence as noted by all of the participants.

Due to the emphasis of testing in reading and math, most of the teachers revealed they decreased the amount of time spent on science and/or social studies. For example, Delta who taught in a low-socioeconomic, rural school shared that she had to be sure she focused on reading and math daily because these two subjects were a necessary for mak-

ing AYP for high-stakes test scores. Whenever there were time restrictions and a subject had to be eliminated, Delta stated, “it was going to be science or social studies.” The amount of teaching of science and social studies decreased because the effectiveness of schools was determined by the scores students made on reading and math as opposed to social studies and science. Similarly, Iota and Eta who both taught in high-socioeconomic, suburban schools agreed they decreased the amount of time spent on social studies and science because, “it was not tested.” Zeta who taught in a low-socioeconomic, rural school also agreed he placed more priority on reading and math, but in contrast to other teachers, he did not decrease the amount of time he spent on any of the subjects including social studies and science. Zeta was the only participant who did not decrease the teaching of social studies and science. Zeta remarked, “I tried hard to hold true to teaching all subjects and not slacking on social studies.” He expressed that he did not let anything deter him from teaching how he believed all subjects should be taught.

Due to the administration of the Alabama Science Assessment (ASA), which was not a high-stakes test but rather a state mandated test for fifth grade students, some teachers included aspects of science in the curriculum. Beta who taught in a low-socio economic, rural school remarked she was able to teach science because her fifth graders were required to take the ASA. Although she taught science, she perceived it was rushed and mentioned “it still got the back burner” because of the emphasis placed on the SAT-10 and ARMT which both assessed reading and math skills. The SAT-10 and the ARMT took precedence over the ASA because it was not mandated by the NCLB enacted by the federal government and because no sanctions were attached for low test scores on the

ASA. Delta, a third grade teacher from a low-socioeconomic, rural school also stated she spent more time on science than social studies. More attention was given to science because of the ASA that was administered to the fifth graders at her school. For example, in February they had an experiment each week that each grade level would broadcast over the television for the entire school to see. Theta, a fourth-grade teacher from a high-socioeconomic, suburban school mentioned reading and math dominated the curriculum as well. However, she also stated the fifth-grade teachers at her school taught science more because it was tested on the ASA.

Similarly, third-grade and fourth-grade teachers who were employed at AMSTI schools were teaching more science than those who were not employed at AMSTI schools. For instance, Gamma who taught in a low-socioeconomic, rural school was able to teach science only because the school where she taught was an AMSTI school, and "teachers were required to teach science 45 min a day." However, she expressed some students were not learning all of the science and social studies as the state required because of the intervention she had to place on the high-stakes test.

Two teachers who taught in different contexts mentioned they alternated the teaching of social studies and science. Eta who taught in a high-socioeconomic, suburban school had forty minutes on her schedule to teach science and social studies. However, due to recess and students having to spend time to pack up belongings to go home for the day she was not able to devote the entire allotted time to these subjects. Eta said, "I really only had about 15 min of science and social studies every day." Hence, she was only able to dedicate half the allotted time to these subjects. She alternated the weeks she taught science and social studies. For example, she taught science for 2 or 3 weeks and after-

wards taught social studies for 2 or 3 weeks. Comparatively, Kappa who taught in a low-socioeconomic, urban school also stated if she had time she would teach social studies one day and science the next alternating between the two subjects as evident in her quote, “I throw in social studies one day and then science the next.”

A few teachers from different socioeconomic statuses and school geographical locations, mentioned they were integrating social studies and science in other areas. Alpha said, “even if I’m short-cutting social studies, I’m implementing it in reading.” She considered this integration as an effective way to teach social studies. However, Gamma who taught in a low-socioeconomic, rural school was also teaching social studies during her reading time, but she considered this to be ineffective because she was expected to give intervention to groups of students who were weak in reading and math and was also expected to teach a whole group lesson during that time. Due to these reasons, she remarked, “it’s really hard to get social studies in.” Eta who taught in a high-socioeconomic, suburban school said that although she only had about fifteen minutes for social studies and science she had been integrating these subjects into other areas of the curriculum. However, she admitted she was less focused on these two subjects because they were not tested in the grade level she taught.

The teaching of social studies was even more limited than science among all of the participants independent of socioeconomic status and school geographical location. Theta who taught in a high-socioeconomic, urban school agreed social studies related courses received the least amount of priority. Subjects that were tested took precedence and as mentioned by Theta, “Alabama history was pretty much thrown to the wayside.” Beta and Gamma who both taught in low-socioeconomic, rural schools deemed social

studies as important but just did not have class time to dedicate to this subject. Beta believed students were losing knowledge in these areas and felt social studies was given even less priority than science. Beta expressed this belief when she said it was important for students "to know where they came from." She, as Gamma previously stated, was not able to effectively teach social studies because her school system suggested it be taught during intervention time. Gamma believed the limitation of teaching social studies may not have a major impact on students right now. Elementary students had been able to function throughout elementary school without teaching much social studies, but it was important because students at the middle and high school level were required to take social studies and science classes. Consequently, the researcher inferred Gamma was concerned when elementary students would enter the middle and high school, they would lack background knowledge to function in social studies and science classes at the middle and high school levels. Gamma predicted it would be "detrimental to the students when they got to the middle and high school."

In summary, the teachers, independent of socioeconomic status and school geographical location, disclosed in their discussions that not all subjects received equal priority. Teachers revealed there were imbalances in the content area. All of the teachers posited they placed more priority on reading and math due to high-stakes testing. However, other subjects such as social studies and science were taught only if they had time to attend to them during the day. For all of the participants who taught fifth grade and/or were employed at an AMSTI school, science was taught more often than social studies. For the most part, the teaching of subjects such as social studies and science were limited thereby narrowing the elementary curriculum.

Breadth versus Depth

Teachers were asked how SAT-10 and ARMT testing influenced their teaching practices. Comments made by the participants varied. Discussions disclosed two of the teachers who taught in low-socioeconomic, rural schools believed they were teaching a broad amount of skills as opposed to going into depth with the concepts they taught. One teacher who taught in a high-socioeconomic, suburban school revealed she also had to teach a broad amount of skills as opposed to going into depth with the concepts she taught. On the opposite, both Theta who taught in a high-socioeconomic, suburban area and Zeta who taught in a low-socioeconomic area perceived testing did not have an influence on their teaching practices.

Beta and Eta remarked they had to cover a vast amount of content instead of being able to go into depth with concepts they taught. Beta who taught in a low-socio economic, rural school said she was not able to go deeper into concepts because she had curriculum guides to follow, and she had to be sure she covered everything on the test. She disagreed with the guideline of having to administer high-stakes tests in late March or early April because teachers still had 9 more weeks of school left. Due to the time of the administration of the test, she had to teach all of the objectives for the year before the test. As a result, the students were put at a disadvantage because they had to learn many concepts very quickly. In reference to reading instruction, Eta who taught in a high-socioeconomic, suburban school also emphasized she had to teach everything before the test. Eta remarked she just "kind of scratched the surface," and then had to move on to another comprehension skill. Consequently, Eta also tried to cover all of the objectives on the high-stakes tests instead of going into more depth about what she was teaching. For

example, Eta was not able to have read aloud time and talk about the characters in the story due to the requirements of testing because students just had to be able to choose the correct answer in the text as tested on the high-stakes tests. Delta who taught third grade in a low-socioeconomic, rural school believed she taught a broad amount of skills instead of going into depth because she had to rush all of the time; there were numerous objectives that she had to teach before the high-stakes test was administered. Specifically, she believed the special education students got left behind because of the "move on" mentality. These third grade students did not have a good foundation and were still counting on their fingers to add. She wondered how "they" expected students below grade level to multiply, but she had to continue teaching new concepts because there were so many objectives that had to be taught before the high-stakes test was administered. In contrast to the experiences of other participants, Kappa from a low-socioeconomic, urban school believed she "went into depth in reading." Kappa believed she taught reading in depth because she usually assigned four activities with the skill or story she taught each day. Even after probing, Kappa did not offer more reasons for why she believed she went into depth versus breadth.

Although some teachers believed testing influenced their teaching practices, others perceived testing really had not changed how they taught. For example, Theta who taught in a high-socioeconomic, suburban school thought technology had changed teaching more so than anything else. However, she mentioned she only taught what was on the test and not any extra material because there were so many math concepts to be taught in fourth grade. Consequently, she had to cover a vast amount of material, but she did not see this as a concern and did not mention that she was not able to go into depth with

teaching it. Similarly, Zeta who taught in a low-socioeconomic, rural school believed testing did not cause him to change how he believed all subjects should be taught. He said, “I basically, honestly continue to flow the way I always do.” Zeta admitted he placed more emphasis on reading and math, but he continued to do what he thought was best for children, and testing did not influence his teaching practices.

In summary, teaching a broad amount of skills as opposed to being able to go into depth with concepts was perceived as a dilemma by several teachers independent of socioeconomic status and school geographical location, but not all participants agreed because some teachers who were from different socioeconomic and geographical school locations believed testing did not influence their teaching practices. The teachers who believed testing influenced teaching practices perceived students did not understand the concepts because they had to cover such a vast amount of material before the test, and the students did not have time to grasp a concept before having to go on to a new concept.

Changes in Critical Thinking Skills

Teachers were asked how SAT-10 and ARMT testing influenced the teaching of critical thinking skills. Comments made by the participants revealed teachers varied in their opinions how high-stakes testing influenced teaching practices. The variation of responses was independent of socioeconomic status and school geographical location.

For instance, Gamma who taught in a low-socioeconomic, rural area and Theta who taught in a high-socioeconomic, suburban area both believed the SAT-10 did not help students develop critical-thinking abilities. Specifically, Gamma mentioned that SAT-10 “worked against” critical thinking skills because students were required to an-

swer straight-forward questions. Theta posited when she first started teaching, she did not teach a lot of critical thinking skills because the SAT-10 was more fact-based.

In contrast, Beta who taught in a low-socioeconomic, rural area and Eta who taught in a high-socioeconomic, suburban area both believed neither the SAT-10 or the ARMT allowed the students to be able to think critically because students just needed to know how to pick an answer to a question. Beta clarified students had to be able to think critically in some ways, but students were not engaged in the higher levels of thinking on Bloom's Taxonomy. Consequently, she did not have the opportunity to teach critical thinking skills needed for real life because she was not able to teach her students how to "think in the world around them," but instead she had to teach test taking strategies such as "go back and find the answer." Eta also remarked both of these tests had decreased the teaching of critical thinking skills, because in the past she taught most of her critical thinking and analytical skills during read aloud time and by giving the students time to talk about the characters. However, because of testing she no longer had time to read aloud to her students and discuss the characters in the story thereby not allowing her time to teach the critical thinking skills students needed.

Most teachers independent of socioeconomic status and school geographical location believed ARMT increased their teaching of critical thinking skills. For instance, Gamma who taught in a low-socioeconomic, rural area reported she taught more problem-solving, higher order thinking lessons due to ARMT testing versus teaching the students test taking strategies such as identifying distracters with the SAT-10. Theta who taught in a high-socioeconomic, suburban area noted students have to think when they write. Consequently, Theta viewed this as a positive influence on the curriculum because

the writing portion of the ARMT encouraged students to be creative by "thinking outside of the box" and writing more than one step answers. Alpha who taught in a low-socioeconomic, urban area mentioned ARMT testing gave students the opportunity to express themselves more because it was not just centered around multiple choice questions. She believed it showed all arrays of testing because the students had to respond to open-ended questions.

Some teachers, on the opposite, independent of school geographical location believed both of these tests increased critical thinking skills. Zeta who taught in a low-socioeconomic, rural school, Iota who taught in a high-socioeconomic, suburban school, and Kappa who taught in a low-socioeconomic, urban school all noted ARMT and SAT-10 testing promoted the increase of critical thinking skills. As a result of testing, Iota who taught in a high-socioeconomic, suburban school offered her students unique situations and opportunities to use the strategies she taught them to be prepared for the SAT-10 and the ARMT. Iota specified that she did not just use worksheets to help her students prepare for the high-stakes tests. Instead, she used open-ended activities to help the students practice their critical thinking skills such as having students read a variety of texts and respond to the texts using post-it notes while they read. Specifically Iota remarked, "instead of just having all rote memorization kind of activities where the kids are required to regurgitate information, I give them unique situations and opportunities to do open-ended activities." Kappa who taught in a low-socioeconomic, rural school thought both the SAT-10 and the ARMT influenced teachers to teach critical thinking skills because students had to "stop, think, and apply" all the skills they knew to analyze a variety of in-

formation and to come up with a reasonable conclusion. Kappa remarked, “testing heightens all of those critical skills that you have taught throughout the year.”

In summary, most teachers independent of socioeconomic status and school geographical location believed they increased their teaching of critical thinking skills as a result of ARMT testing due to the open-ended questions students must respond to on this test because they had to think about their answers rather than just pick out a provided multiple-choice response. However, two teachers did not believe ARMT increased critical thinking skills. In contrast, the comments regarding the SAT-10 varied among the participants independent of socioeconomic status and school geographical location. Some teachers believed SAT-10 decreased their teaching of critical thinking skills whereas other teachers believed SAT-10 increased their teaching of critical thinking skills.

Teaching to the Test

Teachers were asked how testing influenced the type of learning activities they planned for students and the amount of time spent on test preparation in the classroom and in discussion with administrators. Comments revealed most of the teachers, independent of socioeconomic status and school geographical location, perceived they were “teaching to the test.” The teaching to the test consisted of using test preparation activities to prepare students for the test. These test preparation activities included using test preparation booklets to expose students to items similar to the ones on the tests. On the opposite, two high-socioeconomic, suburban teachers reported they did not teach to the test because they did not increase the amount of test preparation activities they used in their classrooms.

Comparatively, three of the four the low-socioeconomic, rural teachers believed they had to teach to the test as presented in the following comments. Only one teacher from a high-socioeconomic, suburban school mentioned she had to teach to the test. Beta who taught in a low-socioeconomic, rural school reported she used the test preparation materials such as the *Alabama Coach Book* and *Old Test Ready* books as well as online Powerpoints that had been created for these purposes. As Beta remarked about the use of test preparation materials, "there are tons of resources out there for testing because we do teach the test." Delta who also taught in a low-socioeconomic, rural school also planned test preparation activities for her students. For instance, 15 weeks before testing, the students received an ARMT type question Monday through Thursday. Delta also taught test taking strategies such as how to answer a question before looking at the answer choices, making students knowledgeable about the distracter, and knowing the importance of using scratch paper. Specifically, with regards to scratch paper Delta said, "I'm trying to show them that the reason that I need them to do that is because especially in math, it is so easy to choose the distracter. It is the difference in what place a digit is in." A daily review the school system provided was also used to help prepare Delta's students for the high-stakes tests. Gamma who also taught in a low-socioeconomic, rural school posited testing influenced how she planned her lessons and how she taught. She gave the students ARMT style questions to complete and practice questions similar to the format of the SAT-10. In addition to test preparation materials, Eta revealed she constructed her classroom assessments similar to the high-stakes test. Eta who taught in a high-socioeconomic, suburban school mentioned that she engaged in this practice "so the students felt comfortable when they saw the SAT-10 and the ARMT."

Another test preparation activity used by most of the teachers was the Discovery Education Assessment. The Discovery Education Assessment consisted of items similar to the ones that were tested on the SAT-10 and the ARMT. The Discovery Education Assessment was administered three times a year in the months of September, December, and February in one particular school system before high-stakes tests were administered in late March or early April. The primary purpose of administering this test was to determine students' areas of strengths and weaknesses on specific objectives on the high-stakes tests and to remediate those weaknesses before the high-stakes tests were administered. One low-socioeconomic, rural teacher had positive views about this test preparation activity whereas a high-socioeconomic teacher viewed this test preparation activity negatively. Zeta who taught in a low-socioeconomic, rural school was required by his school system to use the Discovery Education Assessment program. Within this program, Zeta provided small group intervention to the students who showed weaknesses on the test. Zeta viewed this as a positive influence because it "brought out some good things such as diagnosing some problems that the students were having." Other teachers such as Iota used this test preparation strategy to identify students' weaknesses and to remediate those weaknesses. However, Iota viewed this as a negative influence. She viewed this test preparation as teaching to the test. As she stated, "Even though we are not teaching to test, we are because that's the end result."

The teachers who used a lot of test preparation expressed the concern that there was contradiction between what they taught and what they believed needed to be taught. They believed they had to do a lot of test preparation because the students needed it. Beta who taught in a low-socioeconomic, rural school mentioned she had to mainly focus on

the objectives of the test instead of being able to do certain activities from which the students would benefit. Beta wanted her students to be engaged in more open activities such as literature circles and plays in reading. However, testing placed restrictions on what she taught. Specifically, she said, “the test doesn’t teach that, so I don’t need to teach it either.” As discussed earlier, Eta who taught in a high-socioeconomic, suburban school did not have time to read aloud stories to her students and discuss the characters to teach critical thinking skills. Additionally, Alpha who taught in a low-socioeconomic, urban school also mentioned there was a discrepancy between what was taught and what teachers believed should be taught. Alpha remarked that she and other teachers were “teaching to the test” rather than “actually getting in there to teach.”

