STANDARDS-BASED ASSESSMENT AND HIGH STAKES TESTING:

ACCURACY OF STANDARDS-BASED GRADING

by

Amity C. Hardegree

Liberty University

A Dissertation Presented in Partial Fulfillment
Of the Requirements for the Degree
Doctor of Education

Liberty University

July, 2012

STANDARDS-BASED ASSESSMENT AND HIGH STAKES TESTING:

ACCURACY OF STANDARDS-BASED GRADING

by Amity Christa Hardegree

A Dissertation Presented in Partial Fulfillment
Of the Requirements for the Degree
Doctor of Education

Liberty University, Lynchburg, VA

July, 2012

APPROVED BY:

Kenneth Gossett, Ph.D., Committee Chair

April Howard, Ed.D., Committee Member

Rebecca Harrison, Ed.D., Committee Member

Scott B. Watson, Ph.D., Associate Dean, Advanced Programs

ABSTRACT

Amity C. Hardegree. STANDARDS-BASED ASSESSMENT AND HIGH STAKES TESTING: ACCURACY OF STANDARDS-BASED GRADING. (Under the direction of Dr. Kenneth Gossett) School of Education, Liberty University, Summer, 2012. This quantitative study examines whether standards-based grade reporting accurately informs student academic achievement on standardized criterion-referenced tests for all students. The participants for this study were all fifth graders enrolled in eight elementary schools in a rural system in north Georgia from 2009-2010. Approximately 550 students' standards-based report cards (SBRC) and Criterion Referenced Competency Tests (CRCT) provide the data to determine whether grades on standards-based report cards provide accurate information for all students, regardless of gender, ELL status, or socioeconomic status by comparing mean scores on Georgia's CRCT in the areas of math and reading, based on SBRC indicators. The findings of this study provide strong implications for school systems considering a standards-based grading reporting system in response to the recent movement towards standards-based curriculums. The results show alignment between indicators on standards-based assessment and scores on criterion-referenced standardized tests, used as an indicator for AYP (Adequate Yearly Progress), adding to the body of research on the effectiveness of standards-based grading in showing student mastery of curriculum standards. For math, even with the same SBRC score, students who are on the free/reduced lunch program tend to score lower than students who are not. For reading, even with the same SBRC score, females tend to score higher than males, those with limited English proficiency

tend to score lower than those who are proficient, and those who are on the free/reduced lunch program tend to score lower than those who pay for lunch. The study provides evidence to suggest that standards-based grade reporting provides accurate information regarding student learning that can be used as a measure of student achievement.

Descriptors: Standards-based, Assessment, High-stakes testing, Standards-based grading

Dedication

I dedicate this manuscript to my husband, Steven, who has been my faithful supporter and partner in this journey, and to my children, Rachel, Grant, and Ella who are my inspiration and my joy. I also dedicate this work to my parents, who taught me early on to embrace my dreams and God's promises. Finally, I dedicate this paper to the memory of Dr. Jill Jones, who with her radiant spirit and unwavering faith, reminded me of all the doors God has opened, and whose character and passion I will always aspire to emulate.

Acknowledgements

First, I would like to thank my Lord and Savior Jesus Christ who gives me the strength daily to go about the tasks He has so graciously entrusted to me.

Next, I would like to thank Dr. Kenneth Gossett, my committee chair, for his guidance and support throughout this process. I would also like to thank committee members Dr. April Howard and Dr. Rebecca Harrison and research consultant Dr. Scott Watson for sharing their time, expertise, and wisdom to steer me towards completion. I would like to extend sincere thanks to Dr. Kim Love-Myers and the UGA Statistical Consulting Center for their support throughout the project. Next, I would like to thank my district support, especially Kathy Elrod, Kathy Miller, Rise Hawley, Teresa Strickland, and Miranda Storey for the many ways they helped me develop a focus, collect the data, and process the information throughout the journey. I could not have completed this research project without their direction and support. I would also like to thank the faculty and staff at my school as well as my Liberty cohort group, Mishea Griffeth, Derelle McMenomy, and Steven Hardegree, for being there through all of the ups and downs and for providing valuable suggestions and insights.

Finally, I would like to thank my family and friends who believed in me, encouraged me, and helped care for my children throughout this journey. I truly could not have accomplished a task of this magnitude without such a wonderful support system. And to my husband and children, thank you for your sacrifices and unconditional love. You will always be my greatest accomplishment!

TABLE OF CONTENTS

List of Tables

CHAPTER ONE: INTRODUCTION

Now that standards-based instruction is firmly in place, educators must rethink their assessment practices and determine how to grade and report student learning in terms of the standards. To enrich overall assessment and to ensure mastery of standards, many schools have eliminated traditional report cards and are moving towards Standards Based Report Cards (SBRC). The idea is that Standards Based Report cards provide teachers, parents, and students a clearer picture of student mastery of standards and keep teaching and learning focused on the performance goals. Schools in Georgia measure mastery of standards through the Criterion-Referenced Competency Test (CRCT), and this standardized test also serves as one indicator of adequate yearly progress (AYP).

This non-experimental causal comparative study examines whether grades on standards-based report cards provide accurate information for all students, regardless of gender, ELL status, or socioeconomic status by comparing mean scores on Georgia's CRCT in the areas of math and reading, based on SBRC indicators.

Background

In an effort to close the achievement gap between low performing and high performing schools, public education in the United states has focused attention on developing academic content standards, and using assessment and accountability to make sure that students are not just moving through the system, but that they are learning these core standards (Guskey, 2005). Through the recent comprehensive assessment Grant included in the Federal Race to the Top Assessment Program, President Obama and congress (2010) encourage schools to rethink their assessment programs. According to U.S. Department of Education, the Race to the Top

Assessment Program authorized under the American Recovery and Reinvestment Act of 2009 (ARRA) seeks to accomplish the following:

> …to develop assessments that are valid, support and inform instruction, provide accurate information about what students know and can do, and measure student achievement against standards designed to ensure that all students gain the knowledge and skills needed to succeed in college and the workplace. These assessments are intended to play a critical role in educational systems; provide administrators, educators, parents, and students with the data and information needed to continuously improve teaching and learning; and help meet the President's goal of restoring, by 2020, the nation's position as the world leader in college graduates. (US Dept of Ed. 2010).

Clearly, school reform continues to evolve. The *Elementary and Secondary Education Act of 1965* was reauthorized in 2001 and again in 2007, and renamed The *No Child Left Behind Act*, requiring states to use content standards and show "adequate yearly progress" (AYP) towards meeting improvement goals. According to Guskey (2001), one of the key changes that school reform has recently brought is the need to measure student achievement in terms of the learning standards, rather than by comparing students to each other, essentially moving from a norm-referenced system to a criterion-referenced grading system. While the effectiveness of school reform remains controversial, the impact of standards-based education and assessment affects school districts across the nation, and ultimately seeks to improve teaching and learning. This paradigm shift in assessment leads educators to reevaluate assessment practices to make sure that assessment is being used, not just to label, but also to inform instruction, and to provide a true reflection of student learning.

As educators seek to use assessment as a tool for diagnosing and informing instruction, they move away from comparing students to each other, and instead focus on the student's progress toward meeting academic standards. The research shows the need for more specific feedback in terms of standards and highlights the ineffectiveness of using large-scale assessments that rank order students due to the lack of usable feedback for teachers (Guskey, 2003, Barton, 2002, Kifer, 2001). Standards-based assessment moves from a ranking system to a system that is more sensitive to the diverse needs of students, but with a firm connection to classroom learning goals. Assessments can take a variety of forms, from observations to formal assessments and performance tasks. Formative assessments are used to diagnose a student's skill level and help prescribe further instruction, helping to ensure that instruction is differentiated to meet students' needs. Summative assessments follow instruction and are designed to show mastery of a standard. In order to show most recent evidence of learning, students are given multiple opportunities to show learning, and grades are not averaged, but rather replaced by the most recent evidence. Therefore, students are not required to learn at the same pace and in the same way, but are allowed some flexibility to revisit challenging concepts or objectives. By focusing on student learning, standards-based grading eliminates ranking or ordering students, and instead attempts to provide an accurate picture of student learning, free from non-academic factors such as behavior, attendance, or effort.

In the state of Georgia, the CRCT (Criterion Referenced Competency Test) provides one measure for student achievement of the content standards. While supporters of the standards-based curriculum and standardized testing claim that administering such a test ensures that all students attain an acceptable level of understanding on centralized objectives, opponents claim that this standardized approach narrows teaching and reduces opportunities for deeper learning

and critical thinking. Despite controversy over NCLB and standardized testing, standards-based curriculums and standards-based assessment are realities for public schools across the nation, and school systems seek the best way to structure instruction and assessment to maximize student achievement within the constraints of a limited budget, limited time, and limited resources.

Problem Statement

As school reform focuses on standards and developing assessment practices to provide educators with the information needed to continuously improve teaching and learning, teachers face the challenge of reporting student learning in terms of the standards, rather than showing how students compare to their classmates. To enrich overall assessment and to ensure mastery of standards, many schools have eliminated traditional report cards and are moving towards Standards Based Report Cards (SBRC.) Most educators agree that traditional grading methods have been inadequate in communicating student learning and that standards-based grading provides more specific and useful information to teachers, parents, and students about student mastery of standards (Marzano, 2000, Guskey, 2001). Focusing grading on standards, rather than comparing students to their classmates seems a natural follow up to standards based instruction, and should, ideally, lead teachers to better instructional methods and improve student achievement as shown by standardized test scores. School systems considering standards-based grading seek a less subjective and more informative method of grade reporting that truly reflects student learning and mastery of standards as measured by Georgia's CRCT, a standardized test given to student's grades three through eight in the state of Georgia.

Traditional grading systems have been inconsistent in accurately conveying what students know and are able to do in relation to learning standards. In a traditional system, grades have

been determined by averaging a variety of factors, possibly including effort, organization, or behavior. According to Reeves (2011), in traditional systems using letter grades, grades are only slightly related to performance on external assessments. Reeves also notes the importance of making sure that grading systems are fair: "What we describe as proficient performance must truly be a function of performance, and not gender, ethnicity, or socioeconomic status" (Reeves, 2001, p.1). In today's accountability driven educational system, grading systems must reflect student learning of content standards and provide an accurate measure of what students know, understand and are able to do. In an effort to improve both instruction and assessment, many systems have moved to a standards-based system of grading. However, little research has been done regarding the effectiveness of standards-based grading in assessing student mastery of learning standards in a way that provides a reliable indicator of student performance on high-stakes accountability measures, such as criterion-referenced competency tests. Furthermore, research is needed to determine whether standards-based grading provides a system that is fair to all students, regardless of gender, ELL status, ethnicity, socio-economic status, or grade level. This study examines whether grades on standards-based report cards provide accurate information for all students by comparing mean scores on Georgia's CRCT in the areas of math and reading, based on SBRC indicators.

Purpose Statement

The purpose of this study is to determine the effectiveness of standards-based grade reporting in providing accurate indicators of success for all students, regardless of gender, ethnicity, grade, or socioeconomic status, on high-stakes standardized tests designed to show mastery of academic standards and adequate yearly progress. Assessment is central to the educational reform movement that focuses on a standardized curriculum and accountability.

Research has shown that formative assessment and standards-based grading are key components to improving student achievement. Assessment that reemphasizes the standards and provides feedback to students and teachers facilitates learning (Bloom, Madaus, & Hastings, 1981, Stiggins, 2002). Researchers have also highlighted the possible negative outcomes of over-testing or narrowing the curriculum to focus on a narrow set of standards assessed in high-stakes standardized tests (Brown, 1992, 1993; Romberg, 1989). However, little research has been done on how standards-based grading corresponds to student performance on criterion referenced standardized tests.

Many school systems across the country are moving to a standards-based system of grading in an effort to improve assessment and ensure mastery of standards at each grade level. The study of standards-based grading and grade reporting is needed to determine whether such a grading system is an effective measure of student mastery of standards, as reported on criterion referenced competency test. As systems move away from traditional norm-referenced grading systems, it is important to test the validity and reliability of standards based grading in providing accurate information with respect to student success on standardized tests.

Research Questions and Research Hypotheses

The study will examine the following research questions to determine how standards-based grade reporting accurately informs student performance on the Criterion References Competency Test and to identify any differences in the mean scores based on SBRC indicators, gender or sub-groups.

Research Question 1

Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?

Null Hypothesis 1

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for students in grade five will not be significantly different among students who scored Does not Meet, Meets, and Exceeds on standards-based report cards.

Alternative Hypothesis 1

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Research Question 2

Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?

Null Hypothesis 2

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Alternative Hypothesis 2

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Research Question 3

Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade ELL students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade ELL students different from the differences in mean scores for fifth grade non-ELL students?

Null Hypothesis 3

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for ELL students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for ELL students and non-ELL students.

Alternative Hypothesis 3

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for ELL students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The pattern of these differences for ELL students will be different from the pattern of non-ELL students.

Research Question 4

Are the mean scores on the reading section of the Georgia Criterion Referenced Competency

Test different for fifth grade ELL students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade ELL students different from the differences in mean scores for fifth grade non-ELL students?

Null Hypothesis 4

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for ELL students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for ELL students and non-ELL students.

Alternative Hypothesis 4

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for ELL students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The pattern of these differences for ELL students will be different from the pattern of non-ELL students.

Research Question 5

Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade students who receive free or reduced lunch with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade students who receive free or reduced lunch different from the differences in mean scores for fifth grade students who do not receive free or reduced lunch?

Null Hypothesis 5

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for students who receive free or reduced lunch in grade five will

not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for students who do and do not receive free and reduced lunch.

Alternative Hypothesis 5

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for students who receive free or reduced lunch in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The pattern of these differences for students who receive free or reduced lunch will be different from the pattern of students who do not receive free or reduced lunch.

Research Question 6

Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade students who receive free or reduced lunch with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade students who receive free or reduced lunch different from the differences in mean scores for fifth grade students who do not receive free or reduced lunch?

Null Hypothesis 6

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for students who receive free or reduced lunch in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for students who do and do not receive free and reduced lunch.

Alternative Hypothesis 6

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for students who receive free or reduced lunch in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The pattern of these differences for students who receive free or reduced lunch will be different from the pattern of students who do not receive free or reduced lunch.

Research Question 7

Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade female students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?

Null Hypothesis 7

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for female students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Alternative Hypothesis 7

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for female students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Research Question 8

Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade female students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?

Null Hypothesis 8

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for female students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Alternative Hypothesis 8

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for female students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Research Question 9

Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade male students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade male students different from the differences in mean scores for fifth grade female students?

Null Hypothesis 9

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for male students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for male and female fifth grade students.

Alternative Hypothesis 9

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for male students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The pattern of these differences for male and female students will be different.

Research Question 10

Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade male students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade male students different from the differences in mean scores for fifth grade female students?

Null Hypothesis 10

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for male students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for male and female fifth grade students.

Alternative Hypothesis 10

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for male students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The pattern of these differences for male and female students will be different.

Identification of Variables

Independent variables.

Student gender, male or female, status as an English Language Learner (ELL) or not, and socioeconomic status will serve as independent variables in the study.

Dependent variables.

The dependent variables in the study are the student scores on the math and reading sections of the CRCT in relation to student performance in math and reading on standards-based report card performance indicators.

Definition of key terms.

To ensure clear understanding throughout the study, the following terms and acronyms have been defined:

AYP. Adequate Yearly Progress is a term introduced with the *No Child Left Behind* legislation and indicates whether a school or system has made sufficient progress towards meeting school or system improvement goals.

CRCT. Georgia's Criterion Referenced Competency Test administered to all first through eighth grade students during the years 2001-2010. Beginning in 2011, it will be administered to all third through eighth grade students. The CRCT is considered a high-stakes standardized test that serves as a summative assessment of Georgia Performance Standards and an indicator of AYP. Scores at or above 850 indicate a level of performance that *Exceeds the Standard* set for the test. Scores from 800 to 849 indicate a level of performance that *Meets the Standard* set for the test. Scores below 800 indicate a level of performance that *Does Not Meet the Standard* set for the test (i.e., the state's minimum level of proficiency). Students performing at this level may need additional instructional support.

SBRC. Standards based report card is the grade reporting system used by school systems in which students are scored according to their progress towards meeting specific learning goals or standards. Grades are not reported as an average or letter grade, but rather using a set of indicators such as does not meet, meets, and exceeds the standard.

Significance of the Study

Educators who are considering switching to a standards-based system of grading, or those who have made the transition, may benefit from research that shows the accuracy of SBRC, as measured by student performance on criterion referenced tests. Additionally, such research may indicate the need for additional work in developing performance indicators and rubrics to ensure consistency and reliability in grading. Further understanding is needed in determining how standards based grading might be used to provide accurate information regarding student progress toward meeting standards to help guide instruction and provide information regarding student readiness for high-stakes summative assessments. Furthermore, if standards-based assessments provide accurate information for all students regarding student success or failure on high-stakes testing, this information would guide the decision making of school systems who are considering implementing standards-based grading.

One of the goals of standards-based grade reporting is to reduce subjectivity in grading, thus providing accurate information about student learning that is unaffected by various other factors such as participation, behavior, or parental involvement. An examination of sub-groups is needed to indicate whether standards-based grading truly does provide a system that is fair to all students.

Assumptions and Limitations

Assumptions. The school system has provided adequate professional development in the

use of rubrics, performance tasks, and common assessments as well as ongoing training in reporting student learning on the standards-based report card. Grade level teams, with representatives from each grade level, work collaboratively to develop performance indicators that can be used to assess what students should know, understand, and be able to do. Grade level representatives continue open communication with the teachers at their grade level as well as the local school administration, relaying any concerns or discrepancies in reporting student learning in relation to the standards.

Limitations. In all grading, there is an element of subjectivity as to what degree the student has met or exceeded the standard. Despite efforts at careful training in using rubrics and performance tasks, the study will be subject to some rater variance. Furthermore, because standards-based report cards have only recently been used, many issues are still being resolved. During the implementation of such a system, some teachers become proficient quickly, while others tend to take longer to adjust to a new system of grading. Due to this variance, some reliability in this study may be compromised.

Research Plan

This quantitative study examines whether standards-based grades can provide accurate information with respect to high-stakes tests as measured by Georgia's Criterion Referenced Competency Test (CRCT). Standards-based grades were analyzed to determine whether standards-based report cards provide accurate information for all students by comparing mean scores on Georgia's CRCT in the areas of math and reading, based on SBRC indicators. The Criterion Referenced Competency Test (CRCT), a standardized test given to first through eighth grade students in the state of Georgia, provides a reference point to indicate if the Standards-based Report Card (SBRC) provides an effective measure of student's academic success.

Many school systems across the country are moving to a standards-based system of grading in an effort to improve assessment and ensure mastery of standards at each grade level. The study of standards-based grading and grade reporting is needed to determine whether such a grading system is an effective measure of student mastery of standards, as reported on Georgia's Criterion Referenced Competency Test. As systems move away from traditional norm-referenced grading systems, it is important to examine the validity and reliability of standards based grading in relationships to student success on standards based state assessments to determine whether the grading system provides an accurate tool for measuring student growth. Additionally, there is a need to make sure that the grade reporting system is accurate and fair for all students, regardless of gender, socioeconomic status, and ELL status.

CHAPTER TWO: LITERATURE REVIEW

As reform initiatives seek to improve the quality of instruction and ensure student learning, assessment emerges as a critical component of the educational system. While traditional grading practices often compared students to each other, the U.S. Department of Education emphasizes the need for assessment that supports and informs instruction and provides accurate information about what students know and can do, measuring student achievement against standards designed to ensure that all students gain the knowledge and skills needed to succeed in college and the workplace (US Dept of Ed. 2010). Educators are faced with the challenge of grading and reporting student learning in terms of the standards, rather than showing how students compare to their classmates. To enrich overall assessment and to ensure mastery of standards, many schools have eliminated traditional report cards and are moving towards Standards Based Report Cards (SBRC.) Most educators agree that standards-based grading provides more specific and useful information to teachers, parents, and students about student mastery of standards. Focusing grading on standards, rather than comparing students to their classmates seems a natural follow up to standards based instruction, and should, ideally, lead teachers to better instructional methods and improve student achievement as shown by standardized test scores.