Two high-socioeconomic, suburban teachers had similar experiences with test preparation. Iota and Theta both believed they did not spend a lot of time on test preparation activities. Although Iota believed she was teaching to the test because of the use of the Discovery Education Assessment, she reported she did not spend a lot of time on activities such as reading a passage and answering multiple choice questions as required by the SAT-10. Iota did however use reading passages regularly that necessitated students to answer open-ended questions as required on the ARMT. Similarly, Theta did not spend a lot of time on test preparation activities. Theta emphasized that she did not stress the students out about test preparation, but she reviewed how to take the ARMT every couple of weeks. The students practiced bubbling and writing during intervention time “so there was not panic the week before the test.”

Some positive influences of testing did emerge out of an overwhelming amount of negative influences. Alpha was the only teacher who said testing influenced her to attend

various workshops and various forms of professional development, and she attributed this to testing. Also due to testing, she taught concepts in a variety of ways instead of teaching it one way as she had been taught in school.

All the teachers independent of socioeconomic status and school geographical location engaged in test preparation strategies with their administrators. Most of the participants mentioned the time they spent with administrators regarding testing was during faculty meetings discussing overall school-wide data. In addition to the faculty meetings, Gamma who taught in a low-socioeconomic, rural school and the other teachers at her school were required to do "sit and chats" once or twice a month. The "sit and chats" consisted of talking to the administrators about students' progress and sharing ideas and feedback on whatever occurred in her classroom. Similarly, Delta who taught in a low-socioeconomic, rural school mentioned she met with the assistant principal and other teachers in her grade level during a professional learning community to discuss ways to increase student achievement. Therefore, Delta's administrators were actually involved in planning lessons instead of just discussing test data at faculty meetings. Eta who taught in a high-socioeconomic, suburban school said if there were meetings other than faculty meetings and an administrator was present, the focus was on testing. Specifically, Eta mentioned with regards to meetings, "it was almost always directly related to the standardized tests." Consequently, high-stakes testing was the focus of most of the meetings in which teachers participated.

In summary, discussions with teachers revealed they focused instruction to prepare students for high-stakes tests. The focused instruction was mentioned by one rural, one suburban, and one urban participant thereby indicating this view was independent of

socioeconomic status and school geographical location. As a result, these teachers perceived they were not able to teach what the students really needed. However, other teachers' views were in stark contrast. The teachers who stated their teaching practices were not influenced by testing did not allow testing to conflict with how they believed students should be taught. Furthermore, most of the teachers independent of socioeconomic status and school geographical location reported they participated in various meetings with their administrators, and the main focus of the meetings was high-stakes testing. However, two low-socioeconomic, rural teachers reported their administrators did more than discuss test data in faculty meetings. One of these teachers mentioned she had to participate in sessions in which she discussed the progress of her students and what she could do to increase their progress. The other teacher mentioned the administrators at her school were actually involved in the planning of reading and math lessons.

Summary of the Theme

In essence, all of the teachers independent of socioeconomic status and school geographical location noted high-stakes testing "drove the curriculum." Testing influenced which subjects teachers spent most of their time teaching. Consequently, reading and math dominated the curriculum while social studies and science received a limited amount of time thereby narrowing the curriculum. Additionally, three of the four teachers who taught in low-socioeconomic, rural areas and one teacher who taught in a high-socioeconomic school reported spending a significant amount of time covering the objectives that were tested instead of going into more depth with the concepts taught. Furthermore, there were inconsistencies in responses regarding the teaching of critical thinking

skills under the influence of the SAT-10. Some teachers believed the SAT-10 led to a decrease in the teaching of critical thinking skills. Yet, other teachers believed the teaching of critical thinking skills increased. With regards to the ARMT, most of the teachers independent of socioeconomic status and school geographical location believed ARMT testing led to an increase of critical thinking skills due to the use of open-ended questions on the test. However, two teachers independent of socioeconomic status and school geographical location did not believe ARMT testing led to an increase of critical thinking skills. The increase of critical thinking skills as mentioned by most of the participants due to ARMT testing was one of the few positive aspects that teachers noted about testing. Finally, three of the four low-socioeconomic, rural teachers believed they were "teaching to the test," and these teachers believed this restricted them from being able to teach students other skills and concepts they believed should be taught. Finally, all of the teachers independent of socioeconomic status and school geographical location reported they participated in various meetings with administrators to discuss high-stakes testing, and that this was the main focus of most of their meetings.

Theme: Loss of Creativity in Teaching

"Loss of creativity in teaching" was the theme which emerged through analysis and interpretation of the interview data to answer the research sub-question: "What influence do teachers of grades three through five perceive high-stakes testing has on instructional practices?" The four subthemes associated with this theme were as follows: loss of teachable moment, teaching becomes robotic, rigid scheduling, and loss of student motivation for student learning.

Discussions revealed teachers who taught in low-socioeconomic, rural areas expressed concerns about losing teachable moments as opposed to high-socioeconomic and other school geographical areas. Other participants did not address the loss of a teachable moment as a concern. However, some teachers independent of socioeconomic status and school geographical location discussed that "teaching becomes robotic," that rigid schedules were in place, and some discussed a loss of student motivation for student learning. The comments regarding these subthemes varied among the participants and are discussed in depth in the following paragraphs respectively along with a theme summary.

Loss of Teachable Moment

Teachers were asked how SAT-10 and ARMT testing influenced teaching practices. Comments made by the participants revealed low-socioeconomic, rural teachers expressed concerns about losing a teachable moment. Three of the four teachers who taught in low-socioeconomic rural areas believed testing contributed to a lack of teachable moments. Beta, Zeta, and Delta who were teachers in low-socioeconomic, rural schools pointed out they were not able to use teachable moments in their classrooms. However, some teachers independent of socioeconomic status or school geographical location including Gamma from a low-socioeconomic, rural area did not mention "loss of teachable moment" as a concern.

For example, Beta who was a teacher in a low-socioeconomic, rural school posited she was not able to use teachable moments because of the emphasis on testing due to having to teach the test. Beta believed because she was pressured to teach the test she did not have time to teach students concepts she would have enjoyed teaching. Similarly,

Zeta who also was a teacher in a low-socioeconomic, rural school mentioned he was not able to use teachable moments as much anymore. Zeta believed the lack of teachable moments occurred because his school system required teachers to teach the same way. Zeta believed this "scripted way" of teaching was mandated by the school system because it was "data-proven." Specifically, because it was data-proven students were more likely to score higher on the high-stakes tests. Delta who was a teacher in a low-socioeconomic, rural school mentioned she was not satisfied with the way she had to teach. Specifically, Delta remarked with regards to testing "I don't think that it has changed my philosophy or beliefs about teaching, but what I do think is when I leave school I don't feel like I was the teacher that I want to be." Due to testing, she believed she had to get students prepared for the test instead of listening to things that were important to them such as stories they had to share. Specifically, Delta mentioned, "I don't feel like when my kids come up to me and they want to share stories with me that I can hear those stories. I feel like I have to forget that they are kids, and we have to get them prepared for the test." As a result, this lack of opportunity to listen to students discuss things that were important to them decreased the opportunity for a teachable moment. Delta compared how she taught in third grade to how she taught in second grade. Delta believed in second grade she had more freedom and was able to teach the entire year. Delta also believed her beliefs about how students should be taught were present, but she noted her beliefs were not represented in the classroom "the way she wanted them to be." In essence, there is a discrepancy between how Delta taught and how she believed she should have taught.

In summary, three of the four teachers who taught in low-socioeconomic, rural areas in one school system believed they were not able to use teachable moments due to testing. This lack of teachable moments was attributed to having to teach to the test, being required by the school system to teach "scripted" lessons, and not having the time to listen to students share things with the teacher that were important to them. Other participants independent of socioeconomic status or school geographical location did not mention lack of teachable moments as a concern.

Teaching Becomes Robotic

Teachers were asked how SAT-10 and ARMT testing influenced teaching practices. Comments made by the participants revealed three of the four low-socioeconomic, rural teachers mentioned teachers were expected to teach the same way. Also, one suburban teacher mentioned this as well. The other participants independent of socioeconomic status and school geographical location did not disclose information regarding this subtheme.

Conversations with Beta revealed she was not able to be creative in her teaching due to the mandates of testing. Specifically, Beta said, "It forces me to not be able to think outside the box as much as I was taught to and as much as I want to." Beta believed she was not able to engage the students in activities they would enjoy because students had to know how to take the test. Beta reported she would have liked to provide the students with opportunities to act out plays and create different endings to plays, but she said the test did not require this. Hence, she did not teach those things in her class.

Gamma who taught in a low-socioeconomic, rural school believed there were many regulations and restrictions placed on teachers. She believed everyone was expected to teach the same way. Therefore, Gamma believed she was not allowed to teach how she believed students should be taught. Gamma remarked, “Teachers were not allowed to teach how they feel like they should teach their students. Every class and student is different, but we’re all expected to do the same thing.” Gamma considered this loss of creativity a negative consequence of testing.

Zeta who taught in a low-socioeconomic, rural school posited the school system in which he taught mandated a “scripted way” of teaching reading as mentioned previously. So he also believed there was a loss of creativity in teaching due to testing. Zeta believed teachers should have been able to have more input because he said the students at his school have been taught the same lesson two or three years now using the same stories. Iota who taught in a high-socioeconomic, suburban school in the same school system as Zeta also believed teachers were expected to do the same thing instead of being able to be creative. Iota remarked, “I think teaching has gotten very robotic. I think teachers are not allowed to put their own two cents in as much anymore.” She attributed this decrease in teacher creativity to testing and to teaching to the test.

In summary, three of the four teachers in low-socioeconomic, rural schools and one teacher who taught in a high-socioeconomic school perceived they were not able to be creative and teach the way they would like due to testing. They all believed teachers were expected to teach the same way although their teaching contexts were different. All of the teachers believed there should be more opportunities for them to have more input into learning activities in which students participate.

Rigid Scheduling

Teachers were asked how SAT-10 and ARMT testing influenced their teaching practices. Comments made by the participants revealed two low-socioeconomic, rural teachers, one high-socioeconomic, suburban teacher, and one low-socioeconomic, urban teacher believed testing resulted in rigid scheduling. In essence, low-socioeconomic teachers were more likely to discuss rigid scheduling as a concern. Although, one high-socioeconomic, suburban teacher was concerned about rigid scheduling she mentioned her experiences with testing were not as severe as the experiences of her friends who taught in low-socioeconomic areas. Other participants independent of socioeconomic status and school geographical location did not discuss rigid scheduling as being a problem.

Delta who taught in a low-socioeconomic, rural school remarked she taught Math Investigations which involved the use of manipulatives and learning how to solve problems in different ways. She considered this a positive influence on learning. However, she lacked the time to teach Math Investigations the way she thought it should be taught. Delta believed because she had to teach numerous concepts in such a short period of time before testing, she did not have ample time to teach the concepts she would like for the students to have learned. She also contributed this lack of time to the "time frame" mandated by the school system. For instance, Delta observed, "They give us a time frame, like right now I should be on Unit 5, session 2.6, well you know, my kids may not be ready for Unit 5, session 2.6. We may still be on Unit 3, but that is not okay. If they come in my classroom I better be where they say I am supposed to be." The researcher inferred "they" were the administrators at her school. In essence, teachers were expected to teach certain concepts on specific days. Delta believed that her students were not prepared for

the next lesson and needed to spend more time on certain concepts. She referred to this as "short changing" the students. Delta believed this "time restraint" negatively influenced student learning because she was not able to meet their needs and had to teach the next skill whether or not the students were academically prepared for it.

Similarly, Zeta who taught in a low-socioeconomic, rural school also mentioned time restraints. In the school system in which he taught, teachers were given pacing guides which outlined what objective they were to teach and when they were to teach it. In contrast to Delta, Zeta admitted he did not always follow the pacing guide. Instead, sometimes he sequenced the objectives based on what the students needed. Zeta mentioned that he had to teach what the students were able to do at the time. Specifically, he did not teach a concept the students did not have the skills to learn.

Similarly, Eta who taught in a high-socioeconomic, suburban school mentioned she was allowed to arrange her schedule before the implementation of the ARMT. However, after the implementation of the ARMT she had been given a schedule to follow at the beginning of the year. The administrators dictated when she would teach math, reading, social studies, and science. She attributed these time restraints and rigid schedules to the fact that administrators were pressured and nervous about the performance of the students on the high-stakes tests. Although there were time restraints and rigid scheduling at her school she noted she mentioned she was thankful she taught at a high-socioeconomic school because her friends who taught in low-socioeconomic schools had it much worse. In other words, teachers at low-socioeconomic schools were under much more pressure due to testing than she was at her school.

Alpha who taught in a low-socioeconomic, urban school also mentioned time as a concern. She spent time before school and after school tutoring students on different objectives. Alpha even tutored students during her planning time to remediate those who were having difficulty with a particular concept. Alpha believed there was not enough time to remediate the students who needed help.

In summary, one teacher from each socioeconomic status and school geographical location revealed testing resulted in rigid scheduling. Pacing guides, class schedules, and lack of time for student remediation were concerns expressed by the participants. Although Eta who taught in a high-socioeconomic, suburban school stated rigid scheduling was a negative influence of testing, Eta mentioned she was "fortunate" compared to her friends who taught in low-socioeconomic schools. The other teachers independent of socioeconomic status or school geographical location did not specifically mention scheduling as a concern.

Loss of Student Motivation for Learning

Teachers were asked how testing influenced their student's motivation for learning. Comments made by most of the participants revealed students were not motivated by testing. These views were independent of socioeconomic status and school geographical location. Two opposing views made by teachers who taught in low-socioeconomic, rural areas revealed testing motivated high-achieving students but not students who struggled academically.

Beta who taught in a low-socioeconomic, rural school addressed the concern that students demonstrated a loss of motivation for learning. Beta believed students were not

motivated to learn and did not possess a “love of learning” because testing was not something the students enjoyed. Beta responded, students “need to be able to enjoy what they are learning, and sometimes I cannot make it enjoyable.” Beta noted she was not able to make learning enjoyable because she had to “teach the test.”

On the opposite, Gamma who taught in a low-socioeconomic, rural school believed testing motivated some students to learn but not others. She said testing motivated the high-achieving students to “try harder.” On the opposite, Gamma believed the students who struggled were not as motivated to learn. Specifically, she remarked “the students who already do well testing reinforces them to try harder, but for the students who think it is really hard they just end up giving up.” On the contrary, one of Gamma’s students did not do very well on the Discovery Education Assessment, which is a test preparation activity, to determine students’ strengths and weaknesses. She knew this student was “great” in math. However, because there were 40 questions on the test, he became tired and did not “feel like doing it.” So he did not score well on the test. Similarly, Delta who also taught in a low-socioeconomic, rural school agreed the students who were capable were motivated to learn. However, the students who struggled academically lacked motivation, and as she remarked “they just gave up” because they were so overwhelmed by the amount of objectives they had to learn. As Delta remarked, the “kids who get it, their motivation always seems to be there, but the lower level learners is where I see the problem with motivation.”

Eta who taught in a high-socioeconomic, suburban school did not believe testing motivated students to learn. Specifically, she said “I do not think they gain any motivation from preparing for the test.” Similarly, Theta who also taught in a high-socio eco-

nomic, rural school believed students did not gain motivation from testing. She said students inquired about why they had to do things and if it counted for a grade. Theta had told students if they did not do their best on the Discovery Education Assessment they would be "stuck in intervention groups" and had to do probes which were practice test questions with the objectives assessed on the Discovery Education Assessment to motivate the students to try their best. Iota who taught in a high-socioeconomic, suburban area and Alpha who taught in a low-socioeconomic, urban area concurred with the views of Theta and Eta. As mentioned by Iota, "by no means do I think testing is a huge motivator to do well every day in class. Who cares? It is some test in the spring."

Zeta who taught in a low-socioeconomic, rural school did not mention a lack of motivation, but he did mention he told the students to "show what you know" so that other people know they received a good education. So he offered inspiring words for his students to increase their motivation.

In summary, Beta who taught in a low-socioeconomic, rural school believed testing attributed to a loss of learning. Gamma and Delta who taught in low-socioeconomic, rural schools believed testing motivated high-achievers to learn but not the low-achievers. Most of the teachers independent of socioeconomic status or school geographical location did not view testing as a motivator for learning.

Summary of the Theme

In essence, all but one of the low-socioeconomic, rural teachers believed they were not able to use teachable moments due to testing. The lack of teachable moments were attributed to having to teach to the test, having to teach in a scripted way, and not

having time to listen to what the students considered important which decreased the opportunity for teachable moments. Other participants independent of the socioeconomic status or school geographical location did not mention lack of teachable moments as a concern. Additionally, three out of four teachers in low-socioeconomic, rural schools and one teacher who taught in a high-socioeconomic school perceived they were not able to be creative and teach the way they would like due to testing. These teachers mentioned they were expected to teach the same way although their teaching contexts were different.

Furthermore, teachers from each socioeconomic status and school geographical location revealed testing resulted in rigid scheduling. Pacing guides, class schedules, and lack of time for student remediation were concerns expressed by the participants. Although one participant who taught in a high-socioeconomic, suburban school stated rigid scheduling was a negative influence of testing, she mentioned she was fortunate as compared to her friends who taught in low-socioeconomic schools. Other teachers independent of socioeconomic status or school geographical location did not specifically discuss scheduling as a concern.

Finally, one participant who taught in a low-socioeconomic, rural school believed testing attributed to a loss of learning. Two teachers from low-socioeconomic, rural schools believed testing motivated high-achievers to learn but not the low-achievers. Other teachers independent of socioeconomic status or school geographical location did not view testing as a motivator for learning.