Guskey (2001) addresses the many challenges in implementing standards-based grade reporting but also emphasizes the benefits of such a system, noting the importance of aligning grade reporting to the standards and making sure that grades are a reflection of student learning, not behavior or effort. Marzano (2010) cautions that while standards-based grading is gaining popularity and "seems like good practice, without giving teachers guidance and support on how

to collect and interpret the assessment data with which scores like advanced, proficient, basic, and below basic are assigned, standards-based reporting can be highly inaccurate" (p.18). This study aims to determine whether standards-based report cards provide accurate information about student mastery of standards as measured by a criterion-referenced competency test, and whether there is alignment between grades on standards-based report cards and CRCT scores for all students.

This chapter will establish the theoretical framework that provides context and perspective for the study, examine the current literature and research on the topics of standards-based grading and high-stakes testing, and trace the reform movements that have guided the paradigm shift in how student learning is assessed.

Conceptual or Theoretical Framework

At the heart of thinking about standards-based assessment and its impact on student achievement is the social constructivist theory. Vygotsky established the foundations for the social constructivist theory in his theory of "the zone of proximal development." In his theory, Vygotsky established the idea of stretching learners by making sure that instruction is at their level, challenging but not frustrating. This idea of noting what students can do independently and with help, and then structuring instruction to address any gaps in learning is the very idea behind standards-based instruction and assessment. Rather than averaging grades that may or may not show mastery, standards-based grading seeks to communicate what students know, understand, and are able to do. According to Atherton (2009), it is common in constructing skills check-lists to have columns for "cannot yet do", "can do with help", and "can do alone"(p.1). Vygotsky's (1978) thoughts about how children learn impact the relationship between curriculum, instruction, and assessment, focusing on the learning process and the evidence of

what the student can do independently, with assistance, and how scaffolding can facilitate student learning (Hatch, 2010). Vygotsky's concept of the teacher-student interactions in which the teacher models and instructs, with the learner taking responsibility for learning and developing a skill over time and with practice reflects current ideas of standards-based education and assessment in that students take more of a role in showing learning through a variety of tasks, reflecting on their own learning and taking a role in the process. Today's assessment practices are more reflective of social constructivist perspectives and also reflect the ideas of "dynamic assessment" (Feurstein, 1979), which factor in both the predictive value of assessment as well as the student's potential for gain through instruction (Zane, 2009). The idea of using assessment to guide instruction, as well as to measure student learning, is the basis for standards-based grading.

The literature debates whether standards-based movements support Vygotsky's idea that learning leads development or Piaget's idea that development leads learning, noting that in most elementary schools, there is acknowledgement to both ideas. According to Hatch (2010), the key is to use curriculum, instruction, and assessment as a framework for making sure that elementary education is focused on teaching for learning. Hatch states, "Learning should be the stuff of early education, curriculum content should be the focus of what children learn, and teachers should use as many teaching strategies as necessary to maximize every child's opportunity to learn" (2010, p. 264).

Thomas Guskey(2007), an authority on standards-based grading also points to Benjamin Bloom's theory of mastery learning. Bloom sought ways to close the gap between high achieving and low achieving students, indicating that all students should be held to high standards, but not expected to meet them in the same way or at the same time. Guskey (2007)

notes, "A far better approach, according to Bloom, is for teachers to use their classroom assessments as learning tools, both to provide students with feedback on their learning progress and to guide the correction of learning errors. In other words, instead of using assessments only as evaluation devices that mark the end of a unit, Bloom recommended they be used as part of the instructional process to diagnose individual learning difficulties and to prescribe remediation procedures" (p.11). Since students bring with them different backgrounds, learning styles, and experiences, it is important to assess in a variety of ways to produce multiple pieces of evidence of student progress towards mastering standards. According to the constructivist theory, students should not be expected to show mastery of concepts at the same time or in the same way.

School Reform Leading to Standards-based Education

No Child Left Behind (NCLB) legislation has led to national reforms that include high stakes accountability. When the National Assessment of Educational Progress began reporting state-by-state scores in 1990, the South showed a particular need for reform and accountability, and high-stakes accountability reforms were enacted in many states across the nation (Opfer, Henry, and Mashburn, 2008). President George W. Bush led the Charlottesville Education Summit in 1989, which pulled together governors from the fifty states with the intended purpose of raising educational standards through national goals that allow for state-by-state structuring of accountability and annual reporting on progress. The summit produced six National Education Goals, laying the foundation for America 2000 and Goals 2000, and ultimately leading to defining national standards for academic disciplines. The underlying principles of the current standards-based reform model emerged with the following principles:

Involve educators, parents, and the community in setting common expectations for what students should know and be able to do at key grade levels; develop assessments aligned

> to the standards that students and teachers can use to guide instruction; encourage systems and schools to align instruction to the standards and organize professional development around instructional programs; create accountability systems to determine whether students are meeting the standards (Resnick, 2006).

While the framework was established at the national level through reform legislation, details are left to state and local leadership, including decisions about standards and assessment used to determine proficiency. As the federal government became increasingly involved in pushing for higher standards and accountability, "the concepts of "standards-based reforms and accountability" emerged as the operational definition of systemic reforms at the state level" (Chatterji, 2002, p. 347). While the terms systemic reform, standard-based reforms, national standards, and accountability have been used somewhat interchangeably, standards-based reform has come to be seen as the way states have responded to the push for higher standards and school accountability (Chatterji, 2002). According to Chatterji (2002), who performed a study on the systemic nature of reform initiatives, consistency and coherence in implementing reforms has been lacking, especially in the areas of teacher education and educational research (Chatterji, 2002).

The movement to high-stakes accountability and standards-based instruction impacts teaching and learning through curriculum, instruction, and assessment. Aligning assessment to standards, and building in accountability for school systems, teachers, and students are key features of educational improvement efforts from state to state. Researchers examine the question of whether standards-based reform is affecting instruction in the classroom by looking at curriculum, instructional practices, and assessment. The literature points out an obvious, but key point: "standards, assessments, and accountability can raise achievement only if they

motivate and enable better teaching—presumably the result of curriculum that is aligned with the standards and assessments, along with improved professional development for teachers and administrators" (Briars & Resnick. 2000, p. 1). Once policy-makers put the structures in place for standards-based education, states and local systems must set about the tasks of implementing the standards.

In a standards-based system, the standards are clearly tied to learning outcomes related to the content standards, providing students the opportunity to show in various ways, and at various stages, what they know and can do in relation to the learning goals. According to Tognolini and Stanley (2007), "performance and achievement standards are explicit statements of student performance that describe the levels of achievement within the learning area. These are derived from the outcomes of the curriculum and show development in relation to the construct being assessed" (p. 5). The authors note that while the process may seem straightforward, many aspects of the process require judgment and interpretation, providing considerable challenges to systems implementing standards-based assessment:

> For example, the learning outcomes are intended to describe what it means to grow or progress through an area of learning. This path is not deterministic and hence there is scope for this developmental sequence to be challenged by data; setting examination questions that accurately assess the learning outcomes that are consistent with the requirements of the achievement standards and are technically correct…is a challenging task; setting marking keys that are both fair and accurate is a challenging task; ensuring that marking keys are consistently applied is a challenge; accurately establishing the achievement standards and presenting them to teachers, examiners and students in a manner in which they will all interpret them consistently is a challenge; and operationally

> defining the boundaries of the achievement standards in the context of external and internal assessments is a challenge. It is the ability to control all of these aspects of sound assessment and decision making that lead to 'noise' or 'uncertainty' or measurement error in any assessment situation" (Tognolini & Stanley, 2007, p. 5).

While most states across the country have standards firmly in place, states and even local systems are in various stages of implementing the curriculum, aligning assessment, and supporting grading practices with professional development. Though the process is still ongoing in most systems, evaluation of progressive systems with standards-based instruction and assessment in place will benefit those systems still in transition. Under the legislation of NCLB, states determine the high-stakes testing used for measuring mastery of standards. However, school systems are on their own to develop a standard-based system of grading and reporting student grades. Evaluating the consistency and alignment of standards-based grades to high-stakes assessments indicates the predictive value of standards-based grades in determining student success on high-stakes tests serving as one component of schools' adequate yearly progress (AYP).

Researchers acknowledge the variation in grading practices from teacher to teacher and from school to school. One of the intents of standards-based grading is to decrease the subjective nature of grading through the use of common assessments, performance tasks, and rubrics. Research suggests that variation in grading practices is often the result of a lack of training (Stiggins, 2002). Furthermore, in traditional systems, teachers assign grades that combine behavior, work habits, and effort along with achievement (Brookhart, 1991; Cross & Frary, 1996). Studies that examine the discrepancy between student grades and performance on state accountability assessments have found significant variation between report card grades and

test scores. However, there is limited study of standards-based grading as compared to test scores, due in part to the relative recent implementation of standards-based grading in most systems.

Opposing Views on Assessment-Driven Reform

In the wake of *A Nation at Risk* published in 1983 (National Commission on Excellence in Education), and NCLB (2001), many powerful influences have come to see standardized testing as a necessary force for school improvement. This approach, termed *measurement-driven instruction* (MDI) by Popham and colleagues in 1985 continues as one of the driving forces behind accountability and reform (Pohham, Cruse, Rankin, Sandifer, & Williams, 1985). However, as noted by Wiggins (1990), authentic assessment that links classroom learning and standardized assessment is the key to successful reform. Proponents of high-stakes testing claim that attaching high-stakes to standardized assessment causes teachers to be "more reflective, deliberate, and critical in terms of classroom instruction and assessment" (Cizek, 2001, p. 24). Shanker (1995) advocates that high-stakes accountability is necessary to motivate teachers and students, claiming: "stakes for kids go right to the heart of what motivates them to work and learn" (p. 149). Researchers Carnoy and Leob (2002) found a strong positive relationship between the presence of high stakes attached to standardized assessment and NAEP improvement across 50 states (Carnoy & Leob , 2002). Proponents of the standards movement for reform feel that all students should be held to the same core standards and that these standards should go beyond basic skills, challenging students to a challenging content and skills needed for success, despite race, socioeconomic status, or disability (Thernstrom & Thernstrom, 2003).

Marzano (2001) strongly emphasizes the need for a reduced number of quality standards and quality tests that do not narrow the curriculum. Researchers emphasize that the careful choosing of standards, including precise wording to enrich instruction is key to improving achievement (Marzano, 2001;Gandal & Vranek, 2001). Likewise, tests must be carefully constructed to include more than surface level learning. According to the research of Gandal and Vranek (2001), "elementary and middle school tests, for the most part, are appropriately rigorous, yet the high school tests are often well below the level of challenge that we would expect of all students" (p.11).

One of essential components to bringing about desired changes for all students is making sure that the tools and training are in place to help all students reach desired goals. The research shows that identified "best practices" are key to improving student achievement. These research-based practices include identifying similarities and differences; summarizing and note taking; receiving reinforcement for effort and recognition for achievement; doing homework and practicing; using nonlinguistic representations; learning cooperatively; setting objectives and receiving feedback; generating and testing hypotheses; and using cues, questions, and advance organizers (Sherer, 2001).

The idea that all children can learn, while embraced by educators everywhere, is specifically supported by recent brain research that argues that rather than being genetically fixed, the brain remains malleable and responsive to mental force and enriched experiences well after birth and even into adult life (Ridley, 2003). Research also shows that most teachers support the standards movement and feel that for the most part, state-mandated tests cover content that all teachers should cover in their classrooms (Sunderman, Tracey, Kim & Orfield, 2004). Parents, politicians, and business leaders quite openly support the use of standardized

assessment and feel that a standardized curriculum leads to better instruction (Phelps, 2005). In a study by Lay and Stokes-Brown (2009), a majority of all respondents supported using standardized tests to assure students meet national standards and to identify needed areas of improvement for teachers and students, but most opposed the use of high-stakes tests to determine school funding or for students applying for jobs (Lay & Stokes-Brown, 2009).

Opponents of the assessment-driven reform movements disagree with the practice of relying on standardized test data to evaluate education in America and some even refute the claim that American schools are failing. Critics of high-stakes testing caution that while the idea of accountability may be desirable, the effects that it can have on classroom instruction and the natural curiosity of children can be negative. Kohn states, "endorsing the idea of accountability is quite different from holding students and teachers accountable specifically for raising test scores" (p.46). Kohn also points to the difficulty of holding a teacher accountable for student test scores, considering the many factors from a student's past and home environment that can play a part in performance (Kohn, 2000). Furthermore, opponents of high-stakes standardized tests argue that such tests simply cannot serve the purpose of fostering good teaching and learning, providing accountability, and providing reliable student data for making high-stakes decisions (Maduas, 1995). Though well intentioned, the unattainable goals set forth in NCLB to bring all children to proficient level, naturally lead to negative consequences (Wang, Beckett, & Brown, 2006). Included in the negative consequences of testing, are reduced instructional time and opportunities, lower teacher morale, and heightened anxiety for students (Valenzuela, 2000). Many of these opponents of standardized testing for accountability purposes also frown upon the idea of national or state standards, questioning whether these standards will stifle teacher and student creativity, (Wang et al., 2006). In response to the idea that standards-based education

stifles creativity, Marzano states, “Policy-makers are not telling teachers how to teach; they’re just saying that we must produce results relative to specific content. Using standards-based report cards would alleviate the pressure of the high-stakes tests because decisions could be made about students on the basis of patterns of scores obtained over time” (Sherer, 2001, p. 18).

Opponents of high-stakes testing also claim that such tests discriminate against racial and ethnic minorities (Haney, 1993, Hood, 1998, Jencks, 1998). However, research findings show that Latinos tend to be highly supportive of high-stakes testing, and optimistic about public education in general, while African Americans and Whites tend to be more skeptical. The authors contend that these opinions follow the groups cultural histories with education, in which African Americans tend to be frustrated with failed reform movements and Whites are concerned with distribution of resources (Lay, Stokes-Brown, 2009). Research by Hanushek and Raymond (2004) reported that coupling high-stakes accountability with standards based reforms resulted in higher overall and sub-group performance as measure by the National Assessment of Educational Progress, but gains for Latino and African American students were smaller than gains for White students (Hanushek & Raymond, 2004). While the discrepancies in progress is troubling, research suggests that overall, standards-based reforms and accountability systems have raised expectations for all students, and efforts have increased at improving educational opportunities for all students in core academic areas.

Impact of State Standards and Accountability Measures on ELLs

Under the No Child Left Behind Act of 2001 (NCLB), the progress of all students, including ELLs must be assessed annually in math and reading for grades three through eight (Young, Cho, Ling, Cline, Steinber, & Stone, 2008). While educational equity is stated as the goal of NCLB testing mandates, thus requiring that all students be administered the same

academic achievement assessments, research shows that it may take up to 7 years for ELLs to acquire the academic language needed to succeed on these tests (Hakuta, Goto, & Witt, 2000). The standards for Educational and Psychological testing state, "for all test takers, any test that employs language is, in part, a measure of their language skills. This is of particular concern for test takers whose first language is not the language of the test" (American Educational Research Association, American Psychological Association & National Council on Educational Measurement in Education, 1999, p. 91). Under NCLB legislation, states must develop English language proficiency (ELP) standards to measure student attainment of ELP and they must set measurable achievement objectives (AMAO) for ELL students (Wolf, Farnsworth, and Herman, 2008). Researchers point out varied definitions of academic English proficiency in states' ELP standards and assessments, and note the need for a framework of academic English to fully align standards and assessments for ELLs (Wolf, Farnsworth, and Herman, 2008).

Abedi and Gandara highlight the many factors that challenge the ELL population when it comes to standardized assessment and accountability, including parent education level, poverty, language acquisition, equitable schooling conditions, and ill-equipped measurement tools (Abedi & Gandara, 2008). Additionally, many teachers feel that they are not prepared to teach ELLs(Gandara, Rumberger, Maxwell-Jolly & Callahan, 2003).

A number of studies performed by Albedi and his colleagues show significant achievement gaps between ELLs and native English speakers, and that test scores of ELLs are generally lower across grade and subject levels (Abedi, 2002; Abedi & Lord, 2001; Abedi, Lord, & Hofstetter, 1998). A study by Duran (2006) found that while about 30% of non-ELLs scored at proficient level on the NAEP math and reading tests, only about 10% of ELLs earned a proficient score on these assessments. This study highlighted the magnitude of the gap in

achievement between ELLs and non-Ells as being greatest for tests that require more verbal processing skills, such as language arts, and least for mathematics assessments. A further study by Young, et al. (2008) aimed to test the fairness of standards-based assessments for ELLs, found that for all five math and sciences tests given to ELLs and non-ELLs, the mean scores for ELLs were significantly lower than for non-ELLs; however, the study indicates that there was very little evidence of differential test validity when ELL and non-ELL groups were compared (Young, et al, 2008).

Some educators argue that the federal accountability system under NCLB does not provide a true picture of the quality of education in schools with a large ELL population (Zehr, 2008). The concept of growth models to measure academic gains rather than expecting the same level of proficiency from non-native English speakers may provide an option for schools willing to participate, and would give educators some flexibility to focus on rich instruction rather than rigid test-preparation (Zehr, 2008).

Impact of State Standards and Accountability Measures on Ethnic Minorities

Though the intent of state standards and accountability measures aims at improving education for all students and ensuring that all students meet minimum competency standards, research suggests that current practices that rely on high-stakes testing overlook the needs of ethnic minorities and impoverished students. Furthermore, schools that are designated as "low-performing" based on low percentages of students meeting minimum competency requirements, tend to be largely composed of students from ethnic minorities and impoverished conditions (Townsend, 2002). While the intent of accountability measures such as NCLB was to close the gap between high and low achieving students, critics of NCLB and accountability through high-stakes testing claim that these measures penalize students from ethnic minorities and

impoverished backgrounds and that they do not accurately assess students with language forms that vary from standards English (Hilliard, 2000; Neil & Medina, 1989).

Critics also note the unfortunate impact on teaching learning that stems from high-stakes testing, claiming that the high-pressure to perform encourages more teacher-directed learning, while decreasing opportunities for enrichment, student-centered learning, and higher order thinking (Jones et al., 1999, Passman, 2000). According to Kohn (200), "remedial, skill-oriented instructional approaches are implemented wholesale with students from poverty backgrounds and members of ethnic minority groups while their European American counterparts are exposed to instructional practices involving experiential learning and higher order thinking skills" (p. 325). While standards-based education seeks to ensure that minimum competency standards are met, there is the risk of instructional practices focusing only on minimum competencies rather than immersing students in rich learning experiences. Brenda Townsend (2002) warns that African American children may be at an even greater risk of encountering a narrowed curriculum focusing on basic skills, which ultimately puts them at continued risk of falling behind rather than bridging the gap as intended. Furthermore, the author highlights the negative outcomes associated with using standardized testing to assess students and classify schools, noting both the problem of underachievement and motivation connected to racial identity, and the cycle of negative academic identity that can result when standardized testing is used as the primary means of classifying student and school achievement. According to Townsend, authentic assessments used in conjunction with standardized testing, a culturally responsive curriculum, and higher expectations for academic achievement would provide positive steps for African American students.

The research shows that while efforts have been made to narrow the achievement gap as measured by test scores between Whites and Asians as relative to African Americans and Hispanics, disparities continue to exist. States continue to join in the practice of using high-stakes testing with the intent of ensuring that students meet common standards, but the research shows a clear need to examine closely the effects of such a practice, especially noting possible unintended consequences of a narrowed curriculum and teacher-focused instruction (Madaus & Clarke, 2001).