Theme: Causing Stress or Anxiety

"Causing stress or anxiety" was the theme which emerged through data analysis and interpretation to answer the research sub-question "What specific influence do teachers of grades three through five perceive high-stakes testing has on students and teachers? The four subthemes associated with this theme were as follows: teacher anxiety over student performance, pressure on teachers from administrators for AYP scores, student anxiety over their performance, and teacher burnout.

Discussions disclosed all low-socioeconomic, rural teachers experienced stress due to testing. Alpha and Kappa who both taught in low-socioeconomic, urban schools did not specifically mention stress due to testing. However, Alpha and Kappa observed there were some negative influences of testing.

Teacher Anxiety Over Student Performance

Teachers were asked what influence SAT-10 and ARMT testing had on students and teachers. Comments made by the participants revealed all four of the teachers who taught in low-socioeconomic, rural areas experienced stress due to concerns about student performance on high-stakes testing. Other participants independent of socioeconomic status and school geographical location did not mention teacher stress over student performance on high-stakes testing.

Beta who taught in a low-socioeconomic, rural school expressed concerns that testing was stressful for her. She was stressed because she worried about how her students would perform on the test. Beta was also concerned that the high-stakes tests did not account for the days students did not perform well or as Beta said "had a bad day."

Similarly, Gamma who also taught in a low-socioeconomic, rural school mentioned she was under a lot of stress. She believed it was important for her to have “common ground” with all of the schools, but she believed she spent most of her time testing or assessing students and very little time teaching them. She also believed the pressure on teachers to increase test scores negatively affected job performance as evident by her statement, “I feel like because there is pressure on teachers, that teachers do not perform as well as they would if they were just allowed to teach.” Comparatively, Zeta who taught in a low-socioeconomic, rural school also mentioned that testing was stressful because it was stressful for the students and “took a lot” of the students and him. Zeta said he once observed a student who became frustrated during testing. Zeta stated he knew the student wanted to leave the room, but he did not “pitch a fit” as Zeta expected. Zeta noticed the student answered the questions in a pattern by answering “A for a while then B for a while.” Zeta noted this as one example of a negative experience with SAT-10 and ARMT testing because some students did not apply themselves to the test. Similarly, Delta was also concerned about the performance of her students because she believed the tests were too long resulting in students not applying themselves. For instance, Delta said testing “is very tiring for them, and you just see them after the first 10 min; I mean it is almost like they have just given up.” Delta also believed because third graders are so young, it was unfair to ask them to participate in high-stakes testing because it was “daunting” and very “tiring” for the students.

In contrast, Eta who taught in a high-socioeconomic, suburban school mentioned she was thankful for teaching in a high-socioeconomic school. She said her friends in low-socioeconomic schools had been under a lot of stress and almost “choked from the

pressure" of testing. Eta replied, "Sometimes I do not feel like I have much to complain about because of hearing about the way it is for my friends." In essence, Eta perceived her friends who taught in low-socioeconomic, rural schools were under much more pressure than she. However, she still believed the pressure was intense at her school and things still "seemed bad" to her.

In summary, low-socioeconomic, rural teachers were more likely to mention teacher stress than their geographical counterparts. One high-socioeconomic, suburban teacher mentioned that testing increased teacher stress, but she also confirmed teachers in low-socioeconomic areas perceived they were under more stress than teachers who taught in high-socioeconomic schools.

Pressure on Teachers from Administrators for AYP Scores

Teachers were asked what influence SAT-10 and ARMT testing had on students and teachers. Comments made by the participants revealed various perceptions. One teacher from a high-socioeconomic, suburban school and two teachers from low-socioeconomic, rural schools mentioned pressure from administrators.

Gamma specifically mentioned that because she taught in a low-socioeconomic, rural area she noticed what she was required to do increased each year. She was always asked to increase her performance. Gamma remarked, every year she was asked "what else can you do?" Gamma also mentioned she felt overwhelmed because she had to collect data and remediate students who showed weaknesses in the objectives that were tested. She believed because she had to collect data, remediate, and communicate with administrators and interventionists about students' weaknesses she believed her teaching

was not the best it could be and as a result “suffered.” Gamma believed if the school system would “back off” and allow teachers to do what they were trained to do, teachers would be able to do a better job. Gamma specified, “Teachers can get the same data and come to the same conclusions by just doing observations, checklists, and meeting with students. I feel like it is unnecessary to put all these pressures on teachers and students.” Comparatively, Delta who also taught in a low-socioeconomic, rural school believed that because of where she taught there was a lot of pressure on teachers. Particularly, Delta mentioned because the school where she taught always had the lowest test scores in the system the central office placed more pressure on them. As Delta mentioned, “they stayed on us.” For example, central office administrators were in Delta’s and other teachers’ classrooms observing frequently, and they planned with the teachers regularly because there was a chance the school where she taught would not make AYP. She also mentioned the state department had visited their school due to this chance of not making AYP although the school had made AYP during previous years. Delta believed the pressure began with the administrators and trickled to the students.

Eta who taught in a high-socioeconomic, suburban area once had a principal question her about the low results from the high-stakes tests. She had six or seven special education students in her class that particular year. Therefore, the results of the tests were poor. The principal instructed Eta to write a paper about ways she was going to improve her teaching so that this would not occur again in the future. Eta was the only high-socioeconomic, suburban teacher to mention pressure from administrators

Alpha who taught in a low-socioeconomic, urban area said that her administrator “pushed” the teachers at her school so that their main focus was to make AYP, but she

did not view this as stress from her administrator. The researcher inferred she viewed this more as motivation to do well rather than pressure.

In summary, two of the four participants from low-socioeconomic, rural schools and one high-socioeconomic, suburban teacher believed administrators placed pressure on teachers to increase test scores. These teachers viewed pressure from administrators due to testing as a negative influence. Specifically, one teacher from a low-socio economic, rural school believed the pressure due to testing was unnecessary, and that the same conclusions from high-stakes tests can be gathered from observations, checklists, and meeting with students.

Student Anxiety Over Their Performance

Teachers were asked what influence SAT-10 and ARMT testing had on students and teachers. Comments made by the participants revealed students in low-socio economic, rural schools exhibited more signs of stress, which is similar to the teachers in low-socioeconomic, rural schools. One high-socioeconomic, suburban teacher also mentioned some of her students in the past experienced stress, but because she stopped emphasizing testing to her current students the students did not experience anxiety over their performance. Other participants, independent of socioeconomic status or school geographical location did not mention student anxiety over their performance.

Beta from a low-socioeconomic, rural school believed testing had caused stress and student anxiety among her students. Specifically, on the day of ARMT testing one school year, one of Beta's students cried during the middle of a high-stakes test. The student had to be taken outside because she distracted other students with her crying. The

student was crying because she was concerned she would not finish the test and did not want to get “into trouble” for not finishing. The student expressed to Beta she knew how important the test was, and that she wanted to do well and not “disappoint” Beta. However, Beta explained to her that she only wanted her to do her best. Although Beta attempted to console her, the student “broke out in hives and her heart rate increased.” Consequently, the student had to be sent to the nurse because she could not stop crying. Comparatively, Gamma who taught in a low-socioeconomic, rural school also mentioned students experienced anxiety over their performance because her third graders worried if they answered a question correctly. Gamma also believed students were overwhelmed by testing because of its length and because students had to sit quietly and work.

Eta who taught in a high-socioeconomic, suburban school mentioned in her earlier career she had to write a paper for the administrator because her test scores were low, she believed she emphasized the tests so much that she made the students nervous. She also believed she made a few of her students cry over the test. However, she stated she did not overemphasize testing to her students anymore.

On the opposite, Theta who taught in a high-socioeconomic, suburban school believed testing was not stressful and “not that bad.” However, the first year the ARMT was administered she did have one negative experience. Her students cried, and she was very upset because the instructions on the test were “fifteen pages long.” Theta wrote a letter which stated everything that was wrong with the test and handed it to her principal which was sent to the central office in her school system. The central office sent the letter to the state superintendent, and Theta was then accused of cheating on the exam because she knew five of the math questions because she had to read them “over thirty times” to her

students even though reading the questions on the math section of the ARMT was permissible. However, other teachers in the state of Alabama complained shortly after Theta, and the tests were excluded. Theta eventually received an apology for being accused of cheating. Other than this incident, Theta did not believe her students were stressed over the test because she did not do anything to make the students feel stressed. She included test preparation "a little at a time" throughout the year instead of waiting two weeks before the test to prepare students. Theta remarked, "I do it about every couple of weeks. We will practice a bubble grid or practice a writing one so there is no mass panic the week before." Two weeks before the test Theta did "fun review" activities with the students. She believed including test preparation in a "little bit at time" throughout the school year instead of 2 weeks before the test prevented students from being stressed or worried about the test.

Kappa, who taught in a low-socioeconomic urban school, mentioned the tests were long, and the students became restless. So she believed it was difficult keeping the students focused on the test to finish it, but she, like Theta, did not believe the students were stressed.

In summary, participants revealed students in low-socioeconomic, rural schools exhibited more signs of stress during testing, which is similar to the teachers in low-socioeconomic, rural schools. One high-socioeconomic, suburban teacher also mentioned some of her students in the past experienced stress, but because she stopped emphasizing testing to her current students the students did not experience anxiety over their performance. Other participants, independent of socioeconomic status or school geographical location did not mention student anxiety over their performance.

Teacher Burnout

Teachers were asked what influence SAT-10 and ARMT testing had on students and teachers. Comments made by the participants in low-socioeconomic, rural areas inferred they witnessed teacher burnout at the school where they taught. Other participants, independent of socioeconomic status or school geographical location, did not mention teacher burn out.

The only two teachers who mentioned "teacher burn out" were those who taught in low-socioeconomic, rural schools. These participants did not specify that they experienced "burn out," but they observed teachers at their school who did. Gamma believed testing led to teacher burnout. She observed teachers at her school used more sick days and were fatigued or stressed. Gamma noticed the teachers who had been in the teaching field for only 1 or 2 years get "burned out" because of the pressures associated with testing. For example, one teacher at her school thought she was having a heart attack and had to be rushed to the hospital. However, it was only an anxiety attack which Gamma attributed to the testing and the "pressures they face." Zeta who taught in a low-socio economic, rural school mentioned he had noticed new teachers experienced "burnout" because of testing, and these new teachers contemplated leaving the teaching profession.

In summary, only two teachers who taught in low-socioeconomic, rural areas believed the stress of testing had attributed to teacher "burn out." Other participants did not express a specific concern about teacher "burn out." One teacher from a low-socioeconomic, rural school believed the stress of testing is one of the reasons teachers had considered leaving the teaching profession.

Summary of the Theme

Discussions with teachers revealed low-socioeconomic, rural teachers were more likely to mention teacher stress than their school geographical counterparts. One high-socioeconomic, suburban teacher mentioned that testing increased teacher stress, but she also confirmed teachers in low-socioeconomic areas perceived they were under more stress than teachers who taught in high-socioeconomic schools.

Pressure from administrators was another subtheme that emerged. Specifically, two of four participants from low-socioeconomic, rural schools and one high-socio economic, suburban teacher believed administrators placed pressure on teachers to increase test scores. These teachers viewed pressure from administrators due to testing as a negative influence. Specifically, one teacher from a low-socioeconomic, rural school believed the pressure due to testing was unnecessary, and that the same conclusions from high-stakes tests can be gathered from observations, checklists, and meeting with students.

Interestingly, teachers who experienced stressed provided specific examples of students who experienced stress due to testing. Participants revealed students in low-socioeconomic, rural schools exhibited more signs of stress during testing, which is similar to the experiences of the teachers in low-socioeconomic, rural schools. One high-socioeconomic, suburban teacher also mentioned some of her students in the past experienced stress, but because she stopped emphasizing testing to her current students the students did not experience anxiety over their performance. Other participants, independent of socioeconomic status or school geographical location did not mention student anxiety over their performance.

Finally, only two teachers who taught in low-socioeconomic, rural areas believed the stress of testing has attributed to teacher "burn out." The other participants did not express a specific concern about teacher "burn out." One teacher from a low-socio economic, rural school believed the stress of testing is one of the reasons teachers had considered leaving the teaching profession.

Theme: Promoting Inequities

"Promoting inequities" was the theme which emerged through data analysis and interpretation to answer the research sub-question "What specific influence do teachers of grades three through five perceive high-stakes testing has on students and teachers?" The four subthemes associated with this theme were as follows: inadequate learning resources, inadequate parental support, special needs children, and ELLs.

Discussions disclosed teachers in low-socioeconomic schools believed they had inadequate learning resources and parental support as opposed to their richer counterparts. A few teachers in low-socioeconomic, rural areas believed special education students and ELLs were unfairly assessed on high-stakes tests due to time restraints or ability level. Other teachers independent of the socioeconomic status and school geographical location did not believe any group of students was more influenced by testing than another group.

Inadequate Learning Resources

Teachers were asked how SAT-10 and ARMT testing influenced certain populations such as special education students, ELLs, and minorities. Comments revealed two

of the four participants from low-socioeconomic, rural schools perceived they did not have the learning resources as their richer counterparts thereby putting these students at a disadvantage. The participants who taught in high-socioeconomic, rural schools did not mention inadequate resources as a concern.

Beta who taught in a low-socioeconomic, rural school mentioned her students did not have adequate learning resources at home. Specifically, she said computers were available at the school, but her students did not have computers at home which she believed put them at a disadvantage. Therefore, she attempted to teach them as much as she could because when they went home, they did not have the necessary resources to increase their academic achievement. Beta remarked, “I see the scores of the upper schools, and it is not fair because although we do have resources, we do not have the resources some of the children have at home.” In essence, Beta believed her students did not have the same resources at home as the students in high-socioeconomic schools.

Additionally, Delta who also taught in a low-socioeconomic, rural school mentioned it was unfair that the school where she taught was compared to other schools. She believed they were at a disadvantage because her students did not have the same experiences because the students had never been “fifteen miles from their house.” Delta also reported during the summer when students were out of school, they were usually left at home without adult supervision. Specifically, Delta mentioned her students did not go on vacation, to the movies, or to the bowling alley, for example. She also mentioned that due to a lack of funding she was not able to take the students on as many field trips anymore, thereby contributing to low test scores. For example, she discussed the vocabulary section of the reading test. One strategy teachers used was to look at context clues which are

clues within the text to help students determine the meaning of an unknown word. Due to the lack of inadequate learning resources and life experiences, students were not able to use clues from the text to determine the meaning of the given words resulting in lower test scores than their richer counterparts.

In summary, two of the four teachers who taught in low-socioeconomic, rural schools believed their students lacked learning resources at home. One participant believed due to socioeconomic status, students did not have the life experiences needed to score well on high-stakes tests. This inadequacy of resources was stated only among low-socioeconomic, rural teachers.

Inadequate Parental Support

Teachers were asked how SAT-10 and ARMT testing influenced certain populations such as special education students, ELLs, and minorities. Comments revealed two teachers from low-socioeconomic areas who served in different school geographical locations had similar responses. These views are described in the following paragraphs.

Beta who taught in a low-socioeconomic, rural school remarked that she and the teachers at her school tried to do as much as they could to increase student achievement. However, they did not have as much parent support as other schools because the parents chose not to be involved. Comparatively, Alpha who taught in a low-socioeconomic, urban school mentioned she tried to get the parents involved. However, she did not always have the support from the parents that she needed. So she believed her administrator helped motivate the students by promoting “friendly competition” to get the students motivated about their academic achievement.

In summary, two teachers from low-socioeconomic schools both perceived their school lacked the parental support needed to increase student achievement. These two teachers served in different school geographical locations. One teacher served in a rural area while the other teacher served in an urban area. The other participants independent of socioeconomic status and school geographical location did not mention parental support as a concern.

Special Needs Children

Teachers were asked how SAT-10 and ARMT testing influenced certain populations such as special education students, ELLs, and minorities. Comments revealed two of the teachers in low-socioeconomic, rural areas believed special education students were negatively influenced by testing. Yet, the other participants, independent of socioeconomic status and school geographical location, did not believe testing influenced this group differently from others.

Beta who taught in a low-socioeconomic, rural school expressed concerns about special education students who were mentally retarded being required to take the test. She also stated these students had made great academic gains. However, even though they made great academic gains “They had done as well as they could, but the people who graded SAT-10 and ARMT did not care.” In essence, the progress of the student was not important to the people who scored the test; only that one high-stakes test score mattered for AYP purposes. Similarly, Gamma who also taught in a low-socioeconomic, rural school perceived testing had a negative influence on special education students particularly in the areas in which they struggled. Gamma mentioned unless it was written in a

special education student's Individualized Education Plan (IEP) she could not read the reading test to these students. Gamma mentioned she had a student who read fourteen words a minute. Consequently, Gamma's student became very frustrated when she had to read the reading portion of the high-stakes test.

On the opposite, Zeta who taught in a low-socioeconomic, rural school remarked all of his students regardless of race or other diversity responds to testing the same. No particular group was more influenced by testing than the other. He said that the class "practiced and challenged" together. Similarly, when asked about the influence of high-stakes testing on special education students, ELLs, and minorities, Eta and Iota who both taught in a high-socioeconomic, suburban school did not believe testing really influenced these particular populations more than the regular education students. Specifically, Iota remarked, "I really do not think it does. It may affect some, and it will not affect some." Comparatively, Alpha and Kappa who both taught in low-socioeconomic, rural schools did not believe one particular population was more influenced by testing than the other. Alpha believed even if a student received special education services she expected him to make a certain percentile, which indicated she still had goals for these students.