State Standards, Accountability, and Socioeconomic Standing

With the implementation of standardized testing as a key factor in education reform, varying perspectives emerge regarding the possible unintended consequences of such testing on both the curriculum and student learning (Camilli, 2003). However, despite differing perspectives on testing, research shows that "children from socially disadvantaged backgrounds perform more poorly than their more advantaged peers (Waber, Gerber, Turcios, Wagner & Forbes, 2006, p. 459). The study performed by Waber et al. (2006) revealed that while factors associated with poverty may impair metacognitive skills, aspects of schooling can promote these skills. Furthermore, this study, focusing on fifth grade students from two public schools in the city of Boston found that targeted interventions for at-risk students combined with attending to mental health needs of children can lead to improved academic achievement (Waber, et al, 2006). While curriculums and assessment are becoming standardized, the preparation that children receive before entering kindergarten remains glaringly diverse. While some children enter kindergarten already reading, others begin without any awareness of letters or letter sounds. According to Zill (1996) these differences are closely related to factors of age, sex, and family background. According to a research study by Newman and Chin (2003), students with below

average skills tend to have a family that received food stamps or welfare, a single parent household, or non-English speaking parents.

One study by Yeh (2005) in Minnesota focused on four school districts from varying socioeconomic status to determine the impact of state accountability tests on the curriculum. Yeh found that while teachers did admit to narrowing the curriculum to match the state tests, they also noted benefits of standardized assessment. These benefits included a higher awareness of specific needs of students and more focused evaluation of student progress towards meeting learning goals.

Monsaas and Engelhard (1991) interviewed 186 teachers in Western Georgia, finding that the stress connected to high-stakes testing put increased pressure on teachers to focus on basic skills needed to pass the state tests. The study highlighted the differences between socioeconomic status, finding that teachers of students of lower socioeconomic status reported more test prep activities and direct instruction, whereas teachers of students of higher socioeconomic status felt less pressure to change instructional methods to achieve student success on state tests.

The Changing Role of Assessment

In traditional school structures, assessments have often been thought of as evaluations that follow instructional units. Scores were recorded in a grade-book and averaged together, but those scores were more to label than to inform. In a standards-based grading system, educators move away from this competitive grading system and instead focus on clearly communicating what they expect students to know, understand, and be able to do, and how they will assess that process, achievement, or performance. By basing grades on clearly stated learning goals, educators give meaning to grades and ultimately strengthen the focus of the standards in

teaching, learning, and assessment. Classroom assessments should be used, not just to evaluate students learning, but also to enhance it. While in the past, grades have been a reflection of how students measure up in relation to their classmates, for a clear picture of what students are learning, this approach is no longer adequate or beneficial to the learning process.

The move from norm-referenced to criterion-referenced grade reporting, focuses on student learning and on the standards, giving grades meaning that goes beyond the pass or fail connected with a letter grade. Guskey (2001) contrasts norm-referenced grading to criterion-referenced grading, and shows the necessity of aligning grade-reporting to the standards, rather than comparing students with their classmates. Traditionally, many teachers have distributed grades on a bell-shaped, normal probability curve; however, such grading practices do little to communicate a student's mastery of specific learning goals. Furthermore, grading that is based on a certain percentage of students earning each grade, places students in competition against each other, thus undermining cooperative learning, and creating an atmosphere of "winners and losers" rather than a cohesive group working together to meet goals, (Guskey, 2001).

Since all students enter a classroom with varying backgrounds and prior knowledge, they should not be expected to achieve at the same time. By averaging scores, students who learn later in the game are penalized, even though their total gain may be greater. By giving students this opportunity to meet learning goals, but without the pressure of averaged grades, teachers send an important message about what is important, student learning. Furthermore, grading practices that include attendance, effort and homework, or that average in zeros, may not reflect a student's mastery of standards; but rather, they may reflect student behavior more than student learning. Since a zero is often used as a punishment for lack of responsibility or effort, it is not a reflection of how students are progressing towards meeting learning standards (Canaday &

Hotchkiss, 1989). Guskey (2001) recommends, in a standards-based system, better information can be given and better motivation provided, by assigning an incomplete and requiring after-school work sessions for completion of missing work, without devastating effects to grades. Guskey states, " If grades are to represent information about the adequacy of students' achievement and performance with respect to clear learning standards, then the evidence used to determine grades must denote what students have learned and are able to do. "To allow other factors to influence grades or marks misrepresents students' learning attainment," (Guskey, 200l, p. 27). Patricia Scriffiny (2008) concurs, "The system must not allow students to mask their level of understanding with their attendance, their level of effort, or other peripheral issues" (p.72). By separating learning from behavior, educators are able to make sure that students are actually learning, not just playing the game well. The goal of standards-based assessment is to provide a framework for teachers, students, and parents to understand learning goals and to maintain the focus of teaching and learning on those goals.

When implemented well, the research shows that standards-based instruction and assessment can lead to more appropriate instruction that puts students, teachers, and parents on the same page in terms of what the students know and can do, and what learning goals need additional work. Such a system also has the benefit of allowing teachers and parents to track student growth over multiple years on standards that remain the same. This helps to provide a better picture of true growth and it includes the student in self-improvement by showing the student where he or she falls on a continuum rather than continually comparing the student to peers. Ultimately, the process reinforces the articulation of the standards and emphasizes that idea that education is more about learning and growth than ranking students (Toglioni & Stanley, 2007).

The Shift to Standards-based Grading

Once standards are in place, and educators decide to adopt a standards-based system of grading, they must face the task of developing the tools and procedures to record and communicate student progress and achievement. This often proves to be a challenging task, and can sometimes cause great debate and controversy. Guskey and Jung (2006) provide some recommendations for educators to consider when developing a standards-based report card. According to the authors, first educators must clarify their purpose by considering the questions: "What information do we want to communicate? Who is the primary audience for that information? How would we like that information to be used?" (Guskey & Jung, 2006, p. 2). Next, educators must consider grading criteria in terms of product, process, and progress. Guskey and Jung define product as relating to students' specific achievement or performance, process as relating to students' effort, behavior, or work habits, and progress as relating to how much students gain from learning experiences (2006). When educators combine all three of these elements into one grade, the result is a grade that is hard to interpret in terms of what a student has learned about the standards. In such grading practices, it is hard to determine whether a student earned a D because of missing assignments, poor attendance, or because he does not understand how to apply the standard.

To avoid confusing grades that do not reflect student learning, educators should attempt to establish clear indicators of product, process, and progress, and report each separately (Guskey 1994; Stiggins, 2001; Wiggins, 1996). By separating these categories, teachers are able to give parents more specific information about what students are learning, and what improvement is needed. For teachers, grading is often simplified by reducing the categories that one grade represents. However, for standards-based grading to be successful, the plan, including

criteria and indicators, must be carefully established and communicated to teachers, students, and parents. Choosing performance-level descriptors and maintaining consistency prove to be among the biggest challenges of standards based grade reporting. McTighe and O'Connor (2005) recommend showing criteria and providing rubrics for performance tasks in advance. When students understand evaluative criteria, and have a chance to view models of quality work, students are better prepared to produce quality work. Research supports the use of common rubrics within a school or district to ensure consistent criteria. By showing students what quality work looks like, and demanding quality work, teachers help students gain an understanding that quality is the expectation, and students must continue to work until they achieve that result, rather than just receiving a failing grade. McTighe and O'Connor (2005) recommend that teachers "allow new evidence of achievement to replace old evidence" (p. 213). Clearly, students learn at different rates, and students should not necessarily be rewarded for learning a skill faster; rather, the important thing is that they learn it well.

In an interview with Sherer, Robert J. Marzano (2001) claims that standards-based grading is really the only way to tell if students have met the standards. Marzano emphasizes the importance of timely feedback and states, "schools need to examine multiple data waves throughout the year, at least one data wave every grading period," (Sherer, 2001, p.15). Marzano suggests that teachers develop systems to plot student performance on the standards so that they can see growth over time. Additionally, Marzano urges educators to seek balance in their assessment, using both external assessments and internal classroom assessments to arrive at decisions about proficiency on standards (Sherer, 2001).

The idea of fairness in grading can sometimes be ambiguous, but is defined by Reeves (2011) as evaluation based on performance. Too often, many factors can impact the fairness of

grading, including support or lack of, behavior, available tools, and subjectivity. According to Lehrer (2009), the very idea that grades may not be fair has led to greater reliance on standardized tests. Therefore, a system that could report fairly with regard to student performance would enhance the educational process.

Reporting Student Learning

As the purpose of assessment changes, so must the system for reporting student progress. As educators seek to ensure that all students meet learning goals, they must assess for and report on learning in terms of the standards, rather than sorting students into categories. Stiggins traces the shift in assessment practices from assessment being purely evaluative in nature to the current trend of using assessment more to inform instruction. By making students aware of the learning goals, and making sure that students understand expectations through the use of rubrics and models, educators can draw students into the responsibility of charting their own progress towards meeting goals. In a review of the Kentucky Initiative in standards-based grade reporting, Guskey, Swan, and Jung (2010) report the need for a philosophical shift coupled with pre-service training and continued support throughout the process. Establishing the underlying understanding of the difference between formative and summative assessments is also crucial to a standards-based approach to grading. Ultimately, using SBRC should provide students, parents, and teachers with valuable information about a student's progress towards meeting year-long learning goals, and when implemented well, should provide a reliable tool for communicating mastery of standards.

Accountability and Development of Criterion-referenced Testing

In order to give meaning to their results, tests compare the examinee's score to some point of reference beyond the test, either the scores of other test-takers, as in norm-referenced

testing, or to an identified standard of proficiency, as in criterion-referenced testing. Both methods of measuring achievement have developed over the years, often complementing each other in the educational setting. Glaser and Klaus (1962) first introduced the idea of criterion-referenced measurement in military and industrial training, and later expanded the idea in a way that caught on in educational settings. Glaser's (1963) article in American Psychologist defines the idea of criterion-referenced measurement:

> Underlying the concept of achievement measurement is the notion of a continuum of knowledge acquisition ranging from no proficiency at all to perfect performance. An individual's achievement level falls at some point on the continuum as indicated by the behaviors he displays during testing. The degree to which his achievement resembles desired performance at any specified level is assessed by criterion-referenced measures of achievement or proficiency. The standard against which a student's performance is compared when measured in this manner is the behavior which defines each point along the achievement continuum (462).

Glasner paved the way for more recent thinking about standard-based grading in his ability to see the importance of comparing student achievement to a criterion standard rather than to the performance of other students. Further discussions about criterion-referenced testing uncovered the necessity of determining the objectives and level of acceptance, or a cut score (Mager, 1962). Popham and Husek (1969) developed another widely used definition of criterion-referenced measures: "Criterion-referenced measures are those which are used to ascertain an individual's status with respect to some criterion, i.e. performance standard. It is because the individual is compared with some established criterion, rather than other individuals, that these measures are described as criterion-referenced" (p. 2).

While the idea of comparing students to a set standard seems relatively straightforward, varying definitions and concepts remain, making the construction of a test to measure student proficiency a complex process. Tests that are diagnostic in nature seek to determine the level of behavior by noting if behavior is emerging, if there is a missing component, or an erroneous response (Nitko, 1980). According to Nitko (1980), "If the entire complex performance cannot be performed, the diagnosis consists of identifying which of the component behaviors are present and which are lacking" (p. 476). Nitko (1980) further describes the importance of well-defined domains to organize instructionally relevant behaviors.

In order for criterion-referenced tests to serve well as measures of accountability, assessments must be sensitive to instruction, showing that good teaching on the standards, leads to student learning, which leads to higher scores on the criterion-referenced tests. However, research shows variance in how standards are interpreted and presented to students (Hill, 2001). Researchers D'Agostino, Welsh, and Corson conducted a study in Arizona to determine the degree to which a test reflects instruction on the standards, noting that test scores should reflect instruction on the standards. According to the researchers, sensitivity analysis is vital because "Teachers' commitment to the reform effort will diminish if the assessments fail to register their efforts to provide students the opportunity to learn the standards" (D'Agostino, Welsh, & Corson, 2008, p. 6). The researchers surveyed teachers to determine how and to what degree standards were covered in the classroom, ultimately finding that of great importance to improved scores was that the standards were presented in the classroom in a similar way as they were presented on the test (D'Agostino, et al., 2008). While, clearly curriculum, instruction, and assessment should be well aligned in a standards-based system, Resnick warns of the dangers of educators teaching to the test, stating, "If we do not sharpen standards and measure what we

really mean by them, the nation is likely to wake up in a few years to find that it has created a "fool's gold" system" (Resnick, 2006, p.36). With high stakes come serious pressures to match teaching to the tests and spend valuable instructional time in mere test-prep activities at the expense of rich educational experiences.

Validity in Standards-based Assessment

Frederiksen and White (2004) discuss issues concerning validity in establishing both performance standards and the assessment tasks used to gauge student proficiency on the standards. Going back to the work of Haertel and Lourie (2004), the authors note the important components of validity arguments to be considered:

> (a) that interpretations of students' performance on assessment tasks used in judging whether they have met a performance standard are accurate (defensible criterion-referenced score interpretation), (b) that the proficiency levels stipulated in the performance standard represent knowledge and skills that are truly needed if students are to be prepared for their future learning (defensible normative judgment), and (c) that the attainment of a particular test score, a cut score, accurately represents a student's attainment of a performance standard (Frederiksen and White, 2004, p.1).

The authors note the possible validity threat connected to the presence of multiple approaches to performing tasks of any complexity, and they question whether the tests can actually test a true understanding (Frederiksen & White, 2004). This idea leads to the common fear that if teachers instruct with test performance in mind, they may not be developing students to their full capacity. The authors also raise the concern about the wide range of knowledge and skills that may be important and the difficulty of determining which are truly necessary for success in future educational endeavors. Finally, the authors address the use of cut-scores, noting the

importance of making sure that a cut score can accurately distinguish students who have reached a level of proficiency and those who have not (Frederiksen & White, 2004).

Summary

Researchers agree that traditional grade reporting, often resulting from an average of a variety of factors which may or may not reflect student learning, is seen as an unreliable measurement (Brookhart, 1993; Stiggins, Frisbie, & Griswold, 1989). The literature shows that when nonacademic factors are included in grading, a clear discrepancy exists between students' grades and their performance on state accountability assessments (Brennan, Kim, Wenz-Gross & Siperstein, 2001). This discrepancy has led teachers and administrators to seek alternate grade reporting methods that would use multiple measures of student learning to more accurately show student progress toward meeting academic standards.

Increased pressure for accountability, combined with increased capability to report and track student and school data, suggest that standards will continue to be essential to the educational process. As schools implement standards-based curriculums and align grading practices to the standards, there is a clear paradigm shift in thinking of grading. Educators moving to standards-based grading are no longer comparing students to each other, but instead, measuring student progress in relation to standards. While the literature clearly supports the use of standards-based assessment, noting the benefits of establishing clear indicators of what students should know, understand, and be able to do, and then reporting on each separately (Guskey 1994; Stiggins, 2001; Wiggins, 1996), there is a gap in the literature as to whether standards-based grade reporting reduces the discrepancy between students' grades and their performance on state accountability assessments. This study will address that gap in the literature by evaluating the relationship between standards-based report card data and scores on

the state-mandated criterion-referenced competency test in one school system in Georgia.

CHAPER THREE: METHODOLOGY

Introduction

This quantitative study examines whether grades on standards-based report cards provide accurate information for all students by comparing mean scores on Georgia's CRCT in the areas of math and reading, based on SBRC indicators. Standards-based grades were analyzed to determine the accuracy of information provided by SBRC regarding student learning as measured by comparing mean scores on the CRCT, based on SBRC scores, gender, socioeconomic status, and ELL status. The Criterion Referenced Competency Test (CRCT), a standardized test given to students in grades one through eight in the state of Georgia, is considered a measure of student learning and is used in this study to provide a comparison to mean scores on the Standards-based Report Card (SBRC), showing whether SBRC provide an effective measure of student's academic success. The purpose of this study is to examine whether grades on standards-based report cards provide accurate information for all students, regardless of gender, ELL status, or socioeconomic status by comparing on Georgia's CRCT in the areas of math and reading, based on SBRC scores and demographics.

Many school systems across the country are moving to a standards-based system of grading in an effort to improve assessment and ensure mastery of standards at each grade level. The study of standards-based grading and grade reporting is needed to determine whether such a grading system is an effective measure of student mastery of standards, as reported on Georgia's Criterion Referenced Competency Test. As systems move away from traditional norm-referenced grading systems, it is important to test the validity and reliability of standards based grading in indicating student success on standardized tests.

The purposes of Chapter 3 are to describe the: (a) sample population selected for this study; (b) instruments that were used for data collection; (c) methods, materials, and procedures used to collect the data for the study; (d) selection and use of statistical procedures employed in the analysis of the collected data.

Research Design

This non-experimental causal comparative study seeks to determine whether grades on standards-based report cards provide an accurate measure of student achievement on standardized criterion-referenced tests, and whether SBRC represents all groups of students fairly. Data was collected from over 500 upper-elementary students in fifth grade, from eight elementary schools in rural Northeast Georgia school system. A non-experimental, causal comparative design was used to determine whether grades on standards-based report cards provide accurate information for all students by comparing mean scores on Georgia's CRCT in the areas of math and reading, based on SBRC indicators. Because the researcher is interested in finding out if a significant difference in the means exists based on whether a student scores a does not meet, meets, or exceed on the standards-based report card, and whether a student earns a 1 (does not meet, 2 (meets), or 3 (exceeds) on the CRCT, and also whether demographic variables affect that relationship, a one-way analysis of variance (ANOVA) was used to determine whether the CRCT scores are significantly different among three levels of SBRC. The F statistic is used to do a hypothesis test to find out if the Math CRCT score has a common mean among three Math SBRC groups and again for reading. ANOVA returns the *p*-value from this hypothesis test. A p-value is a measure of how much evidence there is against the null hypothesis, where the null hypothesis in this study is that all Math CRCT scores have the same means among three Math SBRC groups and likewise for reading. The p-value is interpreted as

the probability that a population where all groups of students have the same average Math or Reading SBRC scores could produce a sample where the differences among the groups as large (or larger) than those seen in our study. Therefore, a small p-value (typically 0.05 of smaller) is evidence against the null hypothesis, while a large p-value means little or no evidence against the null hypothesis**.** Tests were conducted to see if there were interactions between SBRC scores and the demographic variables with respect to the CRCT scores. Additionally, data trends in the SBRC and CRCT scores were identified using the Tukey multiple comparisons test to determine which pairs of Math CRCT means and Reading CRCT means are significantly different. The CRCT is treated as the continuous dependent variable and grades on the SBRC serve as the independent variable. Other independent variables include student gender, socioeconomic status, grade level, and ELL status.

Research Questions and Research Hypotheses

The study will examine the following research questions to determine how standards-based grade reporting accurately informs student performance on the Criterion References Competency Test and to identify any differences in the means based on gender or sub-groups.

Research Question 1

Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?

Null Hypothesis 1

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for students in grade five will not be significantly different among students who scored Does not Meet, Meets, and Exceeds on standards-based report cards.

Alternative Hypothesis 1

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Research Question 2

Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?

Null Hypothesis 2

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Alternative Hypothesis 2

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Research Question 3

Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade ELL students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade ELL students different from the **differences in mean scores for fifth grade non-ELL students?**

Null Hypothesis 3

Average student performance as measured on the math section of the Georgia Criterion

Referenced Competency Test for ELL students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for ELL students and non-ELL students.

Alternate Hypothesis 3

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for ELL students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The pattern of these differences for ELL students will be different from the pattern of non-ELL students.

Research Question 4

Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade ELL students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade ELL students different from the differences in mean scores for fifth grade non-ELL students?

Null Hypothesis 4

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for ELL students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for ELL students and non-ELL students.

Alternative Hypothesis 4

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for ELL students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report

cards. The pattern of these differences for ELL students will be different from the pattern of non-ELL students.

Research Question 5

Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade students who receive free or reduced lunch with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade students who receive free or reduced lunch different from the differences in mean scores for fifth grade students who do not receive free or reduced lunch?

Null Hypothesis 5

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for students who receive free or reduced lunch in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for students who do and do not receive free and reduced lunch.

Alternative Hypothesis 5

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for students who receive free or reduced lunch in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The pattern of these differences for students who receive free or reduced lunch will be different from the pattern of students who do not receive free or reduced lunch.

Research Question 6

Are the mean scores on the reading section of the Georgia Criterion Referenced Competency

Test different for fifth grade students who receive free or reduced lunch with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade students who receive free or reduced lunch different from the differences in mean scores for fifth grade students who do not receive free or reduced lunch?

Null Hypothesis 6

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for students who receive free or reduced lunch in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for students who do and do not receive free and reduced lunch.

Alternative Hypothesis 6

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for students who receive free or reduced lunch in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The pattern of these differences for students who receive free or reduced lunch will be different from the pattern of students who do not receive free or reduced lunch.

Research Question 7

Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade female students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?

Null Hypothesis 7

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for female students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Alternative Hypothesis 7

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for female students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Research Question 8

Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade female students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?

Null Hypothesis 8

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for female students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Alternative Hypothesis 8

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for female students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Research Question 9

Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade male students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade male students different from the differences in mean scores for fifth grade female students?

Null Hypothesis 9

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for male students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for male and female fifth grade students.

Alternative Hypothesis 9

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for male students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The pattern of these differences for male and female students will be different.

Research Question 10

Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade male students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade male students different from the differences in mean scores for fifth grade female students? Null

Hypothesis 10

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for male students in grade five will not be significantly different

among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for male and female fifth grade students.

Alternative Hypothesis 10

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for male students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The pattern of these differences for male and female students will be different.

Participants

The population for the study included approximately 567 fifth grade students. All fifth grade students in the Northeast Georgia district who take the CRCT and receive standards-based report cards in math and reading were the subjects of this study. Fifth grade is seen as an important upper elementary grade for ensuring that students have met standards necessary for moving to middle school, and is one of the grades, identified as having higher accountability for students due to Georgia's state requirements that fifth grade students must pass the CRCT in order to progress to the next grade level. All fifth grade classes in the system from will be studied. Only SBRC data from second semester from the 2009-2010 school year will be used in order to represent end-of-year mastery of standards, rather than progress towards mastery of standards. All students participating in the study will be graded on the same state standards, using system-wide, common report cards, rubrics, performance indicators, and assessments. Students in the study will also take the CRCT under similar testing conditions.

Student test and report card data will be drawn from all of the school system's elementary students in grade five, using the county data portal, Infinite Campus. Data will be drawn collectively, without any identification of individual students, classes, or schools.

Setting

This study examines student data from a rural, but growing, Northeast Georgia School System, consisting of eight elementary schools, three middle schools, and two high schools. Four of the eight schools in the study are Title I schools, and system-wide, 50% of all students receive free or reduced lunch. The student population is predominately white, with approximately 12% Hispanic, 5% Black, and 3% other or multi-racial. According to the U.S. Census Bureau (2008), this Northeast Georgia system has a total population of 61,620 and a median annual household income of $52,029.

All fifth grade classes of students in the system from the 2009-2010 school- year will be studied, providing a sample population of approximately 1100 students. The school system uses a standards-based curriculum and has common curriculum maps, pacing guides, rubrics, and performance indicators in place. The system has established curriculum teams for each grade level to periodically review the standards, instructional methods and pacing, and assessments in order to insure that teachers across the system are teaching the standards and that assessment reflects student learning on those standards.

Instrumentation

Criterion referenced competency test.

This study will use scores on the Criterion Referenced Competency Test (CRCT), a standardized test required by the Department of Education of the State of Georgia. The CRCT is designed to measure student mastery of the Georgia Performance Standards (GPS). Implemented in 2000, students in grades one through eight have been required to take the CRCT as a summative assessment in the areas of language arts, reading, and math. Additionally, students in grades three through eight have been assessed in science and social studies since

2002. For this study, the researcher will address the math and reading sections of the CRCT to determine the strength of correlation between standards-based grades and CRCT scores for each area. The math and reading sections of the test were selected due to the guidelines established by the Georgia Department of education regarding promotion and retention, which state:

> Beginning with the current 2004-05 school year, the implementation of the Georgia Promotion, Placement, and Retention law (O.C.G.A. §§ 20-2-282 through 20-2-285) and State Board of Education Rule (160-4-2-.11) will take effect for students in Grade 5. All fifth grade students must achieve grade level scores on the Georgia Criterion Referenced Competency Tests (CRCT) in Reading and Mathematics in order to be promoted to the sixth grade (Georgia Department of Education, 2011).

The CRCT differs from norm-referenced tests in that it is designed to test how well students have learned the knowledge and skills within a specific curriculum and only tests the Georgia Performance Standards, rather than highlighting differences between students on an a general academic continuum. The test is designed to reveal individual strengths and weaknesses as related to the intended curriculum, and is used to inform on academic achievement at the student, class, school, system, and state level. The CRCT scores are reported using the following scale based on cut scores for performance levels based on scaled scores: 1(does not meet), 2 (meets), or 3(exceeds). The scaled scores will be compared to the progress indicators on the SBRC: DNM (does not meet), IP (in progress), M (meets), and E (exceeds).

Reliability is one of the key aspects of testing quality that ensures that the same measurement or comparable result will be given for the same student every time. The reliability of the CRCT is evaluated using statistical methods. For the 2004 CRCT, total test reliabilities ranged from 0.79 to 0.86 for Reading, 0.85 and 0.89 for Mathematics.

The other key component of technical quality in testing is validity, which begins with the purpose of the assessment and continues through item writing and review. All CRCT test items are written by qualified, professional content specialists specifically for the Georgia CRCT. After the items are written, curriculum specialists and committees of Georgia educators review the items. Items are evaluated for overall quality and clarity, content coverage and appropriateness, alignment to the curriculum, and grade appropriate stimuli with an emphasis on higher order thinking skills. In addition, there should be one clear correct answer with appropriate, relevant, and reasonable distractors. Items should be free from bias toward or against any particular group. Great care is taken throughout the item-development process to monitor items for potential bias and to ensure representation of all of Georgia's students. To ensure that the CRCT meet the highest standards of technical quality and defensibility, the Testing Division meets with an independent panel of experts – Georgia's Technical Advisory Committee (TAC) – on a quarterly basis. TAC members are experts in the field of educational measurement who review all aspects of the test development and implementation process on a continual basis. The Georgia Department of Education is confident that the CRCT are both reliable and valid and claims that Georgia stakeholders can have the highest confidence in the CRCT program. (Georgia Dept of Education, 2008).

Standards-based report cards.

The School System uses a common standards-based report card in each school within the district. The use of standards-based assessment and grade reporting was implemented system-wide in Jackson County during the 2008-2009 school year. Student achievement on state standards as reported by progress indicators on the Standards Based Report Cards (SBRC) will be examined and compared to CRCT scores. The progress indicators on the SBRC are as

follows: DNM (does not meet), IP (in progress), M (meets), and E (exceeds). Through teacher input on curriculum teams, common assessment practices for each grade level have been established; therefore, all students in the study will be assessed and graded using common language of the standards, as well as common rubrics, assessments, and performance indicators.

Procedures

After gaining approval from the Internal Review Board (IRB) at Liberty University (Appendix A) and from the school system (Appendix B), the researcher began collecting data to conduct the research. Prior to the implementation of this study, a thorough review of literature was completed. The review of literature focuses on characteristics of school reform, the implementation of high-stakes testing for accountability, the history of standards based curricula, the need for standards-based assessment, and current research on standards-based assessment.

The Board of Education for the school system has granted the researcher permission to use the CRCT and standards-based report card data from the 2009-2010 school year. With the assistance of board office personnel, the researcher used the student information system, Infinite Campus, to access CRCT scores and standards-based report card data on all students in the system in grades three and five. Since the researcher used no personal identifiers, getting parental permission on the subjects was not necessary. All school information, teacher information, and personal information was stripped from the data by board office personnel before the researcher obtained the data to comply with FERPA and protect the privacy of students and teachers.

The researcher identified all fifth grade students enrolled in the school system for the 2009-2010 school year having valid test scores for the CRCT and a standards-based report card for the second semester. Using Infinite Campus, the school system's student information system, the students' standards-based summary marks and CRCT scores in math and reading were

collected and entered into an excel spreadsheet. No personal identification of students or teachers, through names, numbers, or classroom assignments were used.

Scores on the Criterion Referenced Competency Test (CRCT), a standardized test required by the Department of Education of the State of Georgia and progress indicators on the Standards Based Report Cards (SBRC), used by all elementary schools in Jackson County, were examined to determine the accuracy in providing information regarding student learning as measured by between the grades that students receive on standards-based report cards and the scores students receive on the CRCT, and then subgroups were examined to determine differences in the mean CRCT scores that would indicate whether standards-based grading provides a system that is fair to all students.

Data Analysis

In this non-experimental study, data were analyzed using SAS 9.2, and descriptive statistics were calculated for SBRC scores and CRCT scores. The researcher chose to use the scale scores for CRCT, as these provide more information than categorical scores. Therefore, there is a continuous score for CRCT, but SBRC is graded only in categories (DNM, M, E). Due to the categorical nature of the predictors, summary marks, gender, ELL status, free/reduced lunch, an Analysis of Variance, or ANOVA was used. This analysis examines the differences among the groups, such as between DNM, M, and E, and provides a decision, based on the variability in the scores within each group, whether the differences are just caused by chance in this particular sample, or whether there are real differences among these populations. By examining the differences between SBRC performance indicators “does not meet, meets, or exceeds” and the standards to CRCT scores of “does not meet, meets, and exceeds,” the researcher examines whether there is a statistically significant difference in mean performance

on the CRCT based on SBRC performance indicators, which would be expected if SBRC indicators are aligned with CRCT performance, and whether these differences vary between groups of students. Furthermore, data were analyzed to determine if there are significant variations in the difference of average CRCT scores among those receiving different SBRC indicators based on grade, gender, socioeconomic status or status as an English Language Learner.

The F statistic is used to do a hypothesis test to find out if the Math CRCT score has a common mean among three Math SBRC groups. ANOVA returns the *p*-value from this hypothesis test. A p-value is a measure of how much evidence we have against the null hypothesis, where the null hypothesis in this study is that all Math CRCT scores have the same means among three Math SBRC groups. The p-value is interpreted as the probability that a population where all groups of students have the same average Math SBRC scores could produce a sample where the differences among the groups as large (or larger) than those seen in our study. Therefore, a small p-value (typically 0.05 of smaller) is evidence against the null hypothesis, while a large p-value means little or no evidence against the null hypothesis**.**

After testing to determine whether there were interactions of subgroups with the correspondence of SBRC scores to CRCT scores, the researcher used Tukey multiple comparisons test to determine which pairs of means are significantly different for both math and reading groups.

To learn whether the demographic variables indicate differences in the Math and Reading CRCT scores, the researcher used t-tests to compare the means of Math CRCT score by Gender, ELL and Lunch. In a T-test, a t statistic is computed and compared to a critical value. The critical value is chosen so that if the means are really the same in the population, the probability

(p-value) that the t statistic will exceed the critical value is small; 5% (or 0.05) is used in this study, as is typical. When the means are different, the probability that the statistic will exceed the critical value is larger.

CHAPTER FOUR: RESULTS

As the focus of school reform targets standards and assessment practices to provide educators with the information needed to continuously improve teaching and learning, teachers face the challenge of reporting student learning in terms of the standards, rather than showing how students compare to their classmates. To enrich overall assessment and to ensure mastery of standards, many schools have eliminated traditional report cards and are moving towards Standards Based Report Cards (SBRC.) Most educators agree that traditional grading methods have been inadequate in communicating student learning and that standards-based grading provides more specific and useful information to teachers, parents, and students about student mastery of standards (Marzano, 2000, Guskey, 2001). Focusing grading on standards, rather than comparing students to their classmates seems a natural follow up to standards based instruction, and should, ideally, lead teachers to better instructional methods and improve student achievement as shown by standardized test scores. School systems considering standards-based grading seek a less subjective and more informative method of grade reporting that truly reflects student learning and mastery of standards as measured by Georgia's CRCT, a standardized test given to student's grades three through eight in the state of Georgia.

The purpose of this study is to determine whether grades on standards-based report cards provide accurate information for all students by comparing mean scores on Georgia's CRCT in the areas of math and reading, based on SBRC indicators. The participants for this study will be fifth graders enrolled in eight elementary schools in a rural county in north Georgia from 2009-2010. This dataset contains two sections, one for math grades on SBRC and CRCT and one for reading grades. The analysis will be conducted by identifying the statistically

significant differences in average CRCT scores based on SBRC math grades and reading grades respectively.

Research Questions and Research Hypotheses

The study will examine the following research questions to determine how standards-based grade reporting accurately informs student performance on the Criterion References Competency Test and to identify any differences in the means based on gender or sub-groups.

Research Question 1

Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?

Null Hypothesis 1

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for students in grade five will not be significantly different among students who scored Does not Meet, Meets, and Exceeds on standards-based report cards.

Alternative Hypothesis 1

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Research Question 2

Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?

Null Hypothesis 2

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Alternative Hypothesis 2

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Research Question 3

Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade ELL students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade ELL students different from the differences in mean scores for fifth grade non-ELL students?

Null Hypothesis 3

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for ELL students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for ELL students and non-ELL students.

Alternative Hypothesis 3

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for ELL students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The pattern of these differences for ELL students will be different from the pattern of non-ELL students.

Research Question 4

Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade ELL students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade ELL students different from the differences in mean scores for fifth grade non-ELL students?

Null Hypothesis 4

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for ELL students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for ELL students and non-ELL students.

Alternative Hypothesis 4

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for ELL students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The pattern of these differences for ELL students will be different from the pattern of non-ELL students.

Research Question 5

Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade students who receive free or reduced lunch with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade students who receive free or reduced lunch different from the differences in mean scores for fifth grade students who do not receive free or reduced lunch?

Null Hypothesis 5

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for students who receive free or reduced lunch in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for students who do and do not receive free and reduced lunch.

Alternative Hypothesis 5

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for students who receive free or reduced lunch in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The pattern of these differences for students who receive free or reduced lunch will be different from the pattern of students who do not receive free or reduced lunch.

Research Question 6

Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade students who receive free or reduced lunch with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade students who receive free or reduced lunch different from the differences in mean scores for fifth grade students who do not receive free or reduced lunch?

Null Hypothesis 6

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for students who receive free or reduced lunch in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for students who do and do

not receive free and reduced lunch.

Alternative Hypothesis 6

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for students who receive free or reduced lunch in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The pattern of these differences for students who receive free or reduced lunch will be different from the pattern of students who do not receive free or reduced lunch.

Research Question 7

Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade female students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?

Null Hypothesis 7

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for female students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Alternative Hypothesis 7

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for female students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Research Question 8

Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade female students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?

Null Hypothesis 8

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for female students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Alternative Hypothesis 8

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for female students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Research Question 9

Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade male students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade male students different from the differences in mean scores for fifth grade female students?

Null Hypothesis 9

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for male students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for male and female fifth grade students.

Alternative Hypothesis 9

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for male students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The pattern of these differences for male and female students will be different.

Research Question 10

Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade male students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade male students different from the differences in mean scores for fifth grade female students?

Null Hypothesis 10

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for male students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for male and female fifth grade students.

Alternative Hypothesis 10

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for male students in grade five will be significantly different among students who scored **Does Not Meet, Meets, and Exceeds** on standards-based report cards. The pattern of these differences for male and female students will be different.

Data Summary

There are 549 students in the math dataset. First, all five variables were examined to determine their basic properties. Table 1 gives the frequency of Math SBRC among the students

in grade five. DNM means “Does Not Meet” the standards, M means “Meets” the standards and E means “Exceeds” the standards. Approximately 79% of students meet the standards, while 12% of students exceed the standards, and 9% do not meet the standards on the Math SBRC.

Table 1. Frequency Distribution of Math SBRC

MathSBRC	Frequency	Percent
DNM	53	9.65%
M	432	78.69%
E	64	11.66%

Table 2 gives the summary statistics for Math CRCT and the distribution of the CRCT categorized scores. The CRCT scores are reported using the following scale based on cut scores for performance levels based on scaled scores: DNM means “Does Not Meet” the standards, M means “Meets” the standards and E means “Exceeds” the standards. The average CRCT score is 852.37 with standard deviation 40.36. The minimum and maximum score is 729 and 990 respectively; 52.64 % of students exceed the standards.

Table 2.

Summary Statistics for Math CRCT

Variable	N	Minimum	Median	Maximum	Mean	Std Dev
MathCRCT	549	729	850	990	852.37	40.36

Frequency Distribution of Math CRCT

MathCRCT	Frequency	Percent
DNM	36	6.56%
M	224	40.80%
E	289	52.64%

Table 3 gives the distribution of Gender among students (F=Female, M=Male); 47.72% of students are female, 52.28% are male.

Table 3.

Frequency Distribution of Gender

Gender	Frequency	Percent
F	262	47.72%
M	287	52.28%

Table 4 gives the distribution of ELL status. ELL status is a rating of the students' English speaking abilities (N = proficiency, Y = limited proficiency). Only 4.19% of students are determined to be of limited English proficiency.

Table 4.

Frequency Distribution of ELL

ELL	Frequency	Percent
N	526	95.81%
Y	23	4.19%

Table 5 gives the distribution of Lunch (C=At-Cost, F=Free, R=Reduced). Among these students, only 11.11% received reduced lunch, and 45.17% received free lunch, for a combined free/reduced lunch population of 56.28% (these two groups will be combined for later analyses).

Table 5.

Frequency Distribution of Lunch

Lunch	Frequency	Percent
C	240	43.72%
F	248	45.17%
R	61	11.11%

Reading dataset

There are 513 students in the Reading portion of the dataset with 2 students who don't have the reading CRCT score, so only 511 students will be taken into account in this analysis.

The variables are the same as those in the math dataset, including Reading SBRC, Reading CRCT, gender, ELL and socioeconomic status as indicated by lunch. The following tables examine the distributions of all variables and summary statistics of Reading CRCT. For the Reading SBRC, 68.10% of students meet the standards, while 24.46% exceed the standards. For the Reading CRCT score, the average score is 835.71 with a standard deviation 23.27. Categorically, 67.12% of students meet the Reading CRCT standards, and 29.55% exceed the standards. Among these students, 46.77% are female, 53.23% are male. Only 3.13% of students are determined to be of limited English proficiency; most of the students (96.87%) are proficient in the English language. Of these students, 42.86% received free lunch and only 10.76% receive reduced lunch, for a total of 53.62% receiving free or reduced lunch (these categories will be combined for the statistical analyses).

Table 6.

Frequency Distribution of Reading SBRC

ReadingSBRC	Frequency	Percent
DNM	38	7.44%
M	348	68.10%
E	125	24.46%

Table 7.

Summary Statistics for Reading CRCT

Variable	N	Minimum	Median	Maximum	Mean	Std Dev
ReadingCRCT	511	771	834	920	835.71	23.27

Frequency Distribution of Reading CRCT

ReadingCRCT	Frequency	Percent
DNM	17	3.33%
M	343	67.12%
E	151	29.55%

Table 8.

Frequency Distribution of Gender

Gender	Frequency	Percent
F	239	46.77%
M	272	53.23%

Table 9.

Frequency Distribution of ELL

ELL	Frequency	Percent
N	495	96.87%
Y	16	3.13%

Table 10.

Frequency Distribution of Lunch

Lunch	Frequency	Percent
C	237	46.38%
F	219	42.86%
R	55	10.76%

Statistical Analyses and Conclusions

Math dataset.