In summary, most of the teachers independent of socioeconomic status and school geographical location did not believe testing had a greater influence on special education students than their regular education counterparts. However, one teacher who taught in a low-socioeconomic, rural school expressed concerns that the special education students were not able to show the progress they made from the beginning of the year until test time because only the high-stakes test score counted. Another participant who also taught

in a low-socioeconomic, rural school believed special education students became more frustrated during test time as opposed to a normal school day.

English Language Learners

Teachers were asked how SAT-10 and ARMT testing influenced certain populations such as special education students, ELLs, and minorities. Several teachers, independent of socioeconomic status or school geographical location believed testing negatively influenced ELLs. The other participants as mentioned previously did not think one group was more influenced by testing than another.

Beta who taught at a low-socioeconomic, rural school stated ELLs were unfairly assessed because they could not read the test to be able to answer the questions. Specifically, Beta said the vocabulary section of the high-stakes test was confusing for some of her regular education students. Hence, obviously the ELLs would have even more difficulty with the vocabulary on the reading section of the high-stakes tests because English was not their native language. Gamma who taught in a low-socioeconomic, rural school also stated because the SAT-10 is timed, the ELLs experienced anxiety. The researcher inferred from Gamma's statement that the ELLs needed more time to process the questions posed since the test was not administered in their native language. Theta who taught fourth grade in a high-socioeconomic, suburban area also mentioned the ELLs had to take the high-stakes tests even though it was not in their native language. She mentioned numerous Ukrainian orphan children were enrolled in her class, and they were on a kindergarten academic level but had to take the fourth grade SAT-10 and ARMT.

Zeta and Eta, on the opposite, did not believe these students were more influenced by testing than the others. These two teachers both served in different academic teaching contexts. Zeta taught in a low-socioeconomic, rural area whereas Eta taught in a high-socioeconomic, suburban area. Alpha and Kappa who both served in low-socioeconomic, rural schools did not have ELLs at their school. Therefore, they were not able to describe accounts of how ELLs were influenced by testing.

In summary, a few teachers independent of socioeconomic status and school geographical location believed ELLs were unfairly assessed because they were not tested in their native language. One teacher who taught in a low-socioeconomic, rural school believed ELLs experienced increased anxiety due to the time restraint of testing. Other teachers, independent of socioeconomic status and school geographical location did not believe ELLs were more influenced by testing than other students, and two the teachers did not have ELLs so they were not able to reflect upon the influence of high-stakes testing on this particular group.

Summary of the Theme

In essence, two of the four teachers who taught in low-socioeconomic, rural schools believed their students lacked learning resources for the home. One participant believed due to their socioeconomic status, students did not have the life experiences needed to score well on high-stakes tests. This inadequacy of resources was stated only among low-socioeconomic, rural teachers.

Most of the teachers independent of socioeconomic status and school geographical location did not believe testing had a greater influence on special education students

than their regular education counterparts. However, one teacher expressed concerns that special education students were not able to show the progress they made from the beginning of the year until test time because only the high-stakes test score counted. Another teacher believed special education students became more frustrated during test time as opposed to a normal school day.

The majority of the teachers, independent of socioeconomic status and school geographical location, did not believe testing had a greater influence on special education students than their regular education counterparts. However, one teacher who taught in a low-socioeconomic, rural school expressed concerns that special education students were not able to show the progress they made from the beginning of the year until test time because only the high-stakes test score counted. Another low-socioeconomic, rural teacher believed special education students became more frustrated during test time as opposed to a normal school day.

Finally, a few teachers independent of socioeconomic status and school geographical location believed ELLs were unfairly assessed because they were not tested in their native language. One teacher who taught in a low-socioeconomic, rural school believed ELLs experienced increased anxiety due to the time restraint of testing. Other teachers, independent of socioeconomic status and school geographical location did not believe ELLs were more influenced by testing than other students.

Conclusions

The major findings derived from the analysis of the interview data from the second, qualitative phase of the study are as follows:

1. All teachers in the study reported they placed more priority on reading and math, which were subjects assessed by the high-stakes tests and placed less priority on subjects not tested such as social studies and science.

2. Two of the four teachers who taught in low-socioeconomic, rural schools believed they were teaching a broad amount of skills as opposed to going into depth with the concepts they taught. One teacher who taught in a high-socioeconomic, suburban school revealed she also had to teach a broad amount of skills as opposed to going into depth with the concepts she taught.

3. Most teachers independent of socioeconomic status and school geographical location believed they increased their teaching of critical thinking skills as a result of ARMT testing due to the open-ended questions students must respond to on this test because they had to think about their answers rather than just select a multiple-choice answer. However, two teachers did not believe ARMT increased critical thinking skills. On the opposite, most teachers reported as a result of SAT-10 testing they decreased the teaching of critical thinking skills. However, one teacher who taught in a low-socio economic, rural school and one teacher who taught in a low-socioeconomic, urban school believed they increased the teaching of critical thinking skills as a result of SAT-10 testing.

4. With the exception of two high-socioeconomic, suburban teachers all of the other participants perceived they were “teaching to the test” due to high-stakes testing and viewed this as a negative influence. Teachers believed “teaching to the test” included the overreliance on the use of test preparation materials that consisted of items similar to the test items found on the high-stakes tests.

5. All teachers independent of the socioeconomic status and school geographical location engaged in school-wide test preparation activities with school administrators through faculty meetings, and high-stakes testing was mainly the focus of these faculty meetings. However, two low-socioeconomic rural teachers reported their administrators did more than discuss test data in faculty meetings. One of these teachers mentioned she had to participate in sessions with her administrators in which she discussed the progress of her students and what she could do to increase their progress. The other teacher mentioned the administrators at her school were actually involved in the planning of reading and math lessons.

6. Teachers in various school geographical locations perceived there was a loss of creativity in teaching. Specifically, teachers did not have time for a teachable moment. This loss of creativity was also attributed to the required use of scripted lessons for the curriculum mandated by one of the school systems to increase student achievement on high-stakes tests as well as teaching schedules being developed by administrators without teacher input.

7. Comments made by most of the participants revealed students were not motivated by testing. These views were independent of socioeconomic status and school geographical location. Two opposing views made by teachers who taught in low-socioeconomic, rural areas revealed testing motivated high-achieving students but not students who struggled academically.

8. Teachers who taught in low-socioeconomic, rural schools reported being stressed about high-stakes testing as opposed to teachers in urban and suburban locations.

Moreover, low-socioeconomic, rural teachers reported their students were stressed due to pressures of high-stakes testing.

9. Comments revealed two of the four participants from low-socioeconomic, rural schools perceived they did not have the learning resources as their richer counterparts thereby putting these students at a disadvantage.

10. Two teachers both from low-socioeconomic schools but one from a rural area and the other from an urban school pointed out a lack of involvement with school which put their students at a disadvantage for testing.

11. The majority of the teachers, independent of socioeconomic status and school geographical location, did not believe testing had a greater influence on special education students than their regular education counterparts. However, two low-socioeconomic, rural teachers believed special education students were disadvantaged by high-stakes testing.

12. Two teachers from low-socioeconomic rural schools and one from a high-socioeconomic suburban school believed ELLs were unfairly assessed because they were not tested in their native language.

Chapter Summary

Qualitative data from individual interviews with nine teachers were used in this chapter to describe the specific influences testing has on teaching practices, on teachers, and on students. The four themes that emerged from the qualitative data analysis were as follows: testing drives the curriculum, loss of creativity in teaching, causing stress or anxiety, and promoting inequities. The findings from this chapter suggest that although

most teachers were negatively influenced due to testing, low-socioeconomic, rural teachers experienced even more negative influences than teachers in urban and suburban areas. However, most teachers reported a positive influence of testing. Specifically, they noted ARMT testing increased the teaching of critical thinking skills. Chapter 6 is comprised of a detailed discussion of the implication of these findings.

CHAPTER 6

DISCUSSION

This study enlightened the need for increased attention to be given to the influences high-stakes testing has on classroom teaching practices in various socioeconomic schools and in various school geographical locations. The use of the mixed methods sequential explanatory design increased the depth and the richness of this study.

Summary of the Major Findings

The purpose of this mixed methods study was to examine elementary teachers' perceptions of the impact of the SAT-10 and the ARMT on classroom teaching practices from a sample of third, fourth, and fifth grade teachers in three large school systems in Alabama that served rural, urban, and/or suburban communities. The purpose of the first, quantitative phase of the study, was to reveal teachers' perceptions of the impact of high-stakes testing on curriculum and instructional approaches, the amount of time spent on critical thinking skills, the amount of time spent on test preparation activities with administrators and in the classroom, and the perceived impact of state tests on students and on teachers by surveying 123 third through fifth grade teachers in three large Alabama school systems.

In the second, qualitative phase of this study, purposeful sampling strategy and maximum variation sampling strategy were employed to interview nine of the 123 third through fifth grade teachers who responded to the survey in the first, quantitative phase

of the study to explore the results from the statistical tests in more depth. ANOVA and Post Hoc analysis identified the differences among school geographical locations on the factors of critical thinking skills and the impact of high-stakes testing on students and teachers. Based on the results of the first, quantitative phase the interview protocol was developed and nine of the 123 participants, who completed the survey in the first, quantitative phase of the study, were selected using maximum variation sampling for individual interviews. In the second, qualitative phase of the study four themes emerged as follows: testing drives the curriculum, loss of creativity in teaching, causing stress or anxiety, and promoting inequities. The mixed methods explanatory sequential design was used in this study to provide a better understanding of the research problem because the qualitative data extended and elaborated on the initial quantitative results. There was no other known study regarding the impact of the SAT-10 and the ARMT on classroom teaching practices that used a mixed methods sequential explanatory design to obtain a more complete picture of high-stakes testing on classroom teaching practices.

Meta-inferences

Metainferences are inferences developed through the integration of the inferences that are acquired on the basis of quantitative and qualitative data in a mixed methods study (Teddlie & Tashakkori, 2003). The section of this chapter integrates the findings from the first, quantitative phase of the study with the findings of the second, qualitative phase of the study. The findings from the first, quantitative phase were interpreted in chapter four. Second, the findings from the first, quantitative phase were used to develop questions to be explored in the second, qualitative phase of the study. Consequently, the

sequence of the mixed methods sequential explanatory design enabled the findings from the second, qualitative phase discussed in chapter five to further explain the results from the first, quantitative phase of the study (Creswell & Plano Clark, 2007; Ivankova, 2004; Ivankova & Stick, 2006).

In this chapter, the results of the quantitative and qualitative phases of the study are discussed by first addressing the meaning of the quantitative results and then explaining how these findings can be further enhanced and elaborated upon by the qualitative findings. Thereafter, the findings are grouped into each of the six factors: curriculum approaches, instructional approaches, time spent on critical thinking skills, time spent on school-wide test preparation activities with administrators, time spent on classroom test preparation activities, and the perceived impact of state tests on students and on teachers. The interpretations of the study results are augmented by comparing them with the high-stakes testing studies that have been conducted in other states (Creswell & Plano Clark, 2007; Ivankova, 2004; Ivankova & Stick, 2006). As each of the six factors is discussed, meta-inferences that emerged from both the quantitative and qualitative analysis are presented. Finally, the study results are discussed within the perceptual theory and the self-efficacy theory that were used as the theoretical framework. The study conclusions, implications, and recommendations are presented.

Central Quantitative Research Question

Are there differences in mean score on teacher curriculum approaches, teacher instructional approaches, class time spent on critical thinking skills, time spent on school wide test preparation, class time spent on classroom test preparation, and the perceived

impact of state tests on students and teachers when controlled for by school geographical location and by socioeconomic status of the students taught?

There were no differences among the teachers from different school geographical locations and the socioeconomic status of students taught for four of the six factors listed. However, statistically significant differences were found among two of the six factors which consisted of the high-stakes testing impact on students and teachers and the time spent on critical thinking skills. Specifically, ANOVA results and Post Hoc analyses indicated urban teachers perceived students and teachers were not as greatly impacted by high-stakes testing as opposed to suburban and rural teachers. On the opposite, rural teachers perceived students and teachers were more greatly impacted by high-stakes testing than urban and suburban teachers. Additionally, ANOVA results and Post Hoc analyses indicated urban teachers spent more time on critical thinking skills than rural and suburban teachers. In contrast, suburban teachers spent less time on critical thinking skills than rural and urban teachers.

Central Qualitative Research Question

In what ways does the combination of school geographical location, the socioeconomic status of students taught, and high-stakes testing influence the classroom instruction of elementary teachers in grades three through five?

Interview data from the second, qualitative phase explained the quantitative results from the first, quantitative phase of the study. As a result of the analysis, four themes emerged: testing drives the curriculum, loss of creativity in teaching, causing stress or anxiety, and promoting inequities. For the feasibility of reporting, the following

discussion of the meta-inferences that emerged from the integration of the results from quantitative and qualitative analysis is organized by the six factors tested in the quantitative phase: curriculum approaches, instructional approaches, time spent on critical thinking skills, time spent on school-wide test preparation activities with administrators, time spent on classroom test preparation activities, and the perceived impact of state tests on students and on teachers.

Curriculum Approaches

In the first, quantitative phase of the study mean factor score differences by school geographical location for teacher curriculum approaches were tested using ANOVA. The results did not yield statistically significant differences in the curriculum approach mean scale score when factorialized by school geographical location. Also, mean factor score differences by socioeconomic status for curriculum approaches were tested using a Student t-test. The results did not yield statistically significant differences in the mean scores of participants who taught in low-socioeconomic and high-socioeconomic schools.

In the second, qualitative phase of the study, “testing drives the curriculum” emerged as a theme among the participants. The subtheme of covering a broad amount of information instead of going into depth with specific concepts emerged among two of the teachers who taught in low-socioeconomic, rural schools and one teacher who taught in a high-socioeconomic, suburban school. On the opposite, one teacher who taught in a high-socioeconomic, suburban area and one teacher who taught in a low-socioeconomic area perceived testing did not have an influence on their teaching practices. Additionally, im-

balances in the content areas emerged as a subtheme among all of the participants, independent of the socioeconomic status or school geographical location. Teachers revealed they placed more priority on math and reading, and subjects that were not tested such as social studies and science received less precedence as noted by all of the participants.

The emergence of the theme "testing drives the curriculum" in the qualitative phase of this study is congruent with findings from other studies (Au, 2007; Costigan, 2002; Crocco & Costigan, 2007; Hoffman & Paris, 2001; Johnson & Johnson, 2006; Lipman, 2004; Pringle & Martin, 2005). For example, Lipman's qualitative study findings suggested teachers from four elementary schools in Chicago ceased from teaching social studies and focused on reading and math the second semester. Also, a study conducted by Hoffman and Paris that examined teachers' perceptions of the impact of the TAAS on teaching practices reported teachers who administered the TAAS reported they planned the objectives they would teach for the year around the areas that students would be assessed on the TAAS. Additionally, Au conducted a meta-analysis of 49 qualitative studies which examined the influence of testing on the curriculum were analyzed. The findings of the meta-synthesis revealed testing had a narrowing effect on the curriculum influencing teachers to use more lecture-based, teacher-centered instruction. However, Au determined other findings suggested high-stakes testing caused an expansion of the curriculum and influenced student-centered instruction which disaffirmed the results of this study. The fact that testing narrowed the curriculum was also mentioned by other researchers (Costigan; Crocco & Costigan; Johnson & Johnson; Pringle & Martin).

Instructional Approaches

In the first, quantitative phase of the study, mean factor score differences by school geographical location for teacher instructional approaches were tested using ANOVA. The results did not yield statistically significant results on differences in the instructional approach mean scale score when factorialized by school geographical location. Also, mean factor score differences by socioeconomic status for instructional approaches were tested using Student's t-test of independent samples. The results did not yield statistically significant differences in the mean scores of participants who taught in low-socioeconomic and high-socioeconomic schools.

However, in the second, qualitative phase of the study, a loss of creativity for teachers emerged as a theme. Discussions revealed some teachers who taught in low-socioeconomic, rural areas expressed concerns about losing a teachable moment as opposed to their geographical counterparts. In contrast, other participants did not address the loss of a teachable moment as a concern. At the same time, some teachers independent of socioeconomic status and school geographical location discussed that "teaching becomes robotic" and that rigid schedules were in place.

The findings that testing negatively impacted teaching practices are congruent with other research findings (Assaf, 2006; Costigan, 2002; Crocco & Costigan, 2007; Hoffman & Paris, 2001; Johnson & Johnson, 2006). For instance, in the study conducted by Hoffman and Paris, teachers who administer the TAAS reported they planned the objectives they taught for the year around the areas that students were assessed on the TAAS (Hoffman et al., 2001). Although reformers considered teachers' planning of objectives around the TAAS as a positive consequence, teachers regarded this outcome as a

negative impact on their teaching practices. Half of the teachers who responded to the survey did not believe that the increased TAAS scores were a result of increased student learning but instead were a result of having to teach to the test.

Similarly, the findings with regards to the SAT-10 which is a standardized test are also congruent with the findings of the study conducted by Assaf (2006). Assaf conducted a qualitative ethnographic study by observing a teacher in the state of Texas. Assaf observed the teacher change her reading instruction from rich and authentic experiences to test-focused instruction focusing on mastery of low-level test skills. Another qualitative study examined the teaching experience of a student teacher in a low-performing urban school. Lloyd (2007) observed and interviewed a student teacher during her ten week student teaching internship and found the student teacher's instruction was greatly influenced by the mandates of the school that followed a test-centered curriculum consisting of worksheets and structured lessons. However, the student teacher made conscious decisions to plan lessons beyond school workbooks to allow students to be engaged in more activities which was in contrast to the adopted teaching style and curriculum at that school.

Comparatively, Crocco and Costigan (2007) interviewed teachers in New York City and found teachers reported they perceived they had lost control over their teaching practice. As a consequence of losing control over the instruction, teachers reported this was a driving force for leaving the teaching profession. Moreover, in another study the researchers temporarily ceased from university teaching and taught third and fourth grade teachers at a low-socioeconomic school in Louisiana during the 2000-2001 school year (Johnson & Johnson, 2006). The authors discovered there were negative consequences of

the demand for accountability in schools, unreasonable demands were placed on teachers which stifled their creativity and enthusiasm, and led to teachers leaving the teaching profession. Costigan (2002) also noted in her study of six first-year teachers of grades three through five in New York City that they viewed testing had a negative impact on students, the curriculum, and classroom teaching practices. The teachers in this study also noted they were unprepared to deal with the volume of testing, and they also noted a loss of power for teachers.

Time Spent on Critical Thinking Skills

In the first, quantitative phase of the study, mean factor score differences by school geographical location for the amount of class time spent on critical thinking skills were tested using ANOVA. The results yielded statistically significant results on differences on time spent on critical thinking skills mean scale score when factorialized by school geographical location. Additionally, post hoc analyses indicated urban teachers spent more time on critical thinking skills than rural and suburban teachers. In contrast, suburban teachers spent less time on critical thinking skills than rural and urban teachers. Mean factor score differences by socioeconomic status for the amount of time spent on critical thinking skills were tested using Student's t-test of independent samples. The results did not yield statistically significant differences in the mean scores of participants who taught in low-socioeconomic and high-socioeconomic schools.

In the second, qualitative phase of the study comments made by the participants varied. However, most teachers, independent of socioeconomic status and school geographical location, reported they decreased the teaching of critical thinking skills due to

SAT-10 testing. This finding is congruent with research findings from other studies (Assaf, 2006; Cankoy & Tut, 2005; Parke, Lane & Stone, 2006). In their study, Cankoy and Tut also found students who spent more time on test-taking strategies did not perform better on math items that required critical thinking. Hence, instructional approaches which focused on high-stakes testing did not increase the critical thinking skills of students in the area of mathematics. Moreover, in this ethnographic study Assaf (2006) observed a teacher changing her reading instruction from rich and authentic experiences to test-focused instruction which resulted in her focusing on mastery of low-level test skills due to testing.

Interestingly, two low-socioeconomic teachers who taught in a rural and an urban school reported they increased the teaching of critical thinking skills due to SAT-10. Furthermore, most teachers independent of socioeconomic status and school geographical location reported they increased the teaching of critical thinking skills due to ARMT whereas two teachers independent of socioeconomic status and school geographical location reported they decreased the teaching of critical thinking skills due to ARMT testing. This range of views among teachers may be attributable to teachers' varying perceptions of the term critical thinking skills. These positive views about ARMT testing are congruent with the findings of the study conducted by Parke, Lane, and Stone (2006). Researchers examined the impact of the MSPAP and the MLO including principal, teacher, and student beliefs, classroom teaching practices, and student learning in writing and in reading (Parke et al., 2006). The results of the study conducted by Parke et al. indicated principals and teachers were supportive of the MSPAP, and performance gains in the areas of reading and writing occurred in schools that used reform-oriented instruction.

Time Spent on Test Preparation Activities with Administrators

In the first, quantitative phase of the study, mean factor score differences by school geographical location for the amount of time spent on school-wide test preparation activities with administrators were tested using ANOVA. The results did not yield statistically significant results on differences on the amount of time spent on school-wide test preparation activities with administrators mean scale score when factorialized by school geographical location. Additionally, mean factor score differences by socioeconomic status for the amount of time spent on school-wide test preparation activities were tested using Student's t-test. The results did not yield statistically significant differences in the mean scores of participants who taught in low-socioeconomic and high-socioeconomic schools.

The second, qualitative phase of the study revealed that all the teachers independent of socioeconomic status and school geographical location engaged in test preparation strategies with their administrators. Most of the teachers' time spent with administrators regarding testing was during faculty meetings discussing overall school-wide data. One low-socioeconomic, rural teacher mentioned that she participated in "sit and chats" with her administrator once a month while another low-socioeconomic, rural teacher mentioned her administrators were involved in the actual planning of reading and math lessons. In essence, most teachers reported administrators used test data to discuss student progress, and this was discussed at faculty meetings. However, two, low-socioeconomic rural teachers, reported their administrators did more than discuss test data in faculty meetings. One of these teachers mentioned she had to participate in the sessions with administrators in which she discussed the progress of her students and what she could do to

increase their progress. The other teacher mentioned the administrators at her school were actually involved in the planning of reading and math lessons.

The finding that most teachers reported administrators discussed school-wide data at faculty meetings to determine how students were progressing is congruent with Grant's (2000) study in which elementary teachers mentioned principals were "more likely to talk about test scores as part of a bigger picture of how students are progressing" (p. 16). Similarly, in this study two low-socioeconomic, rural teachers reported their rural administrators were more involved in testing than just discussing school-wide data at meetings. As mentioned earlier, one teacher reported she participated in sessions with administrators to discuss student progress, and another teacher reported her administrators were involved in planning reading and math lessons with teachers. Consequently, rural administrators in two schools spent more time on high-stakes testing instructional leadership than suburban and urban administrators. This finding is inconsistent with Egley and Jones (2004) findings. Egley and Jones reported rural elementary administrators spent a similar amount of time daily on instructional leadership as suburban and urban administrators. However, rural administrators perceived the FCAT to have a more positive impact on their ability to increase teacher effectiveness than administrators in urban and suburban communities and found the FCAT more useful than urban administrators in aiding them in assessing teachers' strengths and weaknesses in the areas of math, reading, and writing.

Time Spent on Classroom Test Preparation Activities

In the first, quantitative phase of the study, mean factor score differences by school geographical location for the amount of class time spent on classroom test preparation activities were tested using ANOVA. The analysis did not yield statistically significant results on differences on the amount of time spent on classroom test preparation activities mean scale score when factorialized by school geographical location. Additionally, mean factor score differences by socioeconomic status for the amount of time spent on classroom test preparation were tested using Student's t-test of independent samples. The results did not yield statistically significant differences in the mean scores of participants who taught in low-socioeconomic and high-socioeconomic schools

In the second, qualitative phase of the study all of the participants, except two high-socioeconomic, suburban teachers, perceived they were "teaching to the test" due to high-stakes testing and viewed this as a negative influence. The researcher inferred the "teaching to the test" consisted of using test preparation activities to prepare students for the test. These test preparation activities included using test preparation booklets to expose students to items similar to the ones on the tests. On the opposite, two high-socioeconomic, suburban teachers reported they did not "teach to the test" because they did not increase the amount of test preparation activities they used in their classrooms.

Hence, the findings of this study are congruent with research findings in other studies (Booher-Jennings, 2005; Lipman, 2004; McNeil & Valenzuela, 2001; Moon et al., 2003). In a quantitative study conducted by Moon et al. teachers indicated on questionnaires that they spent a significant amount of time preparing students for state-mandated tests, but teachers of low-socioeconomic students reported spending more time on test

preparation strategies than teachers of high-socioeconomic students. Additionally, Booher-Jennings (2005) collected qualitative data from an urban elementary school in Texas. Their findings suggested resources and intervention were withdrawn from students who were less likely to pass the TAAS which resulted in low socioeconomic and minority students losing opportunities to engage in higher-order thinking, analytical writing, and problem solving skills. Finally, in a study conducted by Lipman (2004), it was reported test preparation was used with students such as engaging in practice tests, filling in bubbles on scantron sheets, becoming familiar with the format of the tests and types of questions posed on the tests, and learning test strategies for eliminating incorrect answers.

Perceived Impact of State Tests on Students and Teachers

In the first, quantitative phase of the study mean factor score differences by school geographical location for the perceived impact of state tests on students and teachers were tested using ANOVA. The analysis yielded statistically significant results on perceived impact of state tests on students and teachers mean scale score when factorialized by school geographical location. Post Hoc analyses indicated urban teachers perceived students and teachers were not as greatly impacted by high-stakes testing as opposed to suburban and rural teachers. On the opposite, rural teachers perceived students and teachers were more greatly impacted by high-stakes testing than urban and suburban teachers. Additionally, mean factor score differences by socioeconomic status for the perceived impact of state tests on students and teachers were tested using Student's t-test of independent samples. The results did not yield statistically significant differences in

the mean scores of participants who taught in low-socioeconomic and high-socioeconomic schools.

The t-test on socioeconomic status did not yield statistically significant results in the first, quantitative phase. However, in the second, qualitative phase of the study themes related to stress or anxiety due to testing emerged among teachers of low-socioeconomic, rural schools as opposed to their geographical counterparts. Comments made by the participants revealed all four of the teachers interviewed who taught in low-socioeconomic, rural areas experienced stress due to concerns about student performance on high-stakes testing. Some of the low-socioeconomic, rural teachers also mentioned their students were stressed as a result of the high-stakes tests. A few teachers also mentioned they were pressured from administrators to increase high-stakes testing and two low-socioeconomic, rural teachers mentioned some of the teachers at their school were experiencing teacher burnout.

The qualitative and quantitative findings in this study with regards to testing causes stress and anxiety are congruent with the quantitative findings in other studies (Berger, 2006; Hoffman & Paris, 2001; Mulvenon et al., 2005; Tripplett & Barksdale, 2005). For instance, the quantitative study conducted by Berger investigated the relationship between teacher stress and high-stakes testing and whether the stress among teachers differed between urban and rural teachers in the state of Virginia. The findings suggested rural teachers scored higher stress scores than urban teachers for the variables of personal stress, frustration with student effort, and teacher morale. With regards to student stress, Tripplett and Barksdale studied elementary students' perceptions of high-stakes testing via drawings and written responses to questions posed by the researchers. The findings

suggested there was an overwhelming amount of negativity which supported previous research reports of students experiencing anxiety during high-stakes testing. Additionally, Hoffman and Paris examined teachers' perceptions of the impact of the TAAS. In Hoffman and Paris' study teachers also reported students were aggressive, irritable, or anxious during testing, and some students developed stomachaches and headaches while taking the TAAS. Furthermore, the study conducted by Mulvenon et al. suggested teachers had strong concerns about standardized testing and reported the greatest amount of anxiety. In contrast, most principals, counselors, parents and students valued standardized testing and did not perceive an increase in stress or anxiety due to testing disaffirming teachers' beliefs that high-stakes testing is stressful for students.

Also, in the second, qualitative phase of the study the theme "promoting inequities" emerged. Discussions disclosed teachers in low-socioeconomic schools believed students had inadequate learning resources, and they lacked parental support as opposed to their richer counterparts which put their students at a disadvantage for testing. Moreover, a few teachers in low-socioeconomic, rural areas believed special education students and ELLs were unfairly assessed on high-stakes tests due to time restraints or ability level. One teacher in a high-socioeconomic, suburban area also believed ELLs were unfairly assessed. Other teachers independent of the socioeconomic status and school geographical location did not believe any particular group of students was more influenced by testing than another group. Additionally, most of the participants independent of socioeconomic status and school geographical location revealed students were not motivated by testing. These views were independent of socioeconomic status and school geographical location. Finally, two opposing views made by teachers who taught in low-

socioeconomic, rural areas revealed testing motivated high-achieving students but not students who struggled academically.

The qualitative subtheme “lack of parental support” that emerged for the teachers in low-socioeconomic schools is congruent with the qualitative findings of the study conducted by Newman and Chin (2003). The findings suggested parents of low-socio economic students and of those who speak little English were unlikely to be able to provide academic instruction at home for their children. The parents of these children had to put the family’s income before the children’s educational needs.

Additionally, the qualitative subtheme “ELLs unfairly assessed” emerged for a few teachers in low-socioeconomic, rural schools and among one teacher in a high-socioeconomic, suburban school. A few low-socioeconomic, rural teachers believed ELLs were unfairly assessed because they were not tested in their native language which put the ELLs at a disadvantage for testing. The teachers’ perceptions in this study are in stark contrast to teachers’ perceptions in a study conducted by Escamilla, Chavez, and Vigil (2005) which found teachers perceived Spanish speaking students were responsible for their low achievement scores on accountability tests. However, the test data showed Spanish-speaking Latino students in ELL programs who took the Spanish version of the state test met state standards. The latter finding from the study conducted by Escamilla et al. affirmed some teachers’ perceptions in this study that ELLs are unfairly assessed on high-stakes testing because they are not tested in their native language.

New Findings

This study generated new findings related to the impact of high-stakes testing on classroom teaching practices. For instance, this study again reinforced the fact that testing narrowed the curriculum which is consistent with prior research (Au, 2007; Costigan, 2002; Crocco & Costigan, 2007; Hoffman & Paris, 2001; Johnson & Johnson, 2006; Lipman, 2004; Pringle & Martin, 2005). In this study, the teachers also reported they placed more emphasis on reading and math. However, although the curriculum had been narrowed to emphasize these two subjects, some teachers mentioned they had to cover a broad amount of skills instead of teaching specific skills in depth. Thus, teachers had insufficient time to go into depth due to the need to cover numerous skills in a short amount of time. These findings add insight related to specific instructional approaches the teachers use to accommodate for the changes that high-stakes testing imposed. Second, previous research mentioned a loss of creativity in teaching due to testing (Hoffman et al., 2001; Johnson & Johnson, 2006). However, no other known studies specifically reported this fact. In this study, it emerged that "teaching becomes robotic" which implied teachers have to teach in a "scripted" way. Moreover, in this study one teacher addressed a concern of rigid scheduling in which teachers had little input on the development of their daily teaching schedules of subjects to be taught. Other teachers reported they were required to teach certain lessons on certain days. As a result, teachers reported a loss of teachable moments which consisted of not being able to listen to students share things that were important to them due to having to adhere to schedules. Finally, because there are no other known studies of teachers' perceptions of the impact of SAT-10 and ARMT testing on classroom teaching practices in Alabama, the finding related to the increase of

the teaching of critical thinking skills due to ARMT is significant. This can be interpreted as an example of a positive influence of high-stakes testing on teaching practices versus the significant amount of research about negative influences of the test. These new findings along with findings from previous research are presented in Table 23.

Table 23

New Findings and Findings Congruent with Previous Research

New Findings	Findings Congruent with Previous Research
Breadth Versus Depth	Testing Drives the Curriculum or Narrowing of the Curriculum (Au, 2007; Costigan, 2002; Crocco & Costigan, 2007; Hoffman & Paris, 2001; Johnson & Johnson, 2006; Lipman, 2004; Pringle & Martin, 2005)
Teaching Becomes Robotic (One School System)	Loss of Creativity in Teaching (Flores & Cooke, 2003; Gardner, 2002; Johnson & Johnson, 2006; Paris & Urdan, 2000; Smyth, 2008)
Loss of Teachable Moment	Causing Stress or Anxiety (Berger, 2006; Hoffman & Paris, 2001; Mulvenon et al., 2005; Tripplett & Barksdale, 2005)
Rigid Scheduling	Inadequate Learning Resources in Rural Schools (Gollnick & Chinn, 2002)
	Inadequate Parental Support in Low-Socioeconomic Schools (Newman & Chin, 2003)
ARMT Increased Critical Thinking Skills	Groups such as ELLs and/or Special Education Students Disadvantaged by High-Stakes Testing (Jones et al., 2003; Menken, 2008; Smyth, 2008)

Addressing the Perceptual and Self-Efficacy Theoretical Framework

Perceptual Theory

Behavior is a result of the personal meanings an individual associates with a situation (Combs, 1976). The factors that determine the behavior of individuals are the experiences of the individual at the time of behavior. Because the teachers all had different experiences with testing, there were variations of the perceptions of the impact of high-stakes testing on classroom teaching practices. Some teachers perceived high-stakes testing greatly influenced their teaching practices, whereas other teachers perceived high-stakes testing had little or no influence on their teaching practices.

The perceptual theory has four properties which include stability, fluidity, direction, and intensity (Combs et al., 1976). The fluidity of the perceptual field makes change in behavior feasible and capable of adapting to changing conditions in an effort to live and obtain satisfaction. Fluidity allows for "learning, reasoning, remembering, forgetting, and creativity" (Combs, Richards, & Richards, 1976, p. 27). All teachers in this study reported testing impacted their teaching to some extent. However, one teacher from a low-socioeconomic, rural school and another from a high-socioeconomic suburban school believed testing did not impact their teaching practices as opposed to their peers. Although this statement was contradictory to other statements made throughout the interview, the researcher inferred that these teachers met the requirements of their school system to some extent, but also ensured they adhered to best teaching practices by going beyond a test preparation curriculum to ensure students received a well-rounded education. Thus, they believed testing requirements did not influence them to completely deviate from their beliefs about teaching and how students should learn. In essence, these

teachers found ways to cope with meeting school requirements but yet remain true to their teaching philosophies by effectively balancing the two. The teachers who believed testing did not impact their teaching practices and continued to do what they believed was best for students possessed fluidity. In essence, the researcher inferred fluidity prevented these teachers from leaving the teaching profession and allowed them to adjust their practices to meet the requirements of the school system but yet adhere to best teaching practices.