In this section, a one-way analysis of variance (ANOVA) is used to determine whether the Math CRCT scores are significantly different among three levels of Math SBRC. The F statistic is used to do a hypothesis test to find out if the Math CRCT score has a common mean among three Math SBRC groups. ANOVA returns the *p*-value from this hypothesis test. A p-value is a measure of how much evidence we have against the null hypothesis, where the null hypothesis in this study is that all Math CRCT scores have the same means among three Math SBRC groups. The p-value is interpreted as the probability that a population where all groups of

students have the same average Math SBRC scores could produce a sample where the differences among the groups as large, or larger, than those seen in our study. Therefore, a small p-value (typically 0.05 of smaller) is evidence against the null hypothesis, while a large p-value means little or no evidence against the null hypothesis**.** From Table 11(a), the overall p-value is smaller than 0.0001. This is a strong indication that the average Math CRCT scores from the different Math SBRC groups are not the same. Differences in Math CRCT scores among the groups as large as those seen in this data set, and therefore an *F* statistic as extreme as the observed F, would occur by chance less than once in 10,000 times if the scores were truly equal in the population. Furthermore, in Table 11(b) a Tukey multiple comparisons test was used to determine which pairs of Math CRCT means are significantly different. The first line compares DNM Math SBRC scores to M Math SBRC scores; the average CRCT score for individuals with a DNM on the SBRC is 796.58, and the average CRCT score for individuals with an M on the SBRC is 852.81. This is a difference of 56.23 points, which is statistically significant at that 0.0001 level. In Table 11, the R^2 of 0.3188 indicates that Math SBRC score is associated with 31.88% of the variation in Match CRCT scores. The RMSE of 33.3710 indicates that if one were to predict the Math CRCT score from Math SBRC, the "typical" error of that prediction would be 33.3710 points. The results show that all three pairs of means are significantly different from one another, since the P-values are all less than 0.0001.

Table 11 (a).

ANOVA Results for Prediction of Math CRCT from Math SBRC Scores

Predictor	**R^2**	**RMSE**	**F-Statistic**	**P-Value**
MathSBRC	0.3188	33.3710	127.74	<.0001

Table 11(b).

Least Squares Means Comparisons for Math CRCT Score by Math SBRC Score

MathSBRC1	Math SBRC2	Average CRCT1	Average CRCT2	Difference	P-value
DNM	**M**	796.58	852.81	56.23	<0.0001
DNM	**E**	796.58	895.58	99.00	<0.0001
M	**E**	852.81	895.58	42.77	<0.0001

Next, to determine whether the demographic variables influence the Math CRCT scores, t-tests were used compare the means of Math CRCT score by Gender, ELL and Lunch. In a T-test, a t statistic is computed and compared it to a critical value. The critical value is chosen so that if the means are really the same in the population, the probability (p-value) that the t statistic will exceed the critical value is small; 5% (or 0.05) is used in this case, as is typical. When the means are different, the probability that the statistic will exceed the critical value is larger. The p-value for Gender is 0.2158, indicating there is no significant difference in Math CRCT scores based on Gender. The p-value for ELL is 0.0055, which is smaller than 0.05. Therefore, the null hypothesis that the Math CRCT mean in each group is equal is rejected at the 0.05 level of significance. We conclude that the mean Math CRCT score from the English proficient group is significantly different from the mean score from limited proficiency group. Specifically, the mean is higher in the group that is proficient in English. For the variable Lunch, the mean Math CRCT score of students who received at-cost lunch is significantly different from the mean score from students who received free or reduced lunch (it is higher) with a p-value of < 0.0001 which is considered very significant.

Table 12.

T-test results of Math CRCT by Gender, ELL and Lunch

Gender	N	Mean	Std Dev	T Statistic	P-Value
F	262	850.02	37.607	1.24	0.2158
M	287	845.39	42.678		

ELL	N	Mean	Std Dev	T Statistic	P-Value
N	526	853.07	40.772	3.01	0.0055
Y	23	836.48	24.998		

Lunch	N	Mean	Std Dev	T Statistic	P-Value
C	240	865.72	39.715	7.13	<.0001
FR	309	842.00	37.779		

A final model was used to predict continuous Math CRCT scores from all four of those variables simultaneously to determine overall what influences Math CRCT scores (Math SBRC, Gender, ELL and Lunch). At the beginning, the researcher included the two-way interaction effect of Math SBRC*Gender, Math SBRC*ELL, and Math SBRC*Lunch in this model. As an example of the interpretation of an interaction, suppose there were a significant interaction between Math SBRC and ELL; this would indicate that not only does ELL influence the CRCT scores, but that CRCT also influences the relationship between CRCT and SBRC. However, no interactions proved to be significant, so all the interaction terms were removed and the researcher chose the variables that have a significant effect on Math CRCT based on the P-values. Table 13 shows that only variables Math SBRC and Lunch significantly affect the Math CRCT scores. Note that ELL was not present in the model along with Lunch status, meaning that once Lunch

status is accounted for ELL status does not influence the results. Further exploration of the data indicated this may be because all but one student with limited English proficiency also received free or reduced lunch, and so the effects of these two variables may be somewhat confounded.

Table 13.

ANOVA Results for Prediction of Math CRCT

Variable	R^2	RMSE	F-Statistic	P-Value
Math CRCT	0.3540	32.5265	99.54	<.0001

Predictor	F-Statistic	P-Value
Math SBRC	113.40	<.0001
Lunch	29.71	<.0001

The presence of both Math SBRC and Lunch status in the model indicate that even when considering Math SBRC scores, whether or not a student is on free/reduced lunch influences his or her Math CRCT score. Additional results indicate that even if students have the same Math SBRC score, a student with At-Cost lunch will, on average, score 15.55 points higher on the Math CRCT than those with Free/Reduced lunch.

Next, two-way tables of categorized Math CRCT score * Math SBRC show a categorical version of the SBRC scores explained in section 2.1, along with a test of symmetry. The test of symmetry determines whether deviations from the diagonal are random, or if they tend to be in one direction. That is, the test of symmetry would indicate whether Math SBRC scores are biased in their prediction of Math CRCT scores. These tests can show whether categorized Math SBRC scores are typically matched to their respective equivalents in the Math CRCT scores. The analysis also provides these two-way tables according to Gender and Lunch, respectively. The researcher did not provide these for ELL status, since so few students fell into fell into the limited proficiency category relative to those who were considered proficient.

First, the researcher provides Tables 14, Table 14(a) to show that there are 33 students who don't meet the Math SBRC standards or the Math CRCT standards, which is 62.26% of students who don't meet the Math SBRC standards. Eighteen students who don't meet the Math SBRC standards (33.96%) meet the Math CRCT standards, and the remaining two students who don't meet the Math SBRC standards (3.77%) exceed the Math CRCT standards. There are three students who meet the Math SBRC standards who do not meet the Math CRCT standards, which are 0.69% of the students who meet the Math SBRC standards, 206 students who meet the Math SBRC standards who meet the Math CRCT standards (47.69%), and 223 students who meet the Math SBRC standards who exceed the Math CRCT standards (51.62%). Finally, 0 students who exceed the Math SBRC standards did not meet or simply met the Math CRCT standards; all 64 (100%) of these students exceeded the Math CRCT standards. The most important thing to note when interpreting this table is that the percentages displayed in this table are the percentages within each Math SBRC group who fall into each Math CRCT group rather than overall percentages.

The P value of the test of symmetry for Table 14(a) is less than <0.0001, which indicates that we need to reject the null hypothesis that the cell proportions are symmetric. The clear reason for this is that Math SBRC Scores appear to be consistently lower than their Math CRCT counterparts. This is clear from looking at the "Meets" category for Math CRCT scores in Table 14(a); 52% of students who are given an M on their Math SBRC in fact exceeded the Math CRCT standards. This is true to a lesser extent with the "Does Not Meet" category of the Math SBRC, where 34% of students who are given a DNM on their Math SBRC meet the Math CRCT standards and 4% even exceed the Math CRCT standards.

Similar results follow for each of the subsets of students, all showing this same significant trend of the Math SBRC score being lower than the Math CRCT categorized score. Tables 14(a)-(e).

Frequency Distributions of Math CRCT vs. Math SBRC and Results of Test of Symmetry

Table 14(a) Overall

Table of Math CRCT by Math SBRC

		Math SBRC		
Math CRCT	**DNM**	**M**	**E**	**Total**
DNM	33	3	0	36
	62.26%	0.69%	0.00%	
M	18	206	0	224
	33.96%	47.69%	0.00%	
E	2	223	64	289
	3.77%	51.62%	100.00%	
Total	53	432	64	549

Test of Symmetry	
Statistic(s)	235.7143
P-Value	<.0001

Table 14(b)

Gender: Female

Table of Math CRCT by Math SBRC

	Math SBRC			
Math CRCT	**DNM**	**M**	**E**	**Total**
DNM	10	1	0	11
	50.00%	0.48%	0.00%	
M	9	94	0	103
	45.00%	44.98%	0.00%	
E	1	114	33	148
	5.00%	54.55%	100.00%	
Total	20	209	33	262

Test of Symmetry	
Statistic(s)	121.4000
P-Value	<.0001

Table 14(c)

Gender: Male

Table of Math CRCT by Math SBRC

	Math SBRC			
Math CRCT	**DNM**	**M**	**E**	**Total**
DNM	23	2	0	25
	69.70%	0.90%	0.00%	
M	9	112	0	121
	27.27%	50.22%	0.00%	
E	1	109	31	141
	3.03%	48.88%	100.00%	
Total	33	223	31	287

Test of Symmetry	
Statistic(s)	114.4545
P-Value	<.0001

Table 14(d)

Lunch: At Cost

Table of MathCRCT by MathSBRC

	MathSBRC			
MathCRCT	**DNM**	**M**	**E**	**Total**
DNM	6	1	0	7
	66.67%	0.52%	0.00%	
M	2	70	0	72
	22.22%	36.08%	0.00%	
E	1	123	37	161
	11.11%	63.40%	100.00%	
Total	9	194	37	240

Test of Symmetry	
Statistic(s)	124.3333
P-Value	<.0001

Table 14(e)

Lunch: Free/Reduced

Table of MathCRCT by MathSBRC

MathCRCT	MathSBRC			
	DNM	**M**	**E**	**Total**
DNM	27	2	0	29
	61.36	0.84	0	
M	16	136	0	152
	36.36	57.14	0	
E	1	100	27	128
	2.27	42.02	100.00	
Total	44	238	27	309

Test of Symmetry	
Statistic(s)	111.8889
P-Value	<.0001

Reading dataset

The same analysis was conducted with the reading dataset. Table 15 shows that the P-value of one-way ANOVA model is smaller than 0.0001, which indicates that the average Reading CRCT scores from the different Reading SBRC groups are not the same. Furthermore, the Tukey multiple comparisons test in Table 15(b) shows that all three pairs of Reading CRCT means are significantly different.

Table 15 (a).

ANOVA Results for Prediction of Reading CRCT from Reading SBRC scores

Predictor	**R^2**	**RMSE**	**F-Statistic**	**P-Value**
ReadingSBRC	0.4148	17.8402	180.01	<.0001

Table 15(b).

Least Squares Means Comparisons for Reading CRCT Score by Reading SBRC Score

ReadingSBRC1	Reading SBRC2	Average CRCT1	Average CRCT2	Difference	P-value
DNM	**M**	806.79	830.23	23.44	<0.0001
DNM	**E**	806.79	859.77	52.98	<0.0001
M	**E**	830.23	859.77	29.54	<0.0001

Table 16 shows that for Gender, ELL, and Lunch, Reading CRCT scores are significantly different from one another. Females score significantly higher than males; those who have limited English proficiency score lower than those who do not; and those on the free/reduced lunch program score lower than those who purchase lunch at cost.

Table 16.

T-test results of Reading CRCT by Gender, ELL and Lunch

Gender	N	Mean	Std Dev	T Value	P-Value
F	239	838.57	21.954	2.62	0.0092
M	272	833.2	24.137		

ELL	N	Mean	Std Dev	T Value	P-Value
N	495	836.33	23.117	3.36	0.0009
Y	16	816.69	20.49		

Lunch	N	Mean	Std Dev	T Value	P-Value
C	237	843.03	23.01	6.91	<.0001
FR	274	829.38	21.623		

A final model predicts continuous Reading CRCT scores from all four of those variables simultaneously to determine overall what influences Reading CRCT scores (Reading SBRC, Gender, ELL and Lunch). Table 17 shows that all variables Reading SBRC, Gender, ELL and Lunch significantly affect the Reading CRCT scores.

Table 17.

ANOVA Results for Prediction of Reading CRCT

Variable	R^2	RMSE	F-Statistic	P-Value
ReadingCRCT	0.4487	17.3666	82.20	<.0001

Predictor	F-Statistic	P-Value
ReadingSBRC	155.41	<.0001
Gender	11.42	0.0008
ELL	4.27	0.0394
Lunch	13.02	0.0003

This result implies that even when accounting for Lunch status and ELL status, there is still an effect of Gender; even when accounting for Gender and ELL status there is still an effect of Lunch status, and even when accounting for Gender and Lunch status there is still an effect of ELL status. Additionally, all of these have an effect in the presence of Reading SBRC, which implies that for males and females, for those with limited proficiency and those who are proficient at English, and for those on free/reduced lunch status or not, the same Reading SBRC score may predict a different Reading CRCT score. Results indicate that on average, even with the same Reading SBRC score, females score 5.22 points higher on average than males on the Reading CRCT; even with the same Reading SBRC score, those with English proficiency score 9.23 points higher on average than those with limited English proficiency; and even with the same Reading SBRC score, those who pay for lunch at cost score 5.80 points higher on average than those on free/reduced lunch.

Tables 18(a)-(e) show the two-way tables of categorized Reading CRCT score * Reading SBRC score, along with a test of symmetry. The p-values of the tests of symmetry are less than 0.05 except when gender= male, which indicates that we need to reject the null hypothesis that the cell proportions are symmetric in most cases. Typically there is a similar trend as in Math that Reading SBRC scores tend to underestimate Reading CRCT score categories, though it

doesn't appear to be as severe as with Math. For males, however, Reading SBRC scores seem to present a balanced picture of Reading CRCT scores.

Tables 18(a)-(e).

Frequency Distributions of Reading CRCT vs. Reading SBRC and Results of Test of Symmetry

Table 18(a)

Overall

Table of ReadingCRCT by ReadingSBRC

		ReadingSBRC			
ReadingCRCT					
		DNM	**M**	**E**	**Total**
	DNM	12	5	0	17
		31.58%	1.44%	0.00%	
	M	25	290	28	343
		65.79%	83.33%	22.40%	
	E	1	53	97	151
		2.63%	15.23%	77.60%	
Total		38	348	125	511

Test of Symmetry

Statistic(s)	22.0494
P-Value	<.0001

Table 18(b) *Gender: Female*

Table of ReadingCRCT by ReadingSBRC

	ReadingSBRC			
ReadingCRCT	**DNM**	**M**	**E**	**Total**
DNM	2	1	0	3
	14.29 %	0.59%	0.00%	
M	11	136	8	155
	78.57 %	80.47 %	14.29 %	
E	1	32	48	81
	7.14%	18.93 %	85.71 %	
Total	14	169	56	239

Test of Symmetry	
Statistic(s)	23.7333
P-Value	<.0001

Table 18(c) *Gender: Male*

Table of ReadingCRCT by ReadingSBRC

	ReadingSBRC			
ReadingCRCT	**DNM**	**M**	**E**	**Total**
DNM	10	4	0	14
	41.67 %	2.23%	0.00%	
M	14	154	20	188
	58.33 %	86.03 %	28.99 %	
E	0	21	49	70
	0.00%	11.73 %	71.01 %	
Total	24	179	69	272

Test of Symmetry	
Statistic(s)	5.5799
P-Value	0.1339

Table 18(d) *Lunch: C*

Table of ReadingCRCT by ReadingSBRC

		ReadingSBRC			
ReadingCRCT					
		DNM	**M**	**E**	**Total**
	DNM	0	3	0	3
		0.00%	2.04%	0.00%	
	M	7	109	17	133
		100.00%	74.15%	20.48%	
	E	0	35	66	101
		0.00%	23.81%	79.52%	
Total		7	147	83	237

Test of Symmetry	
Statistic(s)	7.8308
P-Value	0.0496

Table 18(e) *Lunch: Free and Reduced*

Table of Reading CRCT by Reading SBRC

		Reading SBRC			
Reading CRCT					
		DNM	**M**	**E**	**Total**
	DNM	12	2	0	14
		38.71 %	1.00%	0.00%	
	M	18	181	11	210
		58.06 %	90.05 %	26.19 %	
	E	1	18	31	50
		3.23%	8.96%	73.81 %	
Total		31	201	42	274

Test of Symmetry	
Statistic(s)	15.4897
P-Value	0.0014

Research Questions and Results

The study will examine the following research questions to determine how standards-based grade reporting accurately informs student performance on the Criterion References Competency Test and to identify any differences in the means based on gender or sub-groups.

Research Question 1

Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?

Null Hypothesis 1

Average student performance as measured on the math section of the Georgia Criterion

Referenced Competency Test for students in grade five will not be significantly different among students who scored Does not Meet, Meets, and Exceeds on standards-based report cards.

Alternative Hypothesis 1

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Research Question 2

Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?

Null Hypothesis 2

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Alternative Hypothesis 2

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Research Question 3

Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade ELL students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade ELL students different from the differences in mean scores for fifth grade non-ELL students?

Null Hypothesis 3

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for ELL students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for ELL students and non-ELL students.

Alternative Hypothesis 3

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for ELL students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The pattern of these differences for ELL students will be different from the pattern of non-ELL students.

Research Question 4

Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade ELL students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade ELL students different from the differences in mean scores for fifth grade non-ELL students?

Null Hypothesis 4

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for ELL students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for ELL students and non-ELL students.

Alternative Hypothesis 4

Average student performance as measured on the reading section of the Georgia Criterion

Referenced Competency Test for ELL students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The pattern of these differences for ELL students will be different from the pattern of non-ELL students.

Research Question 5

Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade students who receive free or reduced lunch with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade students who receive free or reduced lunch different from the differences in mean scores for fifth grade students who do not receive free or reduced lunch?

Null Hypothesis 5

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for students who receive free or reduced lunch in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for students who do and do not receive free and reduced lunch.

Alternative Hypothesis 5

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for students who receive free or reduced lunch in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The pattern of these differences for students who receive free or reduced lunch will be different from the pattern of students who do not receive free or reduced lunch.

Research Question 6

Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade students who receive free or reduced lunch with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade students who receive free or reduced lunch different from the differences in mean scores for fifth grade students who do not receive free or reduced lunch?

Null Hypothesis 6

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for students who receive free or reduced lunch in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for students who do and do not receive free and reduced lunch.

Alternative Hypothesis 6

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for students who receive free or reduced lunch in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The pattern of these differences for students who receive free or reduced lunch will be different from the pattern of students who do not receive free or reduced lunch.

Research Question 7

Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade female students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?

Null Hypothesis 7

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for female students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Alternative Hypothesis 7

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for female students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Research Question 8

Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade female students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?

Null Hypothesis 8

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for female students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Alternative Hypothesis 8

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for female students in grade five will be significantly different

among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Research Question 9

Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade male students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade male students different from the differences in mean scores for fifth grade female students?

Null Hypothesis 9

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for male students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for male and female fifth grade students.

Alternative Hypothesis 9

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for male students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The pattern of these differences for male and female students will be different.

Research Question 10

Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade male students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade male students different from the differences in mean scores for fifth grade female students?

Null Hypothesis 10

Average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for male students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for male and female fifth grade students.

Alternative Hypothesis 10

Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for male students in grade five will be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards. The pattern of these differences for male and female students will be different.

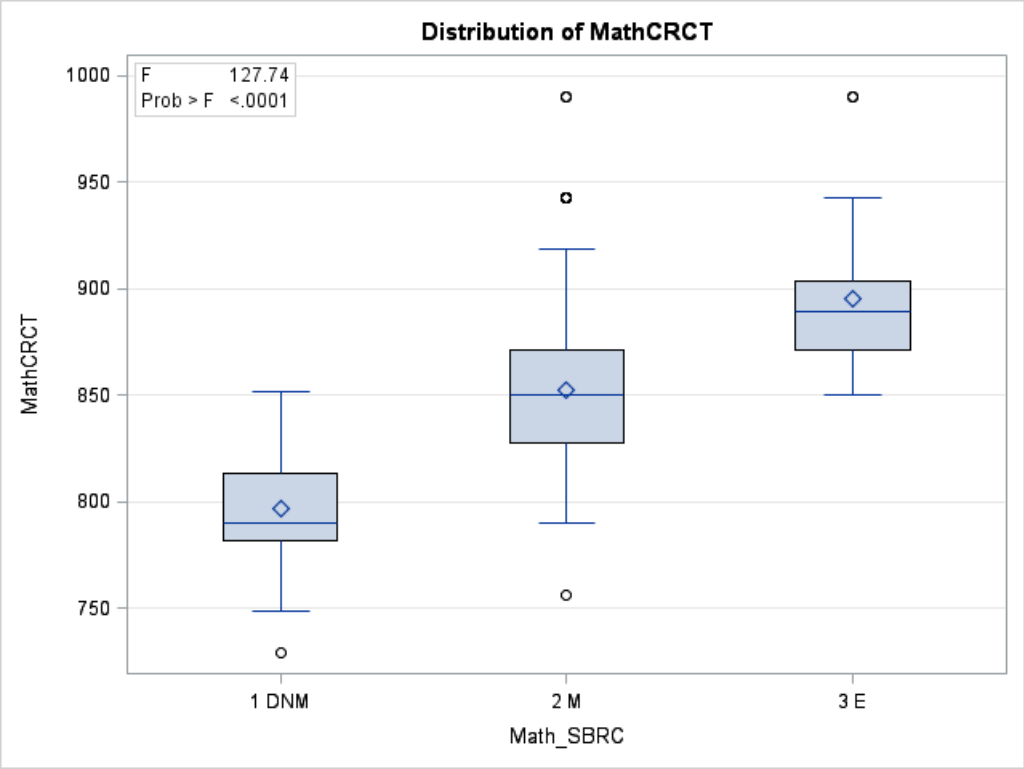
Research question 1. Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?

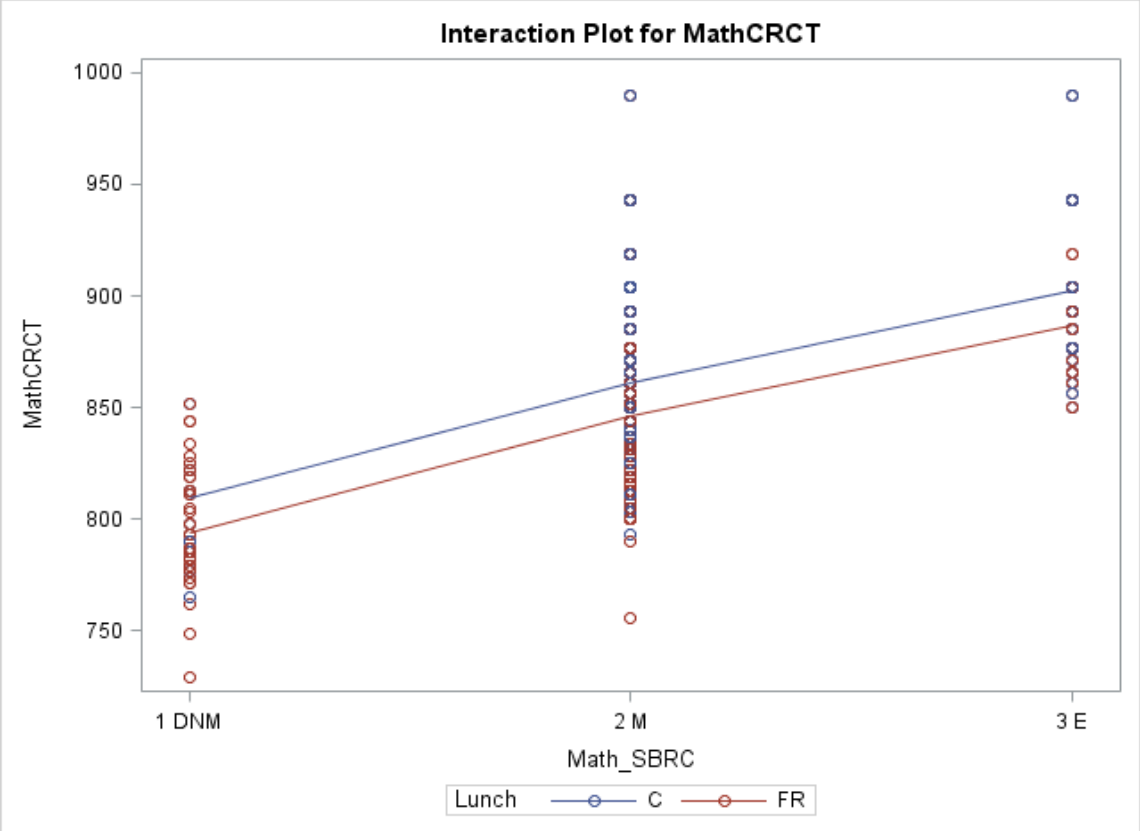
To address research question 1, the researcher retrieved the 2009-2010 SBRC indicators (DNM, M, and E) and CRCT scores of all 5^{th} grade students in the system having both SBRC grades and CRCT scores. The information was pulled from the system's student information system, Infinite Campus.

After retrieving the data, the researcher used a one-way analysis of variance (ANOVA) to determine whether the Math CRCT scores are significantly different among three levels (DNM, M, or E) of Math SBRC. The F statistic was used to do a hypothesis test to find out if the Math CRCT score has a common mean among three Math SBRC groups. ANOVA returns the *p*-value from this hypothesis test. A p-value is a measure of how much evidence there is against the null hypothesis, where the null hypothesis in this study is that all Math CRCT scores have the same

means among three Math SBRC groups. From this test, because the overall p-value is smaller than 0.0001, the researcher rejects null hypothesis 1, concluding that there are significant differences in student performance on the math section of the CRCT based on math grades on standards-based report cards for all fifth grade students.

Furthermore, as shown in Table 11(b) a Tukey multiple comparisons test was used to determine which pairs of Math CRCT means are significantly different (Ollevean, 1999). To interpret this table, examine the first line. The first line compares DNM Math SBRC scores to M Math SBRC scores; the average CRCT score for individuals with a DNM on the SBRC is 796.58, and the average CRCT score for individuals with an M on the SBRC is 852.81. This is a difference of 56.23 points, which is statistically significant at the 0.0001 level. In Table 11, the R^2 of 0.3188 indicates that Math SBRC score is associated with 31.88% of the variation in Match CRCT scores. The RMSE of 33.3710 indicates that if one were to predict the Math CRCT score from Math SBRC, the "typical" error, or "amount off" of that prediction would be 33.3710 points. The results show that all three pairs of means are significantly different from one another, since the P-values are all less than 0.0001. Furthermore, as expected, average CRCT scores are the lowest for students with a math SBRC grade of DNM, and the highest for students with a math SBRC grade of E.

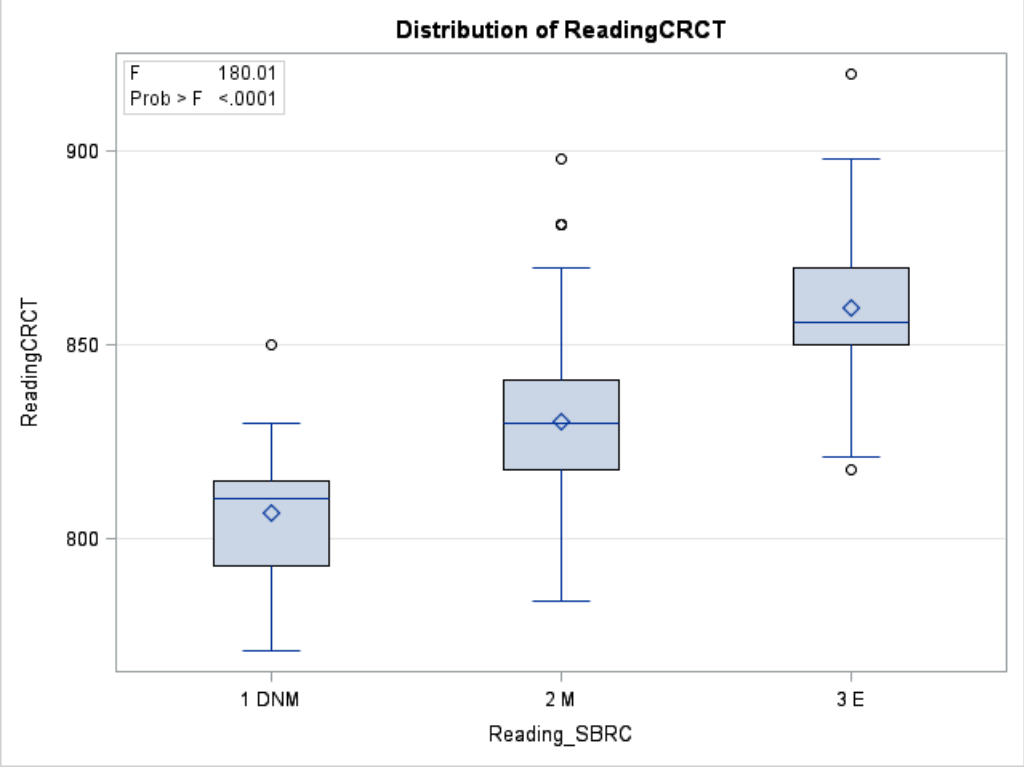

Distribution of MathCRCT
F 127.74
Prob > F <.0001
1000
950
900
850
800
750
MathCRCT
1 DNM
2 M
3 E
Math_SBRC


Interaction Plot for MathCRCT
1000
950
900
850
800
750
MathCRCT
1 DNM
2 M
3 E
Math_SBRC
Lunch
C
FR

Research question 2. Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?

To address research question 2, the researcher retrieved the 2009-2010 SBRC indicators (DNM, M, and E) and CRCT scores of all 5th grade students in the system who had SBRC reading grades and CRCT reading scores. The information was again pulled from the system's student information system, Infinite Campus.

The same analysis was conducted with the reading dataset as for research question one with the math dataset. The researcher conducted a one-way ANOVA, producing a p-value smaller than 0.0001, giving evidence to reject null hypothesis 2. From the tests, the researcher concludes that there is evidence of significant differences of student scores on the reading section of the CRCT based on standards-based report cards.



Research question 3. Are the mean scores on the math section of the Georgia Criterion

Referenced Competency Test different for fifth grade ELL students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade ELL students different from the differences in mean scores for fifth grade non-ELL students?

Using the same dataset, the researcher used a one-way ANOVA to determine that while that there are significant differences in student performance on the math section of the CRCT based on math grades on standards-based report cards for fifth grade ELL students, thus providing evidence to reject the initial part of null hypothesis 3. Since there is need for further investigation into the differences between mean scores for math SBRC, gender, ELL, and socioeconomic status, a final model was used to predict continuous Math CRCT scores from all four of those variables simultaneously to determine whether there was an interaction between ELL and Math SBRC as an indicator of Math CRCT. From this model, the researcher determined that no interactions proved to be significant between ELL and the relationship between math grades on SBRC and math scores on the CRCT. The researcher failed to reject the second part of null hypothesis 3; The lack of differences will be similar for ELL students and non-ELL students.

It is important to note that ELL was not present in the model along with Lunch status, showing that once Lunch status is accounted for, ELL status does not influence the results; however, this could be due to the fact that all but one student with limited English proficiency also received free or reduced lunch, which confounds the effects of these two variables.

Research question 4. Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade ELL students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean

scores for fifth grade ELL students different from the differences in mean scores for fifth grade non-ELL students?

The same analysis was conducted with the reading dataset as for the math dataset. Using the same dataset, the researcher used a one-way ANOVA to determine that that there are significant differences in student performance on the reading section of the CRCT based on reading grades on standards-based report cards for fifth grade ELL students, thus providing evidence to reject the initial part of null hypothesis 4.

Predictor	R^2	RMSE	F-Statistic	P-Value
ReadingSBRC	0.4148	17.8402	180.01	<.0001

Table 16 shows that Reading CRCT scores are significantly different for ELL and non-ELL students, prior to taking SBRC into account. Those who have limited English proficiency score lower than those who do not.

Table 16.

T-test results of Reading CRCT by ELL status

ELL	N	Mean	Std Dev	T Value	P-Value
N	495	836.33	23.117	3.36	0.0009
Y	16	816.69	20.49		

Using an overall model that predicts continuous Reading CRCT scores from all four independent variables simultaneously to determine overall what influences Reading CRCT scores (Reading SBRC, Gender, ELL and Lunch), the researcher gained evidence to reject part two of the null hypothesis 4 as well. Table 17 shows that all variables Reading SBRC, Gender, ELL and Lunch significantly affect the Reading CRCT scores.

Table 17.

ANOVA Results for Prediction of Reading CRCT

Variable	R^2	RMSE	F-Statistic	P-Value
ReadingCRCT	0.4487	17.3666	82.20	<.0001

Predictor	F-Statistic	P-Value
ReadingSBRC	155.41	<.0001
Gender	11.42	0.0008
ELL	4.27	0.0394
Lunch	13.02	0.0003

This result implies that even when accounting for Gender and Lunch status there is still an effect of ELL status. Moreover, ELL status has an effect in the presence of Reading SBRC, which implies that for those with limited proficiency and those who are proficient at English, the same Reading SBRC score may correspond to a different average Reading CRCT score. Results indicate that on average, even with the same Reading SBRC score, those with English proficiency score 9.23 points higher on average than those with limited English proficiency.

Research question 5. Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade students who receive free or reduced lunch with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade students who receive free or reduced lunch different from the differences in mean scores for fifth grade students who do not receive free or reduced lunch?

To address research question 5, the researcher used a one-way analysis of variance to determine evidence against the null hypothesis. The p-value of less than 0.0001 provides evidence to reject null hypothesis 5 stating that average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for students who receive free or reduced lunch in grade five will not be significantly different among students who scored

Does Not Meet, Meets, and Exceeds on standards-based report cards. The lack of differences will be similar for students who do and do not receive free and reduced lunch.

Once this was established, further analysis was done to show interactions between groupings. To determine whether the socioeconomic status corresponds to different Math CRCT scores, a t-test was used compare the means of Math CRCT score by socioeconomic status. In a T-test, a t statistic is computed and compared it to a critical value. The critical value is chosen so that if the means are really the same in the population, the probability (p-value) that the t statistic will exceed the critical value is small; 5% (or 0.05) is used in this case, as is typical. When the means are different, the probability that the statistic will exceed the critical value is larger. For socioeconomic status, the mean Math CRCT score of students who received at-cost lunch is significantly different from the mean score from students who received free or reduced lunch (it is higher) with a p-value of < 0.0001 (very significant).

Table 12.

T-test results of Math CRCT Socioeconomic status

Status	N	Mean	Std Dev	T Statistic	P-Value
At cost	240	865.72	39.715	7.13	<.0001
FR	309	842.00	37.779		

An overall model was used to predict continuous Math CRCT scores from all four of those variables simultaneously to determine overall what influences Math CRCT scores (Math SBRC, Gender, ELL and Lunch). At the beginning, the researcher included the two-way interaction effects of Math SBRC*Gender, Math SBRC*ELL, and Math SBRC*Lunch in this model. As an example of the interpretation of an interaction, suppose there were a significant interaction between Math SBRC and ELL: this would indicate that not only does ELL influence the CRCT scores, but that CRCT also influences the relationship between CRCT and SBRC.

However, no interactions proved to be significant, so all the interaction terms were removed and the researcher chose the variables that have a significant effect on Math CRCT based on the P-values. Table 13 shows that only variables Math SBRC and Lunch significantly affect the Math CRCT scores.

Table 13.

ANOVA Results for Prediction of Math CRCT

Variable	R^2	RMSE	F-Statistic	P-Value
MathCRCT	0.3540	32.5265	99.54	<.0001

Predictor	F-Statistic	P-Value
MathSBRC	113.40	<.0001
Lunch	29.71	<.0001

The presence of both Math SBRC and Lunch status in the model indicate that even when two students have similar SBRC scores, whether or not a student is on free/reduced lunch influences his or her Math CRCT score. Additional results indicate that even if students have the same Math SBRC score, a student with At-Cost lunch will, on average, score 15.55 points higher on the Math CRCT than those with Free/Reduced lunch.

Next, a two-way tables of categorized Math CRCT score * Math SBRC, a categorical version of the SBRC scores, along with a test of symmetry, determines whether deviations from the diagonal are random, or if they tend to be in one direction. That is, the test of symmetry would indicate whether Math SBRC scores are biased in their correspondence to Math CRCT scores. These tests can show whether categorized Math SBRC scores are typically matched to their respective equivalents in the Math CRCT scores. The researcher also examines these two-way tables according to socioeconomic status, respectively.

Table of Math CRCT by Math SBRC

	MathSBRC			
MathCRCT	**DNM**	**M**	**E**	**Total**
DNM	6	1	0	7
	66.67%	0.52%	0.00%	
M	2	70	0	72
	22.22%	36.08%	0.00%	
E	1	123	37	161
	11.11%	63.40%	100.00 %	
Total	9	194	37	240

Test of Symmetry	
Statistic(s)	124.3333
P-Value	<.0001

Table 14(e)

Lunch: Free/Reduced

Table of MathCRCT by MathSBRC

	MathSBRC			
MathCRCT	**DNM**	**M**	**E**	**Total**
DNM	27	2	0	29
	61.36	0.84	0	
M	16	136	0	152
	36.36	57.14	0	
E	1	100	27	128
	2.27	42.02	100.00	
Total	44	238	27	309

Test of Symmetry	
Statistic(s)	111.8889
P-Value	<.0001

Research question 6. Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade students who receive free or reduced lunch with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade students who receive free or reduced lunch different from the differences in mean scores for fifth grade students who do not receive free or reduced lunch?

To address research question 6, the researcher retrieved the 2009-2010 SBRC indicators (DNM, M, and E) and CRCT scores of all 5th grade students in the system with SBRC reading grades and CRCT reading scores. The information was pulled from the system's database, Infinite Campus.

With the reading dataset, the researcher conducted a one-way ANOVA, producing a p-value smaller than 0.0001, giving evidence to reject null hypothesis 6 stating that average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for students who receive free or reduced lunch in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

The researcher conducted a T-test to determine the expected Reading CRCT sores based on gender, ELL, and free/reduced lunch status. Table 16 shows that for Gender, ELL, and Lunch, Reading CRCT scores are significantly different from one another. Females score significantly higher than males; those who have limited English proficiency score lower than those who do not; and those on the free/reduced lunch program score lower than those who purchase lunch at cost.

Table 16.

T-test results of Reading CRCT by Gender, ELL and Lunch

Gender	N	Mean	Std Dev	T Value	P-Value
F	239	838.57	21.954	2.62	0.0092
M	272	833.2	24.137		

ELL	N	Mean	Std Dev	T Value	P-Value
N	495	836.33	23.117	3.36	0.0009
Y	16	816.69	20.49		

Lunch	N	Mean	Std Dev	T Value	P-Value
C	237	843.03	23.01	6.91	<.0001
FR	274	829.38	21.623		

Finally, the research uses a model to predict continuous Reading CRCT scores from all four of those variables simultaneously to determine overall what influences Reading CRCT scores (Reading SBRC, Gender, ELL and Lunch). Table 17 shows that all variables including Reading SBRC, Gender, ELL and Lunch significantly affect the Reading CRCT scores.

Table 17.

ANOVA Results for Prediction of Reading CRCT

Variable	R^2	RMSE	F-Statistic	P-Value
ReadingCRCT	0.4487	17.3666	82.20	<.0001

Predictor	F-Statistic	P-Value
ReadingSBRC	155.41	<.0001
Gender	11.42	0.0008
ELL	4.27	0.0394
Lunch	13.02	0.0003

This result implies that even when accounting for Lunch status and ELL status, there is still an effect of Gender; even when accounting for Gender and ELL status there is still an effect of Lunch status, and even when accounting for Gender and Lunch status there is still an effect of ELL status. Additionally, all of these have an effect in the presence of Reading SBRC, which

implies that for males and females, for those with limited proficiency and those who are proficient at English, and for those on free/reduced lunch status or not, the same Reading SBRC score may predict a different Reading CRCT score. Results indicate that even with the same Reading SBRC score, those who pay for lunch at cost score 5.80 points higher on average than those on free/reduced lunch.