Self-Efficacy Theory

The findings from the quantitative and qualitative phases of the study also corroborated the use of the self-efficacy theory to guide the research for this study. Teacher efficacy influences teachers' enthusiasm, persistence, commitment, and instructional strategies in addition to students' motivation, achievement, and beliefs about self-efficacy (Tschannen-Moran & Hoy, 2001). Individuals who believe they will succeed increase resilient self-efficacy (Bandura, 1994). Bandura concluded people's beliefs about their efficacy influence what they do as a group, the amount of effort they place into a task, their endurance when their efforts do not yield quick results, and their probability of success.

The findings in this study aligned with the concepts of the self-efficacy theory in that high-stakes testing influenced rural teachers more than their geographical counterparts thereby impacting their self-efficacy. Specifically, rural teachers perceived students and teachers were more greatly impacted by high-stakes testing than urban and suburban teachers, and all four of the low-socioeconomic rural teachers reported being stressed due

to testing. Consequently, three out of the four teachers who taught in low-socioeconomic, rural areas believed they were not able to use teachable moments or teach the way they believed students should be taught due to testing. In other words, these teachers perceived they had to divert from their beliefs about teaching and best practices and meet the mandates of the school system. Hence, the finding that low-socioeconomic, rural teachers were more greatly impacted by high-stakes testing may be attributed to the stress experienced by three of the four rural teachers thereby influencing their self-efficacy.

In summary, teachers' perceptions of the impact of high-stakes testing varied among the participants. This variation of perceptions was due to the different experiences of each individual. One teacher from a low-socioeconomic, rural school and another teacher from a high-socioeconomic, suburban school both stated testing did not influence their teaching practices at all whereas the influence of testing seemed to have different degrees of influence on other teachers. Three of the four low-socioeconomic, rural teachers believed they had to deviate from their beliefs to meet the mandates of testing. Survey data and discussions in individual interviews disclosed low-socioeconomic, rural teachers appeared to be more influenced by testing than teachers from urban and suburban areas.

Implications

Implications for Policymakers

In order to judge the effectiveness of schools, policymakers must form a panel of various stakeholders to discuss how schools should be evaluated (Paris & Urdan, 2000). A panel should consist of policymakers, educators from various socioeconomic schools, educators from various school geographical locations, parents, and students before poli-

cies are implemented. Assessments that are used to judge the effectiveness of schools should demonstrate a variety of ways for schools to demonstrate their effectiveness. The comparisons among schools are ineffective because students come from different backgrounds and socioeconomic statuses. As this study noted there are inadequacies of resources in rural and urban schools. Instead of implementing standardized tests which compare schools to each other, criterion-referenced tests should be used to document student progress.

Implications for Educational Leadership Programs

All of the teachers in this study noted there were imbalances in the curriculum as a result of testing. Specifically, subjects that were tested such as reading and math receive more precedence as opposed to subjects that were not tested such as social studies and science. In this era of high-stakes testing, educational leadership programs must prepare administrators to effectively lead schools in the endeavor to provide a well-balanced curriculum to prepare students for the professional world and society. Specifically, educational leadership programs should offer courses that focus specifically on accountability and ways to identify and resolve current problems related to accountability measures.

Additionally, because some teachers believed administrators placed pressure on teachers to increase test scores, educational leadership programs should offer courses that prepare educational leaders to create a school environment which promotes high morale. Because rural teachers reported stress as a result of high-stakes testing and teachers who taught in low-socioeconomic schools disclosed there were inadequacies of resources, educational leaders should also be required to complete internships in a variety of schools

with varying demographics to experience situations unique to certain school geographical locations or a specific socioeconomic status.

Implications for School Administrators

An overwhelming amount of the perceptions examined in this study were negative, and some teachers noted they were stressed as a result of testing. Consequently, school administrators should decrease stress and anxiety by creating a support system for teachers. Specifically, administrators should encourage teachers to engage in peer observations so that teachers can learn from each other in creative ways to adhere to developmentally appropriate practices while yet meeting AYP as mandated by NCLB. Also to reduce stress and anxiety, administrators should seek ways to increase morale by seeking and implementing teacher input. For instance, as a result of the qualitative finding that rigid schedules were in place due to testing, administrators should grant teachers more flexibility and input in scheduling and in the pacing of the curriculum. Additionally, administrators should conduct focus groups with teachers (Berger, 2006) on a regular basis to identify and resolve problems that may arise as a result of high-stakes testing and to gain a plethora of ideas for instructional strategies that are considered best practices but will also help schools make AYP.

Implications for Teachers

Teachers should voice their concerns about the impact of high-stakes testing on classroom teaching practices to policymakers and should ask to be included in the development of the policies regarding the accountability measures of schools. The accountabil-

ity measures should require a variety of ways to assess students and judge the effectiveness of schools (Paris & Urdan, 2000). The accountability measures employed should consist of the use of portfolios, essays, open-ended response questions, and multiple choice questions. This variety of assessments will provide a wider scope of the possible growth and capabilities of students.

Until standards for judging the effectiveness of schools are changed, educators must find creative ways to provide students with a balanced curriculum that includes developmentally appropriate practices but yet also covers the objectives on the high-stakes tests. To increase the use of best practices, teachers should adhere to the standards for National Board Certification to enhance their instruction. One of the components of National Board Certification for elementary teachers is to integrate the subjects of math and science (National Board for Professional Teaching Standards, 2010). Teachers must effectively integrate these subjects as well as other subjects to create a well-balanced curriculum. To include social studies and science in the curriculum, teachers must integrate these subjects into other areas such as reading to ensure these subjects are not deemphasized and to ensure students are not engaged in a narrow curriculum.

Recommendations for Future Research

As a result of the findings, conclusions, and implications, in this study, recommendations for future research are presented in the following paragraphs. Most teachers reported ARMT testing had increased the teaching of critical thinking skills. Hence, an in-depth study of ARMT testing should be examined by the use of survey data, interview data, and observational data in the classroom by the researcher to determine how teachers

increase the teaching of critical thinking skills due to ARMT testing. Additionally, teachers' perceptions of what is considered critical thinking skills should be examined in more depth because perceptions of this term may have varied among the participants.

The sample population of teachers in this study was limited to 123 participants. Specifically, the sample population of urban teachers in this study consisted of only 22 of the 123 participants (17.9%). Because of this limited sample, the impact of high-stakes testing on urban teachers' teaching practices should be further explored to determine if additional trends and themes might emerge as a result of a larger sample size and more diverse perspectives.

This study was limited to elementary teachers' perceptions of the impact of high-stakes testing on classroom teaching practices. Teachers discussed their perceptions of the impact of high-stakes testing on instructional practices, on themselves, and on their students. There are limited studies on high-stakes testing in Alabama at the high school level and from administrators' perspectives. Because of the limited research of high-stakes testing in Alabama and because only elementary teachers were included in this study, additional and more comprehensive studies should be conducted to examine the perceptions of administrators, teachers, students, and parents beliefs of how high-stakes testing has impacted students to obtain a variety of experiences and perceptions and to ultimately help students avoid the stress of high-stakes testing and succeed.

The findings of this study concluded low-socioeconomic, rural teachers experienced stressed due to testing, and these teachers also perceived their students experienced stress. Consequently, future studies should be conducted to obtain teacher perceptions and student perceptions of stress due to testing. The perceptions of the teachers and

the students should be examined to confirm or disconfirm the relationship between teacher stress and student stress as a result of high-stakes testing.

Summary

This study revealed teaching practices at elementary schools were influenced by the implementation of high-stakes testing. In the first, quantitative phase ANOVA and Post Hoc analysis indicated urban teachers perceived students and teachers were not as greatly impacted by high-stakes testing as opposed to suburban and rural teachers. On the opposite, rural teachers perceived students and teachers were more greatly impacted by high-stakes testing than urban and suburban teachers. Additionally, ANOVA results and Post Hoc analysis indicated urban teachers spent more time on critical thinking skills than urban and suburban teachers. In contrast, suburban teachers spent less time on critical thinking skills than rural and urban teachers.

In the second, qualitative phase all teachers independent of socioeconomic status or school geographical location reported they increased their focus on reading and math, which were the subjects assessed on high-stakes tests and de-emphasized subjects not tested such as social studies and science. Additionally, the study revealed teachers of low-socioeconomic, rural schools mentioned stress in the qualitative phase of the study as opposed to their geographical counterparts. New findings that were not reported in any other known studies emerged in the second, qualitative phase which included: breadth versus depth, teaching becomes robotic due to the use of “scripted” lessons in one school system, a loss of a teachable moment, rigid scheduling, and ARMT increased the teaching of critical thinking skills.

The perceptual and self-efficacy theories guided this study. Because the teachers all had different experiences with testing, there were variations of the perceptions of the impact of high-stakes testing on classroom teaching practices. In the second, qualitative phase, some teachers perceived high-stakes testing greatly influenced their teaching practices, whereas other teachers perceived high-stakes testing had little or no influence on their teaching practices. The two teachers who reported high-stakes testing did not impact their teaching practices possessed fluidity which allowed them to adjust their practices to meet the requirements of the school system but yet adhere to best teaching practices. Three of the four teachers who taught in low-socioeconomic, rural schools believed they had to deviate from their beliefs to meet the mandates of testing. Survey data and discussions in interviews disclosed low-socioeconomic, rural teachers appeared to be more influenced by testing than teachers from urban and suburban areas. Hence, the finding that rural teachers were more greatly impacted by high-stakes testing may be attributed to the stress experienced by rural teachers thereby influencing their self-efficacy.

Based on the results of this study, it is recommended that future studies focus on the ARMT and the increase of critical thinking skills due to ARMT testing. The study should include observational data of classroom teaching to confirm or disconfirm the use of critical thinking skills. Additionally, future studies should be conducted to obtain teacher perceptions and student perceptions of stress due to testing. The perceptions of the teachers and the students should be examined to confirm or disconfirm the relationship between teacher stress and student stress as a result of high-stakes testing.

LIST OF REFERENCES

Abedi, J., & Dietal, R. (2004). Challenges in the no child left behind act for English language learners. *CRESST Policy Brief 7*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Alabama Association of Schools Boards. (2007). *Is no child left behind working in Alabama*? Retrieved February 25, 2008 from http://alabamaschool boards.org/ASB_NCLBworking.html

Alabama Education News. (2008). *83% of Alabama schools make AYP!* Retrieved September 28, 2008 from http://www.alsde.edu/general/aen.pdf

Alabama School Journal (2008). AYP: *The good, the bad, and the wrong.* Retrieved November 13, 2008 from http://www.myaea.org/PDFfile/ASJ/ AS-JAug182008.pdf

Alabama State Department of Education. (2010). Assessment and Accountability. Retrieved January 21, 2010 from http://www.alsde.edu

Albrecht, S., & Joles, C. (2003). Accountability and access to opportunity: Mutually exclusive tenets under a high-stakes testing mandate. *Preventing School Failure, 47*(2), 86-91.

Allport, F. H. (1955). *Theories of perception and the concept of structure*. New York: John Wiley & Sons.

Assaf, L. (2006). One reading specialist's response to high-stakes testing pressures. *The Reading Teacher, 60*(2), 158-167.

Arriaza, G. (2004). Making changes that stay made: School reform and community involvement. *High School Journal, 87*(4), 10-24.

Ashton, P. (1985). Motivation and the teacher's sense of efficacy. In C. Ames & R. Ames (Eds.) *Research on motivation in education: The classroom milieu* (pp. 141-171). New York: Academic Press.

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher, 36*(5), 258-267.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewoood Cliffs, NJ: Prentice-Hall, Inc.

Barksdale-Ladd, M., & Thomas, K. (2000). "What's at stake in high-stakes testing; Teachers and parents speak out." *Journal of Teacher Education, 51*, 384-397.

Bazeley, P. (2003). Computerized data analysis for mixed methods research. In A. Tashakorri & C. Teddlie (Eds.), *Handbook on mixed methods in the behavioral and social sciences* (pp. 385-422). Thousand Oaks, CA: Sage Publications.

Berger, A. (2006). *High-stakes testing and its relationship to stress between rural and urban elementary school teachers*. Ph.D. dissertation, Walden University.

Booher-Jennings, J. (2005), Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal*, *42*(2), 231-268.

Cankoy, O., & Tut, M. (2005). High-stakes testing and mathematics performance of fourth graders in north cyprus. *Journal of Educational Research, 98*(4), 234-243.

Canner, J. (1992). Regaining the public trust: A review of the school testing programs, practices. *NASSP Bulletin, 76*, 6-15.

Cimbricz, S. (2002, January 9). State-mandated testing and teachers' beliefs and practice. *Education Policy Analysis Archives, 10*(2). Retrieved July 5, 2007 from http://epaa.asu.edu/epaa/v10n2.html

Cizek, G. (2001). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice*, *21*(4), 18-27.

Combs, A. W., & Snygg, D. *Individual behavior*. New York: Harper & Row, 1959.

Costigan, A.T. (2002). Teaching the culture of high stakes testing: Listening to new teachers. *Action in Teacher Education, 23*(4), 28-34.

Creswell, J. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage Publications.

Creswell, J. (2005). *Educational research: Planning, conducting and evaluating quantitative and qualitative research.* Columbus, OH: Merrill Prentice Hall.

Creswell, J.W. (2008). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches.* Thousand Oaks: SAGE.

Creswell, J. W. & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, *39*(3), 1-7.

Creswell, J. W., Plano-Clark, V. L., Gutmann, M., & Hanson, W. (2003). Advanced mixed methods research designs. In A. Teddlie & Tashakkori (Eds.), *Handbook on mixed methods in the behavioral and social sciences* (pp. 209-240). Thousand Oaks, CA: Sage.

Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research.* Thousand Oaks, CA: Sage.

Crocco, M. S., & Costigan, A.T. (2007). The narrowing of curriculum and pedagogy in the age of accountability. *Urban Education*, *42*(6), 512-535.

Crocker, L. (2003). Teaching for the test: Validity, fairness, and moral action. *Educational Measurement: Issues & Practice*, *22*(3), 5-11.

Cronbach, L. J. (1984). *Essentials of psychological testing* (4th ed.). New York: Harper & Row.

Dillman, D. A. (2006). *Mail and internet surveys: The tailored design method* (2nd ed). Hoboken, NJ: John Wiley and Sons, Inc.

Dillman, D. A. (2008). *Internet, mail, and mixed mode surveys: The tailored design method* (3rd ed). Hoboken, NJ: John Wiley Co.

Egly, R., & Jones, B. (2004). Rural elementary administrators' view of high-stakes testing. *The Rural Educator*, *26*(1), 30-39.

Escamilla, K., Chávez, L., & Vigil, P. (2005). Rethinking the "gap": High-stakes testing and Spanish-speaking students in Colorado. *Journal of Teacher Education*, *56*(132), 132-144.

Fletcher, J., Francis, D., Boudousquie, A., Copeland, K., Young, V., Kalinowski, S., et al. (2006). Effects of accommodations on high-stakes testing for students with reading disabilities. *Exceptional Children*, *72*(2), 136-150.

Flores, B., & Clarke, E. (2003). Texas voices speak out about high-stakes testing: Preservice teachers, teachers, and students. *Current Issues in Education*, *6*(3), Retrieved July 18, 2009 from http://cie.asu.edu/volume6/number3

Garcia, O. & Menken, K. (2006). The English of Latinos from a plurilingual transcultural angle: Implications for assessment and schools, In S. Nero (ed.) *Dialects, Englishes, Creoles, and Education*. Clevedon: Multilingual Matters.

Gardner, S. (2002). Forecasting and managing student achievement on high-stakes tests. *T H E Journal*, *29*(6), 40-41.

Gayler, K. (2005). *How have high school exit exams changed our schools? Some perspectives from Virginia and Maryland*. Center on Education Policy: Washington, DC.

Gerwin, D. (2004). Preservice teachers report the impact of high-stakes testing. *The Social Studies*, *95*(2), 71-74.

Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construction validation. *Journal of Educational Psychology*, *76*, 569-582.

Glaser, B. G., & Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research.* Chicago: Aldine Publishing Company.

Goertz, M., & Duffy, M. (2003). Mapping the landscape of high-stakes testing and accountability programs. *Theory Into Practice*, *42*(1), 4-11.

Goldson, D. (2005). Elementary science: Left behind? *Journal of Science Teacher Education*, *16*, 185-187.

Gollnick, D., & Chinn, P. (2002). *Multicultural Education in a Pluralistic Society*. Columbus, OH: Merrill Prentice Hall.

Gordon, S., & Reese, M. (1997). High-stakes testing: Worth the price? *Journal of School Leadership*, *7*, 345-368.

Grant, S. (2000). Teachers and tests: Exploring teachers' perceptions of changes in the New York state testing program. *Education Policy Analysis Archives, 8*(14), 1-28.

Grant, S. (2004). Oppression, privilege, and high-stakes testing. *Multicultural Perspectives*, *6*(1), 3-11.

Grant, S. (2006). *Measuring history: Cases of state-level testing across the United States*. Greenwich, CT: Information Age.

Greene, J., Winters, M., & Forster, G. (2004). Testing high-stakes tests: Can we believe the results of accountability tests? *Teachers College Record, 106*(6), 1124-1144.

Greene, J., Winters, M., & Forster, G. (2005). *Education myths: What special interest groups want you to believe about our schools – and why it isn't so.* Lanham: MD: Rowman and Littlefield Publishers, Inc.

Gulek C. (2003). Preparing for high-stakes testing. *Theory Into Practice, 42*(1), 42-50.

Hansen, W. E., Creswell, J. W., Plano Clark, V. L., Petska, K. P., & Creswell, J. D. (2005). Mixed methods research designs in counseling psychology. *Journal of Counseling Psychology, 52*(2), 224-235.

Hatch, L. A. (2002). *Doing qualitative research in education settings*. Albany, NY: State University of New York Press.

Haynes, K. (2008). Through the nominations of principals: Effective teachers of African American students share limitations of high-stakes testing. *Teaching and Teacher Education: An International Journal of Research and Studies, 24*(8), 2157-2167.

Healey, J. (2002). *Statistics: A tool for social research*. Sydney, Australia: Wadsworth.

Henning, J. (2006). Teacher leaders at work: Analyzing standardized achievement data to improve instruction. *Education, 126*(4), 729-737.

Hoffman, J., Assaf, L., & Paris, S. (2001). High-stakes testing in reading: Today in Texas, tomorrow? *The Reading Teacher, 54*(5), 482-492.

Hogan, T. (1983). Measurement implications of "a nation at risk." *Educational Measurement: Issues and Practice 2*(4), 32-43.

Hong, W., & Youngs, P. (2008). Does high-stakes testing increase cultural capital among low-income and racial minority students?. *Education Policy Analysis Archives, 16*(6). Retrieved July 27, 2009 from http://epaa.asu.edu/epaa/v16n6/

Horn, C. (2003). High-stakes testing and students: Stopping or perpetuating a cycle of failure? *Theory Into Practice*, *42*(1), 30-41.

Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed methods sequential explanatory design: From theory to practice. *Field Methods, 18*(1), 3-20.

Jimerson, L. (2005). Placism in NCLB –How rural children are left behind. *Equity and Excellence in Education, 38*(3): 211-219.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 14-26.

Johnson, B., & Turner, L. A. (2003). Data collection strategies in mixed methods research. In: A. Tashakkori & C. Teddlie (Eds.), *Handbook on mixed methods in the behavioral and social sciences* (pp. 297-320). Thousand Oaks, CA: Sage Publications.

Johnson, D. D., & Johnson, B. (2006). *High stakes: Poverty, testing, and failure in American schools*. New York: Rowman & Littlefield.

Jones, M., Jones, B., & Hargrove, T. (2003). *The unintended consequences of high-stakes testing*. Lanham, MD: Rowman and Littlefield Publishers, Inc.

Kornhaber, M., & Orfield, G. (2001). *Raising standards or raising barriers? Inequality and high-stakes testing in public education*. New York: Century Foundation Press.

Lane, S., & Stone, C. (2002). Strategies for examining the consequences of assessment and accountability programs. *Educational Measurement: Issues and Practice, 21*(1), 23-30.

Lane, S. (2004). Validity of high-stakes assessment: Are students engaged in complex Thinking? *Educational Measurement: Issues and Practice, 23*(3), 6-14.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. Beverly Hills: CA: Sage.

Lipman, P. (2004). *High stakes education: Inequality, globalization, and urban school reform*. New York: Routledge Palmer.

Lloyd, G. M. (2007). Strategic compromise: A student teacher's design of kindergarten mathematics instruction in a high-stakes testing climate. *Journal of Teacher Education, 58*(4), 328-347.

Marchant, G. (2004). What is at stake with high stakes testing? A discussion of issues and research. *Ohio Journal of Science*. *104*(2), 2-7.

Matthews, J., & Argetsinger, A. (2000). Cheating on rise along with testing: Teachers complain of pressure to do well on standardized exams. *Washington Post*, 2 June, A1.

Maxcy, S. J. (2003). Pragmatic threads in mixed methods research in the social sciences: The search for multiple modes of inquiry and the end of the philosophy of formalism. In A. Tashakkori & C. Teddlie (Eds.), *Handbook on mixed methods in the behavioral and social sciences* (pp. 51-89). Thousand Oaks, CA: Sage.

Maxwell, J. (2005). *Qualitative research design*. Thousand Oaks, CA: Sage.

McNair, S., Bhargava, A., Adams, L., Edgerton, S., & Bess, K. (2003). Teachers speak out on assessment practices. *Early Childhood Education Journal, 31*(1), 23-31.

McNeil, L. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.

McNeil, L. M., & Valenzela, A. (2001). The harmful impact of the TAAS system of testing in Texas: Beneath the accountability rhetoric. In M. Kornhaber & G. Orfield (Eds.), *Raising standards or raising barriers? Inequality and high stakes testing in public education* (pp. 127-150). New York: Century Foundation.

Menken, K. (2008). *English learners left behind: Standardized testing as language policy*. Cleveland, UK: Multilingual Matters.

Mertler, C. A., & Vannatta, R. A. (2005). *Advanced and multivariate statistical methods* (2nd ed.). Los Angeles, CA: Pyrczak.

Moon T., Callahan, C., & Tomlinson, C. (2003). Effects of state testing programs on elementary schools with high concentrations of student poverty-Good news or bad news? *Current Issues in Education , 6*(8), 1-25. Retrieved February 18, 2006 from http://cie.asu.edu/volume6/number8

Morse, J., & Richards, L. (2002). *Readme first: For a user's guide to qualitative methods*. Thousand Oaks, CA: Sage Publications.

Mulvenon, S., Stegman, C., & Ritter, G. (2005). Test Anxiety: A multifaceted study on the perceptions of teachers, principals, counselors, students, and parents. *International Journal of Testing, 5*(1), 37-61.

National Board for Professional Teaching Standards. (2010). Generalist/ Early Childhood Portfolio Instructions. Retrieved May 10, 2010 from http://www.nbpts.org /userfiles /File/EC_Gen_Portfolio_Instructions.pdf

National Center for Education Statistics (2007). *The nation's report card: America's high school graduates-Results from the 2005 National Assessment of Educational progress high school transcript study*. Washington, DC: U.S. Government Printing Office.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw Hill.

O'Neill, P. (2003). High stakes testing law and litigation. *Brigham Young University Education and Law Journal, 2*, 623-662.

Onwuegbuzie, A., & Johnson, B. (2006). The validity issue in mixed research. *Research in the Schools, 13*(1), 48-63.

Onwuegbuzie, A. J., & Teddlie, C. (2003). A framework for analyzing data in mixed methods research. In A. Tashakkori and C. Teddlie (Eds.), *Handbook on mixed methods in the behavioral and social sciences* (pp. 351-384). Thousand Oaks, CA: Sage.

Onwuegbuzie, A., & Collins, K. (2007). A typology of mixed methods sampling designs in social science research. *The Qualitative Report*, *12*(2), 281-316. Retrieved October 23, 2008 from http://www.nova.edu/ssss/QR/QR12-2/onwuegbuzie2.pdf

Orlich, D. (2004). No child left behind an illogical accountability model. *Clearing House, 78*(1), 6-11.

Paris, S. (2000). Trojan horse in the schoolyard. *Issues in Education*, *6*(1/2), 1-12.

Paris, S., & McEvoy, E. (2000). Harmful and enduring effects of high-stakes testing. *Issues in Education, 6*(1/2), 145-200.

Paris, S., & Urdan, T. (2000). Policies and practices of high-stakes testing that influence teachers and schools. *Issues in Education, 6*(1/2), 83-108.

Parke, C., Lane, S., & Stone, C. (2006). Impact of a state performance assessment program in reading and writing. *Educational Research and Evaluation*, *12*(3), 239-269.

Pennington, J. (2004). Teaching interrupted: The effect of high-stakes testing on literacy instruction in a Texas elementary school. In F.B. Boyd & C.H. Brock (Eds.), *Multicultural and multilingual literacy and language*. New York: Guilford.

Plano Clark, V., & Creswell, J. (2008). *The mixed methods reader*. Thousand Oaks, CA: Sage Publications.

Popham, J. (2001). *The truth about testing*. Alexandria, VA: Association for Supervision and Curriculum Development.

Pringle, R. M., & Martin, S. C. (2005). The potential impacts of upcoming high-stakes testing on the teaching of science in elementary classrooms. *Research in Science Education*, *35*, 347-361.

Punch, K. F. (1998). Mixed methods and evaluative criteria. In K. F. Punch, *Introduction to social research* (pp.239-250). Thousand Oaks, CA: Sage Publications.

Rock, T., Heafner, K., O'Connor, J., Passe, S. Oldendorf, A., Good, et al. (2006). One state closer to a national crisis: A report on elementary social studies education in North Carolina schools. *Theory and Research in Social Education*, *34*(4), 455-483.

Roderick, M., & Nagaoka, J. (2005). Retention under Chicago's high-stakes testing program: Helpful, harmful, or harmless? *Educational Evaluation and Policy Analysis*, *27*(4), 309-340.

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 181-210.

Sandelowski, M. (2001). Real qualitative researchers don't count: The use of numbers in qualitative research. *Research in Nursing & Health, 24*, 230-240.

Savage, T. (2003). Assessment and quality social studies. *The Social Studies, 94*(5), 201-206.

Shepard, L. (2002). The hazards of high-stakes testing. *Issues in Science and Technology, 19*(2), 53-58.

Sloane, F., & Kelly, A. (2003). Issues in high-stakes testing programs. *Theory Into Practice, 42*(1), 12-17.

Solano-Flores, G., & Trumball, E. (2003). Examining language in context: The need for new research paradigms in the testing of English-language learners. *Educational Researcher, 32*(2), 3-13.

Stecher, B. M., & Hamilton, L. S. (2002). Putting theory to the test, RAND Review, Spring, 17-23. Retrieved July 18 2008, from http://www.rand.org/ publications/randreview/issues/rr.04.02/theory.html

Stiggins, R., & Chappuis, J. (2005). Using student-involved classroom assessment to close assessment to close achievement gaps. *Theory Into Practice, 44*(1), 11-18.

Tabachnick, B., & Fidell, L. (2007). *Using multivariate statistics*. Boston, MA: Pearson.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches.* Thousand Oaks, CA: Sage.

Tashakkori, A., & Teddlie, C. (Eds.). (2003). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks: Sage.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences.* Thousand Oaks: SAGE.

Thompson, S., & Thurlow, M. (2001). State special education outcomes: A report on state activities at the beginning of a new decade. Minneapolis, Minn.: University of Minnesota, National Center on Educational Outcomes. Retrieved February 3, 2010 from Education.umn.edu/NCEO/OnlinePubs/2001StateReport.html

Thorndike, R. M. (2005). *Measurement and evaluation in psychology and education* (7th ed). Columbus, OH: Merill Prentice Hall.

Triplett, C. F., & Barksdale, M. A. (2005). Third through sixth graders' perceptions of high-stakes testing. *Journal of Literacy Research, 37*(2), 237-260.

Tschannen-Moran, M., & Hoy, A. (2001). Teacher efficacy: Capturing and elusive construct. *Teaching and Teacher Education, 17*, 783-805.

United States Department of Education. (2010). Accountability. Office of Elementary and Secondary Education. Washington, D.C., Retrieved January 10, 2010, from http://www.ed.gov/nclb/accountability/index.html

VanFossen, P. (2005). "Reading and math take so much of the time…": An overview of social studies instruction in elementary classrooms in Indiana. *Theory and Research in Social Education, 33*(3), 376-403.

Vogler. K. (2006). Impact of a high school graduation examination on Mississippi social studies teachers' instructional practices. *In Measuring history: Cases of state-level testing across the United States*, ed. S. G. Grant, 273-302. Greenwich, CT: Information Age.

Vogler, K., & Virtue, D. (2007). "Just the facts, ma'am": Teaching social studies in the era of standards and high-stakes testing. *The Social Studies, 98*(2), 54-58.

Wolf, P. (2007). Academic improvement through regular assessment. *Peabody Journal of Education, 82*(4), 690-702.

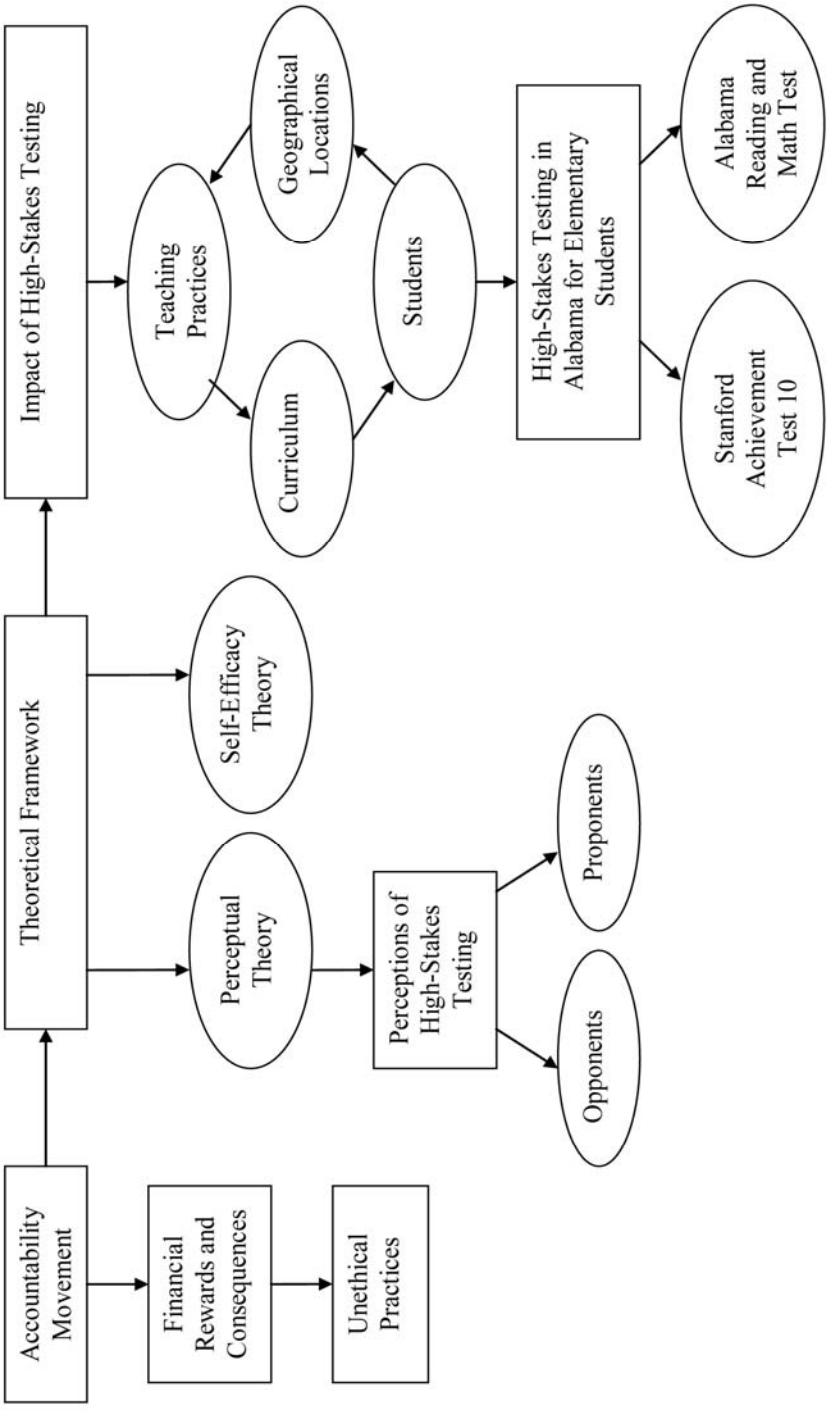
Yarbrough, T. (1999). *Teacher perceptions of the North Carolina ABC Program and the relationship to classroom practice*. Ph.D. dissertation, University of North Carolina at Chapel Hill.

Yin, R. K. (2006). Mixed methods research: Are the methods genuinely integrated or merely parallel? *Research in the Schools, 13*(1), 41-47.

Zimmerman, B., & Dibenedetto, M. (2008). Mastery learning and assessment: Implications for students and teachers in an era of high-stakes testing. *Psychology in the Schools, 45*(3), 206-216.

APPENDIX A

CONCEPTUAL FRAMEWORK FOR STUDY


Accountability Movement
Financial Rewards and Consequences
Unethical Practices
Theoretical Framework
Self-Efficacy Theory
Perceptual Theory
Perceptions of High-Stakes Testing
Proponents
Opponents
Impact of High-Stakes Testing
Teaching Practices
Geographical Locations
Curriculum
Students
High-Stakes Testing in Alabama for Elementary Students
Alabama Reading and Math Test
Stanford Achievement Test 10

APPENDIX B

PERMISSION TO USE QUESTIONNAIRE

CURRY
UNIVERSITY of VIRGINIA

Curry School of Education
405 Emmet Street South
P.O. Box 400265
Charlottesville, VA 22904-4265
Phone: 434-924-3160
Fax: 434-924-3866
www.curry.edschool.virginia.edu

Department of Leadership, Foundations, and Policy

June 5, 2006

Tonya Borden-Hudson
1144 Forest Lakes Way
Sterrett, AL 35147

Dear Ms. Borden-Hudson:

This letter is giving you permission to use the high-stakes questionnaire that was developed at the University of Virginia. If you make any changes to the survey, please provide a copy of the changes.