Research question 7. Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade female students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?

Using the same math dataset, the researcher used one-way ANOVA model to return a P-value smaller than 0.0001, giving evidence to reject null hypothesis 7 stating that Average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for female students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Then, further analysis was conducted to determine trends in the subgroups.
To determine whether the demographic variables influence the Math CRCT scores, t-tests were used compare the means of Math CRCT score by Gender, ELL and Lunch. In a T-test, a t statistic is computed and compared it to a critical value. The critical value is chosen so that if the means are really the same in the population, the probability (p-value) that the t statistic will exceed the critical value is small; 5% (or 0.05) is used in this case, as is typical. When the means are different, the probability that the statistic will exceed the critical value is larger. The p-value for Gender is 0.2158, indicating there is no significant difference in Math CRCT scores based on Gender.

Research question 8. Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade female students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?

Using the reading dataset, the researcher used one-way ANOVA model to return a P-value smaller than 0.0001, giving evidence to reject null hypothesis 8 stating that average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for female students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Once the determination was made that there is a difference in CRCT means based on SBRC for students in all groups, further analysis was conducted to determine trends in the subgroups.

The researcher use the Tukey multiple comparisons test in Table 15(b) to show that all three pairs of Reading CRCT means are significantly different.

Table 15 (a).

ANOVA Results for Prediction of Reading CRCT from Reading SBRC scores

Predictor	R^2	RMSE	F-Statistic	P-Value
ReadingSBRC	0.4148	17.8402	180.01	<.0001

Table 15(b).

Least Squares Means Comparisons for Reading CRCT Score by Reading SBRC Score

ReadingSBRC 1	Reading SBRC2	Average CRCT1	Average CRCT2	Difference	P-value
DNM	**M**	806.79	830.23	23.44	<0.0001
DNM	**E**	806.79	859.77	52.98	<0.0001
M	**E**	830.23	859.77	29.54	<0.0001

Table 16 shows that for Gender, ELL, and Lunch, Reading CRCT scores are significantly different from one another. Females score significantly higher than males.

Table 16.

T-test results of Reading CRCT by Gender, ELL and Lunch

Gender	N	Mean	Std Dev	T Value	P-Value
F	239	838.57	21.954	2.62	0.0092
M	272	833.2	24.137		

ELL	N	Mean	Std Dev	T Value	P-Value
N	495	836.33	23.117	3.36	0.0009
Y	16	816.69	20.49		

Lunch	N	Mean	Std Dev	T Value	P-Value
C	237	843.03	23.01	6.91	<.0001
FR	274	829.38	21.623		

An overall model predicts continuous Reading CRCT scores from all four of those variables simultaneously to determine overall what influences Reading CRCT scores (Reading SBRC, Gender, ELL and Lunch). Table 17 shows that all variables Reading SBRC, Gender, ELL and Lunch significantly affect the Reading CRCT scores.

Table 17.

ANOVA Results for Prediction of Reading CRCT

Variable	R^2	RMSE	F-Statistic	P-Value
ReadingCRCT	0.4487	17.3666	82.20	<.0001

Predictor	F-Statistic	P-Value
ReadingSBRC	155.41	<.0001
Gender	11.42	0.0008
ELL	4.27	0.0394
Lunch	13.02	0.0003

This result implies that even when accounting for Lunch status and ELL status, there is still an effect of Gender; even when accounting for Gender and ELL status there is still an effect

of Lunch status, and even when accounting for Gender and Lunch status there is still an effect of ELL status. Additionally, all of these have an effect in the presence of Reading SBRC, which implies that for males and females, for those with limited proficiency and those who are proficient at English, and for those on free/reduced lunch status or not, the same Reading SBRC score may predict a different Reading CRCT score. Results indicate that on average, even with the same Reading SBRC score, females score 5.22 points higher on average than males on the Reading CRCT.

Tables 18(a)-(e) show the two-way tables of categorized Reading CRCT score * Reading SBRC score, along with a test of symmetry. The p-values of the tests of symmetry are less than 0.05 except when gender= male, which indicates that we need to reject the null hypothesis that the cell proportions are symmetric in most cases. Typically there is a similar trend as in Math that Reading SBRC scores tend to underestimate Reading CRCT score categories, though it doesn't appear to be as severe as with Math.

Tables 18(a)-(e).

Frequency Distributions of Reading CRCT vs. Reading SBRC and Results of Test of Symmetry

Table 18(a) *Overall*

Table of ReadingCRCT by ReadingSBRC

		ReadingSBRC			
ReadingCRCT					
		DNM	**M**	**E**	**Total**
	DNM	12	5	0	17
		31.58 %	1.44%	0.00%	
	M	25	290	28	343
		65.79 %	83.33 %	22.40 %	
	E	1	53	97	151
		2.63%	15.23 %	77.60 %	
Total		38	348	125	511

Test of Symmetry	
Statistic(s)	22.0494
P-Value	<.0001

Table 18(b) *Gender: Female*

Table of Reading CRCT by Reading SBRC

		Reading SBRC			
Reading CRCT					
		DNM	**M**	**E**	**Total**
	DNM	2	1	0	3
		14.29 %	0.59%	0.00%	
	M	11	136	8	155
		78.57 %	80.47 %	14.29 %	
	E	1	32	48	81
		7.14%	18.93 %	85.71 %	
Total		14	169	56	239

Test of Symmetry	
Statistic(s)	23.7333
P-Value	<.0001

Research question 9. Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade male students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade male students different from the differences in mean scores for fifth grade female students?

To address research question 9, the researcher used the same math data set. First, a one-way analysis of variance (ANOVA) was used to determine whether the Math CRCT scores are significantly different among three levels of Math SBRC. The F statistic is used to do a hypothesis test to find out if the Math CRCT score has a common mean among three Math SBRC groups. ANOVA returns the *p*-value from this hypothesis test. A p-value is a measure of how much evidence we have against the null hypothesis, where the null hypothesis in this study is that all Math CRCT scores have the same means among three Math SBRC groups. The p-value is interpreted as the probability that a population where all groups of students have the same average Math SBRC scores could produce a sample where the differences among the groups as large (or larger) than those seen in our study. Therefore, a small p-value (typically 0.05 of smaller) is evidence against the null hypothesis, while a large p-value means little or no evidence against the null hypothesis**.** From Table 11(a), the overall p-value is smaller than 0.0001. This is a strong indication that the average Math CRCT scores from the different Math SBRC groups are not the same. Differences in Math CRCT scores among the groups as large as those seen in this data set, and therefore an *F* statistic as extreme as the observed F, would occur

by chance less than once in 10,000 times if the scores were truly equal in the population. Furthermore, in Table 11(b) a Tukey multiple comparisons test was used to determine which pairs of Math CRCT means are significantly different. To interpret this table, we examine the first line. The first line compares DNM Math SBRC scores to M Math SBRC scores; the average CRCT score for individuals with a DNM on the SBRC is 796.58, and the average CRCT score for individuals with an M on the SBRC is 852.81. This is a difference of 56.23 points, which is statistically significant at that 0.0001 level. In Table 11, the R^2 of 0.3188 indicates that Math SBRC score is associated with 31.88% of the variation in Match CRCT scores. The RMSE of 33.3710 indicates that if one were to predict the Math CRCT score from Math SBRC, the "typical" error (or "amount off") of that prediction would be 33.3710 points. The results show that all three pairs of means are significantly different from one another, since the P-values are all less than 0.0001, giving evidence to reject the initial part of null hypothesis 9 that average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for male students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards.

Table 11 (a).

ANOVA Results for Prediction of Math CRCT from Math SBRC Scores

Predictor	R^2	RMSE	F-Statistic	P-Value
MathSBRC	0.3188	33.3710	127.74	<.0001

Table 11(b).

Least Squares Means Comparisons for Math CRCT Score by Math SBRC Score

MathSBRC1	Math SBRC2	Average CRCT1	Average CRCT2	Difference	P-value
DNM	**M**	796.58	852.81	56.23	<0.0001
DNM	**E**	796.58	895.58	99.00	<0.0001
M	**E**	852.81	895.58	42.77	<0.0001

Next, to determine whether the demographic variables influence the Math CRCT scores, t-tests were used compare the means of Math CRCT score by Gender, ELL and Lunch. In a T-test, a t statistic is computed and compared it to a critical value. The critical value was chosen so that if the means were really the same in the population, the probability (p-value) that the t statistic will exceed the critical value is small; 5% (or 0.05) is used in this case, as is typical. When the means are different, the probability that the statistic will exceed the critical value is larger. The p-value for Gender is 0.2158, indicating there is no significant difference in Math CRCT scores based on Gender, giving the researcher evidence to also reject the second part of null hypothesis 9 stating that the lack of differences will be similar for male and female fifth grade students.

Furthermore, the researcher concludes that there is no significant effect of gender on Math CRCT scores.

Table 12.

T-test results of Math CRCT by Gender, ELL and Lunch

Gender	N	Mean	Std Dev	T Statistic	P-Value
F	262	850.02	37.607	1.24	0.2158
M	287	845.39	42.678		

Research question 10. Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade male students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade male students different from the differences in mean scores for fifth grade female students?

To address research question 10, the researcher first rejected the initial part of null hypothesis 10 stating that average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for male students in grade five will not be significantly different among students who scored Does Not Meet, Meets, and Exceeds on standards-based report cards, based on the information gained using the reading dataset and the one-way ANOVA model to return a P-value smaller than 0.0001. After determining that there is an expected alignment between grades on SBRC and performance on the GA CRCT for fifth grade male students, the researcher completed further analysis to whether there are differences in mean CRCT scores for fifth grade male and female students, based on SBRC.

Table 15 shows that the P-value of one-way ANOVA model is smaller than 0.0001, which indicates that the average Reading CRCT scores from the different Reading SBRC groups are not the same. Furthermore, the Tukey multiple comparisons test in Table 15(b) shows that all three pairs of Reading CRCT means are significantly different.

Table 15 (a).

ANOVA Results for Alignment of Reading CRCT to Reading SBRC scores

Predictor	R^2	RMSE	F-Statistic	P-Value
ReadingSBRC	0.4148	17.8402	180.01	<.0001

Table 15(b).

Least Squares Means Comparisons for ReadingCRCT Score by ReadingSBRC Score

ReadingSBRC 1	Reading SBRC2	Average CRCT1	Average CRCT2	Difference	P-value
DNM	**M**	806.79	830.23	23.44	<0.0001
DNM	**E**	806.79	859.77	52.98	<0.0001
M	**E**	830.23	859.77	29.54	<0.0001

Table 16 shows that for Gender, ELL, and Lunch, Reading CRCT scores are significantly different from one another. Females score significantly higher than males; those who have limited English proficiency score lower than those who do not; and those on the free/reduced lunch program score lower than those who purchase lunch at cost.

Table 16.

T-test results of Reading CRCT by Gender, ELL and Lunch

Gender	N	Mean	Std Dev	T Value	P-Value
F	239	838.57	21.954	2.62	0.0092
M	272	833.2	24.137		

ELL	N	Mean	Std Dev	T Value	P-Value
N	495	836.33	23.117	3.36	0.0009
Y	16	816.69	20.49		

Lunch	N	Mean	Std Dev	T Value	P-Value
C	237	843.03	23.01	6.91	<.0001
FR	274	829.38	21.623		

A final model predicts continuous Reading CRCT scores from all four of those variables simultaneously to determine overall what influences Reading CRCT scores (Reading SBRC, Gender, ELL and Lunch). Table 17 shows that all variables Reading SBRC, Gender, ELL and Lunch significantly affect the Reading CRCT scores.

Table 17.

ANOVA Results for Prediction of Reading CRCT

Variable	R^2	RMSE	F-Statistic	P-Value
ReadingCRCT	0.4487	17.3666	82.20	<.0001

Predictor	F-Statistic	P-Value
ReadingSBRC	155.41	<.0001
Gender	11.42	0.0008
ELL	4.27	0.0394
Lunch	13.02	0.0003

This result implies that even when accounting for Lunch status and ELL status, there is still an effect of Gender; even when accounting for Gender and ELL status there is still an effect of Lunch status, and even when accounting for Gender and Lunch status there is still an effect of ELL status. Additionally, all of these have an effect in the presence of Reading SBRC, which implies that for males and females, for those with limited proficiency and those who are proficient at English, and for those on free/reduced lunch status or not, the same Reading SBRC score may predict a different Reading CRCT score. Results indicate that on average, even with the same Reading SBRC score, females score 5.22 points higher on average than males on the Reading CRCT; even with the same Reading SBRC score, those with English proficiency score 9.23 points higher on average than those with limited English proficiency; and even with the same Reading SBRC score, those who pay for lunch at cost score 5.80 points higher on average than those on free/reduced lunch.

Tables 18(a)-(e) show the two-way tables of categorized Reading CRCT score * Reading SBRC score, along with a test of symmetry. The p-values of the tests of symmetry are less than 0.05 except when gender= male, which indicates that we need to reject the null hypothesis that the cell proportions are symmetric in most cases. Typically there is a similar trend as in Math that Reading SBRC scores tend to underestimate Reading CRCT score categories, though it

doesn't appear to be as severe as with Math. For males, however, Reading SBRC scores seem to present a balanced picture of Reading CRCT scores.

Tables 18(a)-(e).

Frequency Distributions of Reading CRCT vs. Readomg SBRC and Results of Test of Symmetry

Table 18(a) *Overall*

Table of Reading CRCT by Reading SBRC

		Reading SBRC			
Reading CRCT		**DNM**	**M**	**E**	**Total**
	DNM	12	5	0	17
		31.58%	1.44%	0.00%	
	M	25	290	28	343
		65.79%	83.33%	22.40%	
	E	1	53	97	151
		2.63%	15.23%	77.60%	
Total		38	348	125	511

Test of Symmetry

Statistic(s)	22.0494
P-Value	<.0001

Table 18(c) *Gender: Male*

Table of Reading CRCT by Reading SBRC

		Reading SBRC			
Reading CRCT					
		DNM	**M**	**E**	**Total**
	DNM	10	4	0	14
		41.67 %	2.23%	0.00%	
	M	14	154	20	188
		58.33 %	86.03 %	28.99 %	
	E	0	21	49	70
		0.00%	11.73 %	71.01 %	
Total		24	179	69	272

Test of Symmetry	
Statistic(s)	5.5799
P-Value	0.1339

Summary

For both Math and Reading CRCT scores, it was clear that there are large and significant differences in these scores that depend on one's Math or Reading SBRC score. Additionally, for Math, even with the same SBRC score, students who are on the free/reduced lunch program tend to score lower than students who are not. For Reading, this result extended even further: even with the same SBRC score, females tend to score higher than males, those with limited English proficiency tend to score lower than those who are proficient, and those who are on the free/reduced lunch program tend to score lower than those who are. This means that, for some reason, SBRC scores do not represent all groups similarly when it comes to providing accurate information that would correspond with CRCT scores.

Additionally, for almost every group, when CRCT scores are categorized as DNM, M, or E, SBRC scores tend to underestimate these CRCT score categories. Often, students who do not meet SBRC standards will meet CRCT standards, and students who meet SBRC standards will exceed CRCT standards. This was more significant and noticeable for Math than for Reading; the only exception to this general trend was that for males, Reading SBRC scores tend to be fairly representative of Reading CRCT scores.

CHAPTER FIVE: DISCUSSION

Sound assessment and grade reporting procedures provide the foundation for effective standards-based instruction. Nationally, standards-based instruction is at the forefront of reform, with the introduction of Common Core State Standards providing states a clear picture of how standards can provide a focus for learning that allows students to build on previous learning while continuously working towards a deeper level of understanding. By clearly establishing what students should know and be able to do at each grade level and focusing on conceptual understandings, teachers are able to pinpoint instruction, thus individualizing the learning process for the student (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010). This shift in thinking about instruction requires sound assessment practices that report learning in terms of the standards.

Restatement of the Problem

Traditional grading systems often include many factors that may not actually reflect student mastery of standards or most current learning. Often, grades can be inflated by effort, homework, or participation, thus having little predictive value on student success. In a system that requires accountability on all levels, and seeks to individualize instruction to meet the needs of students, assessment and grade-reporting must provide a clear picture of student learning. Standards-based assessment and grade-reporting provide a natural follow up to standards-based instruction; however, little research has been completed to show whether standards-based grading gives an accurate predictor of student success on high-stakes accountability measures and whether standards-based grading provides a fair grading system for all students.

One of the goals of standards-based grade reporting is to reduce subjectivity in grading, thus providing accurate information about student learning that is unaffected by various other

factors such as participation, behavior, or parental involvement. An examination of sub-groups is needed to indicate whether standards-based grading truly does provide a system that is fair to all students.

Research Questions

Research question 1. *Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?*

The researcher used a one-way analysis of variance (ANOVA) to determine whether the patterns of Math CRCT scores are significantly different among three levels (DNM, M, or E) of Math SBRC. The researcher rejects null hypothesis 1, stating that average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for students in grade five will not be significantly different among students who scored Does not Meet, Meets, and Exceeds on standards-based report cards. The results of the Tukey multiple comparisons test show that all three pairs of means are significantly different from one another, since the P-values are all less than 0.0001.

Furthermore, the data shows that for almost every group, when CRCT scores are categorized as DNM, M, or E, SBRC scores tend to underestimate these CRCT score categories. Often, students who do not meet SBRC standards will meet CRCT standards, and students who meet SBRC standards will exceed CRCT standards. This was more significant and noticeable for Math than for Reading.

This finding supports the idea found in the Review of Literature that standards-based report cards provide a quality alternative to relying completely on high-stakes tests to show

student learning, and that decisions could be made about students on the basis of patterns of scores obtained over time through standards-based assessment (Sherer, 2001).

Research question 2. *Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?*

Through the same analysis conducted with the reading dataset, the researcher confirmed that the P-value of one-way ANOVA model is smaller than 0.0001, indicating that while there is alignment between SBRC and CRCT scores for all students, the average Reading CRCT scores from the different Reading SBRC groups are not the same. Furthermore, the Tukey multiple comparisons test shows that all three pairs of Reading CRCT means are significantly different, thus providing evidence to reject null hypothesis 2, stating that average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for students in grade five will not be significantly different among students who scored DNM, M, and E on standards-based report cards.

This finding concurs with the research of D'Agostino and Welsh (2007) which investigated the question of whether standards-based grading yields accurate information about student performance on standardized tests and lends support to the practice of standards-based grading for the purpose of strengthening the relationship between teacher judgments through grading and proficiency levels on the Arizona Instrument to Measure Standards. Furthermore, the data highlights previous findings of Brennan, Kim, Wenz-Gross, & Siperstein (2001) which shows that traditional grades, not based on standards, and state test results are only moderately related due to inclusion of non-achievement factors that factor in to traditional grading practices.

Research question 3. *Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade ELL students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade ELL students different from the differences in mean scores for fifth grade non-ELL students?*

Using the same math dataset, the researcher confirmed, through an ANOVA, that there is an expected alignment between math grades on standards-based report cards and student performance on the math section of the CRCT for fifth grade ELL students, thus providing evidence to reject null hypothesis 3, stating that the average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for ELL students in grade five will not be significantly different among students who scored DNM, M, and E on standards-based report cards.

However, there is need for further investigation into the differences between mean scores for math SBRC, gender, ELL, and socioeconomic status. The researcher further determined that no interactions proved to be significant between ELL and the relationship between math grades on SBRC and math scores on the CRCT, and given a particular SBRC score ELL and non-ELL students score similarly, giving the researcher reason to fail to reject the second part of null hypothesis 3.

This finding is consistent with the information in the review of literature showing that ELLs are generally lower across grade and subject levels (Abedi, 2002; Abedi & Lord, 2001; Abedi, Lord, & Hofstetter, 1998). Additionally, this supports the research showing the magnitude of the gap in achievement between ELLs and non-Ells as being greatest for tests that

require more verbal processing skills, such as language arts, and least for mathematics assessments, as shown in the NAEP report for grades 4 and 8 (U.S. DOE, 2011).