Good luck with your studies and your research.

Regards,

Tonya R. Moon

Tonya R. Moon, Ph.D.
Associate Professor

APPENDIX C

MODIFIED ELEMENTARY TEACHER QUESTIONNAIRE

Modified Elementary Teacher Questionnaire

(State Testing Program)

1. In what community do you teach: _____rural _____urban _____suburban

2. Circle the grade(s) you currently teach: 3 4 5

3. Circle the grade(s) you taught last year: K 1 2 3 4 5 6

4. Years you've been teaching (including this year): _____ years

5. Years you've been teaching at this school (including this year): _____years

6. Please indicate your estimate of your class's academic achievement/ability level.

   ____Much above grade level ____Somewhat above grade level

   ____At grade level ____Somewhat below grade level

   ____Much below grade level

7. How would you characterize the emphasis on the outcomes of the SAT-10 and the ARMT in your school during the past year?

   _____extremely high _____very high _____moderate

   _____low _____no emphasis

8. To what extent do you agree the following curriculum and instructional approaches are affected by the need to increase SAT-10 and ARMT scores? (*Select the option that best describes your choice*.)

	Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
a. Focusing clearly and consistently on the topics covered on the SAT-10 and ARMT					
b. Using an interdisciplinary curriculum					
c. Differentiating the curriculum (e.g. using tiered assignment, lessons targeted to student past achievements, accelerating pace of learning for some students)					
d. Doing hands-on work					
e. Using alternative (e.g. performance) assessments					
f. Providing broad coverage across many areas of basic knowledge					
g. Teaching which allows for in-depth exploration where one critical exemplar of a concept or principle can be understood as a basis to generalize to other exemplars					
h. Making sure the content and skills covered on the SAT-10 and ARMT are reviewed prior to the test administration					
i. SAT-10 and ARMT help clarify and specify learning goals					
j. SAT-10 and ARMT tests give me important feedback about how well I am teaching the curricular area(s)					
k. I teach to the SAT-10 and ARMT more than I normally would					
l. I omit certain information because there is not enough time to fit it in because of the SAT-10 and ARMT					
m. I do not do certain things that look interesting or beneficial for students unless they are on the SAT-10 and ARMT					
n. I do not do anything differently because of the SAT-10 and ARMT					

9. To what extent do you agree on how frequently the following practices are used in your classroom? (*Select the option that best describes your choice*.)

	Daily	Twice a Week	Once a Week	Once a Month	None
a. Use of constructed response items (short essays)					
b. Use of multiple-choice items					
c. Use of long term projects (e.g., research or other projects requiring a week or more)					
d. Use of performance type items (e.g., presentations, science experiments)					
e. Test preparation (e.g., homework and classwork)					
f. Making sure the content and skills covered on the SAT-10 and ARMT are reviewed prior to the test administration					
g. Adjusting the curriculum sequence based on the content coverage of the SAT-10 and ARMT					

10. To what extent do you agree with how much attention you are able to give to the following curricular areas in your classroom? (*Select the option that best describes your choice*.)

	Daily	Twice a Week	Once a Week	Once a Month	None
a. Higher-order thinking skills					
b. Problem-solving skills					
c. Topics which are not assessed on the SAT-10 and the ARMT					
d. The fine and performing arts (e.g., music, art, drama)					
e. Basic skills (e.g., computations, grammar, vocabulary					
f. Factual knowledge					
g. Enrichment or extension of the curriculum according to student interest and/or ability to delve deeper					

11. To what extent do you agree with how often you use the following test preparation activities during the **first 1/3** of the year? (*Select the option that best describes your choice*.)

	Daily	Twice a Week	Once a Week	Once a Month	None
a. Student worksheets					
b. Instruction for students on test-taking strategies					
c. Review/practice using state released test items					
d. Student practice in the kinds of item formats that are on the SAT-10 and the ARMT					

during the **second 1/3** of the year?

	Daily	Twice a Week	Once a Week	Once a Month	None
e. Student worksheets					
f. Instruction for students on test-taking strategies					
g. Review/practice using state released test items					
h. Student practice in the kinds of item formats that are on the SAT-10 and the ARMT					

during the **month prior** of the test?

	Daily	Twice a Week	Once a Week	Once a Month	None
i. Student worksheets					
j. Instruction for students on test-taking strategies					
k. Review/practice using state released test items					
l. Student practice in the kinds of item formats that are on the SAT-10 and the ARMT					

during the **period following** SAT-10 and ARMT testing?

	Daily	Twice a Week	Once a Week	Once a Month	None
m. Student worksheets					
n. Instruction for students on test-taking strategies					
o. Review/practice using state released test items					
p. Student practice in the kinds of item formats that are on the SAT-10 and the ARMT					

12. To what extent do you agree with how often during the year your school administration engages in the following activities with teachers? (*Select the option that best describes your choice*.)

	Daily	Twice a Week	Once a Week	Once a Month	None
a. Reviews test scores at staff meetings					
b. Discusses ways to improve test scores					
c. Provides materials to improve test scores					
d. Checks to see that teachers are emphasizing areas which showed weakness from past test results					
e. Introduces or discusses important new instructional ideas					

13. To what extent do you agree students are affected by the need to increase scores on the SAT-10 and ARMT? (*Select the option that best describes your choice*.)

	Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
a. Test results are an accurate picture of student learning					
b. Many of the students I teach are not capable of learning the material on the SAT-10 and ARMT					
c. SAT-10 and ARMT testing is improving student learning					
d. Students are treated as test-takers rather than learners					
e. Students are under too much pressure to increase test scores					
f. Students see learning as a chore because of pressure from SAT-10 and ARMT testing					
g. Students feel bad if they do not have high test scores					

14. To what extent do you agree teachers are affected by the need to increase scores on the SAT-10 and ARMT? (*Select the option that best describes your choice*.)

	Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
a. Having to prepare students for the SAT-10 and ARMT impacts my approach to teaching					
b. My current students' most recent test results impact my approach to teaching					
c. Our school's overall test results impact my approach to teaching					
d. Learning outcomes measured by SAT-10 and ARMT are the most important ones to measure					
e. My school is more interested in increasing student test scores on the SAT-10 and ARMT than in improving overall student learning					
f. Teachers in my school feel there is discrepancy between what they think should be taught and what the SAT-10 and ARMT emphasize					
g. Teachers are under too much pressure to increase test scores					

15. If you have any additional comments on the way SAT-10 and ARMT testing is helping or hurting your instructional practices, please use the space below.

Using modified version of the survey developed by Moon et al. in
Moon T., Callahan, C. & Tomlinson, C. (2003). Effects of state testing programs on elementary schools with high concentrations of student poverty- Good news or bad news? *Current Issues in Education* [Electronic Version], 6(8), 1- 25. Retrieved February 18, 2006 from http://cie.asu.edu/volume6/number8.

APPENDIX D

INSTITUTIONAL REVIEW BOARD FOR HUMAN USE APPROVAL

THE UNIVERSITY OF ALABAMA AT BIRMINGHAM

Institutional Review Board for Human Use

Form 4: IRB Approval Form
Identification and Certification of Research
Projects Involving Human Subjects

UAB's Institutional Review Boards for Human Use (IRBs) have an approved Federalwide Assurance with the Office for Human Research Protections (OHRP). The UAB IRBs are also in compliance with 21 CFR Parts 50 and 56 and ICH GCP Guidelines. The Assurance became effective on November 24, 2003 and expires on October 26, 2010. The Assurance number is FWA00005960.

Principal Investigator:	BORDEN-HUDSON, LATONYA
Co-Investigator(s):	IVANKOVA, NATALIYA V SEARBY, LINDA
Protocol Number:	**X081208006**
Protocol Title:	*Examining Teachers' Perceptions of the Impact of High-stakes Testing on Classroom Teaching Practices: A Mixed Methods Study*

The IRB reviewed and approved the above named project on 12/23/08. The review was conducted in accordance with UAB's Assurance of Compliance approved by the Department of Health and Human Services. This Project will be subject to Annual continuing review as provided in that Assurance.

This project received EXPEDITED review.

IRB Approval Date: 12-23-08

Date IRB Approval Issued: 12/23/08

Marilyn Doss

Marilyn Doss, M.A.
Vice Chair of the Institutional Review
Board for Human Use (IRB)

Investigators please note:

The IRB approved consent form used in the study must contain the IRB approval date and expiration date.

IRB approval is given for one year unless otherwise noted. For projects subject to annual review research activities may not continue past the one year anniversary of the IRB approval date.

Any modifications in the study methodology, protocol and/or consent form must be submitted for review and approval to the IRB prior to implementation.

Adverse Events and/or unanticipated risks to subjects or others at UAB or other participating institutions must be reported promptly to the IRB.

470 Administration Building
701 20th Street South
205.934.3789
Fax 205.934.1301
irb@uab.edu

The University of
Alabama at Birmingham
Mailing Address:
AB 470
1530 3RD AVE S
BIRMINGHAM AL 35294-0104

APPENDIX E

INSTITUTIONAL REVIEW BOARD PROTOCOL RENEWAL

THE UNIVERSITY OF ALABAMA AT BIRMINGHAM

Institutional Review Board for Human Use

Form 4: IRB Approval Form
Identification and Certification of Research
Projects Involving Human Subjects

UAB's Institutional Review Boards for Human Use (IRBs) have an approved Federalwide Assurance with the Office for Human Research Protections (OHRP). The Assurance number is FWA00005960 and it expires on October 26, 2010. The UAB IRBs are also in compliance with 21 CFR Parts 50 and 56 and ICH GCP Guidelines.

Principal Investigator:	BORDEN-HUDSON, LATONYA
Co-Investigator(s):	
Protocol Number:	**X081208006**
Protocol Title:	*Examining Teachers' Perceptions of the Impact of High-stakes Testing on Classroom Teaching Practices: A Mixed Methods Study*

The IRB reviewed and approved the above named project on 11-17-09. The review was conducted in accordance with UAB's Assurance of Compliance approved by the Department of Health and Human Services. This Project will be subject to Annual continuing review as provided in that Assurance.

This project received EXPEDITED review.

IRB Approval Date: 11-17-09

Date IRB Approval Issued: 11-17-09

Marilyn Doss

Marilyn Doss, M.A.
Vice Chair of the Institutional Review
Board for Human Use (IRB)

Investigators please note:

The IRB approved consent form used in the study must contain the IRB approval date and expiration date.

IRB approval is given for one year unless otherwise noted. For projects subject to annual review research activities may not continue past the one year anniversary of the IRB approval date.

Any modifications in the study methodology, protocol and/or consent form must be submitted for review and approval to the IRB prior to implementation.

Adverse Events and/or unanticipated risks to subjects or others at UAB or other participating institutions must be reported promptly to the IRB.

470 Administration Building
701 20th Street South
205.934.3789
Fax 205.934.1301
irb@uab.edu

The University of
Alabama at Birmingham
Mailing Address:
AB 470
1530 3RD AVE S
BIRMINGHAM AL 35294-0104

APPENDIX F

RECRUITMENT LETTER

January 21, 2009

Dear Teacher,

My name is LaTonya Borden-Hudson. I am currently a first grade teacher, and I am also a doctoral student at UAB. As my final research project, I will examine teachers' perceptions of the impact of high-stakes testing on classroom teaching practices. In order for me to do this study, I will need your participation. To participate please sign the consent forms, and also obtain the signature of a witness. Next, complete the survey which will only take up to 15 minutes. Please place one of the signed consent forms in the self-addressed stamped envelope along with the completed survey. Please keep one of the consent forms for your records. Some of the participants will be contacted later for follow-up interviews that will last for up to 60 minutes.

I am sending two consent forms and a survey to each third through fifth grade teacher in three school systems. The scores from the survey will be compared to teachers in your school system as well as to teachers in two other school systems.

After surveys have been completed and returned, I will select 9 – 12 participants to be interviewed. The interview will increase the understanding of teachers' responses to the study.

I would like to ensure you that all names and personal information will remain confidential. Numerical codes will be assigned to each survey to keep track of who has completed the survey, but that information will be available only to me. No school names will be given, and no one will be able to identify you in the study. The names of teachers who are willing to participate in an interview will be assigned a letter of the Greek alphabet to protect anonymity.

If you have any questions or concerns, please contact me at hudson@uab.edu. Thank you in advance for your time and cooperation.

Sincerely,
LaTonya Borden-Hudson

APPENDIX G

INFORMED CONSENT

Informed Consent

TITLE OF RESEARCH: Examining Teachers' Perceptions of the Impact of High-stakes Testing on Classroom Teaching Practices

IRB PROTOCOL NUMBER: X081208006

INVESTIGATOR: LaTonya Borden-Hudson

SPONSOR: University of Alabama at Birmingham Department of Educational Leadership

Explanation of Procedures

You are being asked to participate in a study. The purpose of this study is to examine teachers' perceptions of the impact of high-stakes testing on classroom teaching practices in grades three, four, and five. Two school systems will be the location of this study. You will be asked to complete the survey that will take about 15 minutes. The scores from the survey will be compared to the scores of other teachers in your school system in addition to the scores of other teachers in another school system. After the analysis of the survey results, 9-12 participants who completed the survey will be asked to participate in a face to face interview to discuss the impact of high-stakes testing on their classroom teaching practices. The participants' responses to the survey will be used as the criteria to select interview participants. The interview will last approximately 60 minutes and will be conducted at the school where you work or at a location chosen by you.

Risks and Discomforts

The risks and discomforts from participating in this study are no greater than the risks and discomforts of daily living.

Benefits

You will receive no direct benefit from participating in this study. The awareness regarding the impact of high-stakes testing on classroom teaching practices could result in greater understanding about the issues facing elementary school teaching practices.

Alternatives

Your participation in the study is strictly voluntary. The only alternative to participating in this study is not to participate.

Participant's Initials:_____

Confidentiality

Your name will not be seen by anyone but the principal investigator. No one else will be able to identify any participants. The only reason your name is needed is to keep track of which teachers have completed the survey and to contact you if you are chosen to participate in the interview phase of this study. The following groups will have access to private information that identifies you by name: the Office for Human Research Protections (OHRP), and the University of Alabama at Birmingham (UAB) Institutional Review Board (IRB). All data will be stored and protected in a locked metal file cabinet in the investigator's home.

Refusal or Withdrawal without Penalty

You are free to withdraw from this study at any time. Your participation in this study may be ended without your consent if it is determined by the investigator that it is in your best interest.

Cost of Participation

There is no cost for participating in this study.

Payment for Participation in Research

You will not be paid for completing the survey. If you complete the survey and are selected for follow-up interviews, you will receive a $20 Visa gift card.

Questions

If you have any questions about the research, LaTonya Borden-Hudson will be glad to answer them. Mrs. Borden-Hudson's number is (205) 515-1877. If you have questions about your rights as a research participant, or concerns or complaints about the research, you may contact Ms. Sheila Moore. Ms. Moore is the Director of the Office of the Institutional Review Board for Human Use (OIRB). Ms. Moore may be reached at (205) 934-3789 or 1-800-822-8816. If calling the toll-free number, press the option for "all other calls" or for an operator/attendant and ask for extension 4-3789. Regular hours for the Office of the IRB are 8:00 a.m. to 5:00 p.m. CT, Monday through Friday. You may also call this number in the event the research staff cannot be reached or you wish to talk to someone else.

Participant's Initials:_____

Legal Rights

You are not waiving any of your legal rights by signing this informed consent document.

Signatures

Your signature below indicates that you agree to participate in this study. You will receive a copy of this signed document.

Signature of Participant Date

Signature of Investigator Date

Signature of Witness Date

APPENDIX H

INTERVIEW PROTOCOL

1. Please tell me about yourself.

   Probes:
   Why did you choose teaching?
   How many years of experience do you have?
   How long have you taught at your current school, and what grade do you currently teach?
   How long have you been involved with SAT-10 and ARMT testing?

2. What are your experiences with SAT-10 and ARMT testing?

   Probes:
   Please describe one example of a positive experience with SAT-10 and ARMT testing.
   Please describe one example of a negative experience with SAT-10 and ARMT testing

3. Please describe any influences SAT-10 and ARMT testing might seem to have on the curriculum.

   Probes:
   What influence does SAT-10 and ARMT testing have on reading, math, social studies, and science. How does it influence the priority of the subjects you teach?

4. How does SAT-10 and ARMT influence the teaching of critical thinking skills?

5. How do you think SAT-10 and ARMT testing influences your teaching practices.

   Probe:
   How does SAT-10 and ARMT testing influence the types of learning activities you plan for students?

6. Please describe any influences SAT-10 and ARMT testing might seem to have on the amount of time you spend on test preparation in the classroom and with administrators.

   Probe:
   Please describe examples of the types of test preparation in which students are engaged or in which you are involved in as a teacher or in faculty meetings.

7. How has SAT-10 and ARMT testing influenced you as a teacher?

8. How has SAT-10 and ARMT testing influenced your students?

9. How does SAT-10 and ARMT testing influence student's motivation for learning?

10. How does SAT-10 and ARMT testing influence certain populations such as special education students, ELLs, and minorities?

11. Is there any other information you would like to share regarding your experiences with SAT-10 and ARMT testing?