Research question 4. *Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade ELL students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade ELL students different from the differences in mean scores for fifth grade non-ELL students?*

The analysis conducted with the reading dataset indicated that the average Reading CRCT scores from the different Reading SBRC groups are not the same for ELL students, and that there is sufficient evidence to reject null hypothesis 4, based on the expected alignment between SBRC grades and CRCT scores for all students. While the researcher confirmed that there is a significant correspondence between SBRC and CRCT for all students, The Tukey multiple comparisons test in Table 15(b) shows that all three pairs of Reading CRCT means are significantly different. Furthermore, the researcher confirmed that those who have limited English proficiency score lower than those who do not, even with the same SBRC score, which concurs with the research in the review of literature showing that ELLs are generally lower across grade and subject levels (Abedi, 2002; Abedi & Lord, 2001; Abedi, Lord, & Hofstetter, 1998).

The researcher determined through the use of t-tests that all variables including Reading SBRC, Gender, ELL and Lunch significantly affect the Reading CRCT scores.
Results indicate that on average, even with the same Reading SBRC score, those with English proficiency score 9.23 points higher on average than those with limited English proficiency.

Research question 5. *Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade students who receive free or reduced lunch with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade students who receive free or reduced lunch different from the differences in mean scores for fifth grade students who do not receive free or reduced lunch?*

Using an analysis of variance, the researcher gained evidence to reject null hypothesis 5 stating that the average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for students who receive free or reduced lunch in grade five will not be significantly different among students who scored DNM, M, and E on standards-based report cards.

The data shows there is evidence of alignment between CRCT and SBRC in every case, so further analysis was done to show patterns of interactions between groupings. The presence of both Math SBRC and Lunch status in the model indicated that even when considering Math SBRC scores, whether or not a student is on free/reduced lunch influences his or her Math CRCT score. Additional results indicate that even if students have the same Math SBRC score, a student with At-Cost lunch will, on average, score 15.55 points higher on the Math CRCT than those with Free/Reduced lunch. This finding supports the research of Waber, Gerber, Turcios, Wagner & Forbes (2006), indicating that children from disadvantaged backgrounds perform more poorly than their more advantaged peers.

To determine whether the demographic variables influence the Math CRCT scores, t-tests were used compare the means of Math CRCT score by Gender, ELL and Lunch. The researcher confirmed for the variable Lunch, the mean Math CRCT score of students who received at-cost

lunch is significantly different from the mean score from students who received free or reduced lunch. Those students who purchase lunch were found to have significantly higher mean scores, with a p-value of < 0.0001.

Research question 6. *Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade students who receive free or reduced lunch with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade students who receive free or reduced lunch different from the differences in mean scores for fifth grade students who do not receive free or reduced lunch?*

The researcher rejected null hypothesis 6 stating that the average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for students who receive free or reduced lunch in grade five will not be significantly different among students who scored DNM, M, and E on standards-based report cards.

This result implies that even when even with the same Reading SBRC score, those who pay for lunch at cost score 5.80 points higher on average than those on free/reduced lunch. This finding concurs with research by Newman and Chin (2003), showing that students with below average skills tend to have a family that received food stamps or welfare, a single parent household, or non-English speaking parents.

Through T-tests, the researcher determined the expected Reading CRCT scores based on gender, ELL, and free/reduced lunch status, and confirmed that for Gender, ELL, and Lunch, Reading CRCT scores are significantly different from one another. Furthermore, those on the free/reduced lunch program score lower than those who purchase lunch at cost. Finally, the

researcher determined that all variables including Reading SBRC, Gender, ELL and Lunch significantly affect the Reading CRCT scores.

Research question 7. *Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade female students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?*

Using the same math dataset, the researcher used an ANOVA to reject null hypothesis 7 stating that the average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for female students in grade five will not be significantly different among students who scored DNM, M, and E on standards-based report cards. Based on the use of ANOVA, the researcher confirmed that there is alignment between SBRC and CRCT performance for students in all groups. However, further analysis was conducted to determine differences among subgroups.

Research question 8. *Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade female students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards?*

Using the reading dataset, the researcher used an ANOVA to reject null hypothesis 8 stating that the average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for female students in grade five will not be significantly different among students who scored DNM, M, and E on standards-based report cards. Once the determination was made that there is a correspondence between SBRC and CRCT performance for students in all groups, further analysis confirmed through a Tukey multiple comparisons test that all three pairs of Reading CRCT means are significantly different.

Research question 9. *Are the mean scores on the math section of the Georgia Criterion Referenced Competency Test different for fifth grade male students with math grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade male students different from the differences in mean scores for fifth grade female students?*

Through a one-way analysis of variance (ANOVA) the researcher determined that while there is an expected alignment between SBRC and CRCT math scores, the Math CRCT scores are significantly different among three levels of Math SBRC. The F statistic was used to do a hypothesis test to find out if the Math CRCT score has a common mean among three Math SBRC groups, and then a Tukey multiple comparisons test determined which pairs of Math CRCT means are significantly different. From this test, the researcher found that the average CRCT score for individuals with a DNM on the SBRC is 796.58, and the average CRCT score for individuals with an M on the SBRC is 852.81. This is a difference of 56.23 points, which is statistically significant at that 0.0001 level. The data also indicate that Math SBRC score is associated with 31.88% of the variation in Match CRCT scores. The researcher confirmed that all three pairs of means are significantly different from one another, since the P-values are all less than 0.0001.

The analysis confirmed that there is an expected correspondence between SBRC and CRCT for all student, regardless of gender, thus leading the researcher to reject null hypothesis 9 stating the average student performance as measured on the math section of the Georgia Criterion Referenced Competency Test for male students in grade five will not be significantly different among students who scored DNM, M, and E on standards-based report cards. The lack

of differences will be similar for male and female fifth grade students. Furthermore, the researcher concludes that there is no significant effect of gender on Math CRCT scores.

This finding concurs with the research of Entwisle, Alexander, and Olson (1994) stating that math achievement of boys and girls is about equal in elementary school. Additionally,this finding parallels the trends in NAEP mathematics at grade 4 and 8 in 2009 which show no significant change in score based on gender (U.S. DOE, 2011).

Research question 10. *Are the mean scores on the reading section of the Georgia Criterion Referenced Competency Test different for fifth grade male students with reading grades of Does Not Meet, Meets, and Exceeds on standards-based report cards? Are the differences in mean scores for fifth grade male students different from the differences in mean scores for fifth grade female students?*

Through a one-way ANOVA, the researcher rejected the null hypothesis stating that the average student performance as measured on the reading section of the Georgia Criterion Referenced Competency Test for male students in grade five will not be significantly different among students who scored DNM, M, and E on standards-based report cards, and the lack of differences will be similar for male and female fifth grade students. After determining that there is an expected alignment between grades on SBRC and performance on the CRCT for fifth grade male students, the researcher completed further analysis to determine any differences between the scores of males on SBRC and CRCT.

The researcher confirmed that for Gender, ELL, and Lunch, Reading CRCT scores are significantly different from one another. Females score significantly higher than males; those who have limited English proficiency score lower than those who do not; and those on the free/reduced lunch program score lower than those who purchase lunch at cost.

The data confirmed that even when accounting for Lunch status and ELL status, there is still an effect of Gender; even when accounting for Gender and ELL status there is still an effect of Lunch status, and even when accounting for Gender and Lunch status there is still an effect of ELL status. Additionally, the analysis revealed an effect in the presence of Reading SBRC, which implies that for males and females, for those with limited proficiency and those who are proficient at English, and for those on free/reduced lunch status or not, the same Reading SBRC score may predict a different Reading CRCT score. The analysis indicated that on average, even with the same Reading SBRC score, females score 5.22 points higher on average than males on the Reading CRCT; even with the same Reading SBRC score, those with English proficiency score 9.23 points higher on average than those with limited English proficiency; and even with the same Reading SBRC score, those who pay for lunch at cost score 5.80 points higher on average than those on free/reduced lunch.

The data revealed a similar trend as in Math that Reading SBRC scores tend to underestimate Reading CRCT score categories, though it does not appear to be as severe as with Math. For males, however, the researcher shows that Reading SBRC scores seem to present a balanced picture of Reading CRCT scores. The idea that teachers may be holding students to a higher standards than state tests supports research findings of Peterson and Xabel Lastra-Anadon (2010) that every state, for both reading and math (with the exception of Massachusetts for math), deems more students “proficient” on their own state assessments than the National Assessment of Educational Progress NAEP does, showing that state tests typically set the bar low. The idea that teachers have higher expectations for their students than does the state parallels the finding in recent studies that there is a clear contrast between NAEP standards and state standards as measured by state tests (and Xabel Lastra-Anadon, 2010).

This finding concurs with the National Assessment of Education Progress (NAEP) results, 10% more female students scored proficient in reading, and in 2000, females outperformed boys at all age levels, a larger gap than in 1998, despite accommodations (Donahue et al 2001). According to a study by Kafer (2005) this trend continues in middle school, high school, and post-secondary with females dominating males on honor rolls and college admissions.

Discussion

For the rural school system in Northeast Georgia, as well as for systems and states across the nation, standards-based instruction and assessment drives school improvement and reform. The school system's commitment of resources in the form of personnel, software, training, and continued reflection show a deep belief that if we are to improve student learning, we must accurately assess and report progress in terms of mastery of the standards. In order to close the gaps in learning, those gaps must be accurately identified and reported. The system in which this study was conducted prides itself on being proactive in developing and improving standards-based grade reporting, and is a system that leads the way for others in the successful implementation of standards-based grading in grades K-8.

As states across the nation adopt Common Core standards, many more systems may find it necessary to evaluate their grade-reporting systems, and will look to systems that have undergone such a transition from traditional grading to standards-based grading to provide direction. The implementation of Common Core Standards marks an exciting shift in thinking about how students learn, and at the core is the realization that in order to maximize learning, instruction must be targeted to meet student need. Accurate, fair, high quality assessment is critical in this process. This study was designed to determine whether the standards-based grade

reporting system could accurately predict student achievement on high-stakes assessments and whether that system is fair for all students, regardless of gender, socioeconomic status, or limited English proficiency.

Limitations of the Study

There are limitations to this study, but these limitations do not affect the outcome of the study. One limitation is the fact that only students from one system in Northeast Georgia composed the sample population and only one year of data was utilized. Therefore, the results may not be generalized to other grade levels or to other school systems. While the whole-system's fifth grade population was used, a larger population, including students from a more demographically diverse population, would certainly expand the study.

Another limitation is the use of data from only one school year and one grade level. CRCT tests do vary from year to year, as well as by grade, so testing multiple grade levels over multiple years would expand the results. Additionally, there is a varying level of teacher comfort with standards-based grading that provides a limitation to the study. Since the data was collected in the second full year of implementation of standards-based grading, some teachers may develop proficiency with this type of grading system quickly, while others may have difficulty assessing and reporting by standards, allowing for some subjectivity or variance in grading.

Implications for Future Research

The current study is significant because it provides information about the correspondence between standards-based grade reporting, a relatively new approach to reporting student learning, and student success on high-stakes accountability assessments, specifically, the CRCT. The study concluded that standards-based grade reporting can provide accurate information regarding student learning that may serve as a reliable indicator of student success

on high-stakes standardized assessments. This finding is significant because it contradicts the previous research on traditional grading systems showing evidence that traditional grades and state test results are only moderately related (Brennan, Kim, Wenz-Gross, & Sipperstein, 2001; Conley, 2000).

Little research has been done on standards-based grade reporting, as it is a relatively new practice, with many systems still working through implementation. However, this study provides a good indication that the practice of standards based grading warrants further research and continued study as valuable system for reporting on student learning. Additional research should evaluate a larger sample population that includes students of various grade levels (1-5, 6-8), students from assorted school districts and/or states, and students that represent a different demographic background than those found in the rural North Georgia system.

Further Qualitative studies are also recommended to identify the level of subjectivity in grading student work, and to observe the process of implementation of standards-based grading. A case study of a system's implementation would greatly benefit systems considering making the transition to a standards-based grade reporting system. Qualitative research seeks to gain insight and deeper understandings of the how and why of test results. While many studies have been done on the benefits of assessing by standards, few studies show the predictive relationship between standards-based grading and high-stakes tests, and few chronicle the implementation of such a system. The study of standards-based grading is particularly significant as forty-five states prepare to implement the national Common Core standards and federal and state funding levels continue to be tied to assessment achievement levels.

As school systems, administrators, and other educational leaders seek to improve student achievement, the need for quality assessment and accurate reporting systems is greater than ever.

Moreover, as school leaders seek to maximize the positive impact of their resources, quality assessment in terms of the standards drives school and system improvement plans, informing important decisions. The results of this study reveal the clear benefit of reporting on student learning in terms of standards as a way to ensure that all stakeholders are well informed of student progress on the standards and that the grades students earn reflect mastery of the standards rather than a meaningless average that may or may not reflect learning.

Conclusions

The purpose of this quantitative non-experimental causal comparative research study was to determine whether SBRC scores provide accurate information in regard to student academic achievement on high-stakes standardized summative assessments, and whether standards-based grade reporting represents all students similarly, regardless of gender, socioeconomic status, or ELL status. Specifically, the study examined whether grades on standards-based report cards provide accurate information for all students, regardless of gender, ELL status, or socioeconomic status by comparing mean scores on Georgia's CRCT in the areas of math and reading, based on SBRC indicators.

The participants for this study were fifth graders enrolled in eight elementary schools in a rural county in north Georgia from 2009-2010. This dataset contained two sections, one for math grades on SBRC, one for reading grades. The analysis identified the statistically significant differences in mean scores on the CRCT, based on SBRC grades for math grades and reading grades respectively. Ten research questions examined the differences in the mean CRCT scores for all fifth grade students in an entire system in the areas of math and reading, based on SBRC indicators, gender, socioeconomic status, and ELL status. Likewise, there were ten null

hypotheses that attempted to negate a significant difference in means of SBRC data in relation to student achievement on the CRCT based on student demographics.

The study revealed that grades on a SBRC accurately inform student performance on the CRCT for all groups of students, though for both math and reading CRCT scores, it was clear that there are large and significant differences in these scores that depend on one's math or reading SBRC score. Additionally, for Math, even with the same SBRC score, students who are on the free/reduced lunch program tend to score lower than students who are not. For Reading, this result extended even further: even with the same SBRC score, females tend to score higher than males, those with limited English proficiency tend to score lower than those who are proficient, and those who are on the free/reduced lunch program tend to score lower than those who are. This means that, while SBRC scores do not represent all groups similarly when it comes to predicting CRCT scores, there is a significant relationship between SBRC scores and CRCT scores for all students.

Additionally, for almost every group, when CRCT scores are categorized as DNM, M, or E, SBRC scores tend to underestimate these CRCT score categories. This indicates that the grading expectations of the classroom teachers in the system seem to exceed the required levels for proficiency on the state test. The data showed that often, students who do not meet SBRC standards will meet CRCT standards, and students who meet SBRC standards will exceed CRCT standards. This was more significant and noticeable for Math than for Reading; the only exception to this general trend was that for males, Reading SBRC scores tend to be fairly representative of Reading CRCT scores.

In summary, as expected, scores on standards-based report cards do provide accurate information about student learning as shown by success on high-stakes accountability

assessments. While there are differences in the subgroups, showing some variance in the relationships between SBRC and CRCT for students based on gender, socioeconomic status, and limited English proficiency, a significant correspondence between standards-based grades and CRCT scores exists, showing a grading system that provides accurate information about student achievement on a high-stakes state test. While it may be impossible to completely eliminate subjectivity in grading, it is clear that standards-based grading provides an accurate and reliable measure of student learning. By focusing assessment and grade-reporting on the standards, educators continue to tighten the learning cycle, providing a clearer picture of student learning that is necessary to increase student achievement.

REFERENCES

Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometric issues. *Educational Assessment, 8,* 231-257.

Abedi, J. & Gandara, P. Performance of English language learners as a subgroup in large-scale assessment: Interaction of research and policy. *Educational Measurement: Issues and Practices, 25 (4),* 36-46.

Abedi, J. & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14,* 219-234.

Abedi, J. Lord, C. & Hofstetter, C. (1998). *Impact of selected background variables on students' NAEP math performance.* Lost Angeles: UCLA Center for the Study of Evaluation/National Center for Research on Evaluation, Standards, and Student Testing.

American Educational Research Association, American Psychological Association & National Council on Educational Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington D.C.: American Psychological Association.

Ary, D., Jacobs, L.C., Razavieh, A., & Sorensen, C. (2006). *Introduction to research in education,* 7th ed. Belmont, CA:Thomson & Wadsworth.

Atherton, J.S. (2009). Learning and Teaching; Constructivism in learning [On line] UK: Retrieved from: http://www.learningandteaching.info/learning/constructivism.htm

Barton, P.E. (2002). *Staying on course in education reform.* Princeton, NJ: Information

Center, Educational Testing Service.

Bloom, B.S., Madaus, G. F., & Hastings, J.T. (1981). *Evaluation to improve learning.* New York: McGraw-Hill.

Brennan, R.T., Kim, J., Wenz-Gross, M. & Siperstein, G.M. (2001). The relative equitability of high-stakes testing versus teacher-assigned grades: An analysis of the Massachusetts Comprehensive Assessment System (MCAS. *Harvard Educational Review, 71 (2),* 173-216.

Briars, D.J, &Resnick, L.B. (2000). Standards, assessments—and what else? The essential elements of standards-based school improvement. *CSE Technical Report 528.* Los Angeles, CA: University of California.

Brookhart, S.M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice, 10* (1), 35 -36.

Brookhart, S.M. (1993). Teachers'grading practices: Meaning and values. *Journal of Educational Measurement, 30 (*2), 123-142.

Brown, D.F. (1992). Altering curricula through state-mandated testing: Perceptions of teachers and principals. Paper at the annual meeting of the American Education Research Association. April 1992. San Francisco, CA.

Brown, D.F. (1993). The political influence of state-mandated testing reform through the eyes of principals and teachers. ERIC Document Reproduction Service No. ED 360737.

Canady, R.L. & Hotchkiss, P.R. (1989). It's a good score, just a bad grade. *Pi Delta Kappan, 73* (4), 68-71.

Carnoy, M. & Leob, S. (2002). Does external accountability affect student outcomes? A cross state analysis. *Educational Evaluation and Policy Analysis, 24,* 305-331.

Chatterji, M. (2002). Have the tools of inquiry answered pressing questions on

improving schools? *Review of Educational Research, 72 (3),* 345-386.

Cizek, G.J. (2001). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice, 20(4),* 19-27.

Cross, L.H. & Frary, R.B. (1996). *Hodgepodge grading: Endorsed by students and teachers alike.* Paper presented at the annual meeting of the National Council on Measurement in Education, New York.

D'Agostino, J.V., Welsh, M.E., Corson, N.M. (2007). Instructional sensitivity of a state's standards-based assessment. *Educational Assessment, 12(1),* 1-22.

Duran, R. (2006). *State implementaion of NCLB policies and interpretation of the NAEP performance of English language learners (*Paper commissioned by the NAEP Validity Studies (NVS) Panel). Palo Alto, CA: American Institutes for Research.

Donahue, P., Finnegan, R., Lutkus, A., Allen, N., and Campbell, J. (2001). *The nation's report card: Fourth-grade reading 2000* (NCES 2001-499). Washington, DC: U.S. Government Printing Office.

Entwisle, D., Alexander, K., and Olson, L. (1994). *American Sociological Review,* 59(6), 822-838.

Feuerstein, R. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques.* Baltimore: University Park Press.

Frederiksen, J., & White, B. (2004). Addressing the complexity of validity arguments in standards-based assessment. *Measurement 2(4),* 255-261.

Gandal, M. & Vranek, J. (2001). Standards: Here today, here tomorrow. *Educational Leadership, 59(1),* 6-13.

Gandara, P., Rumberger, R., Maxwell-Jolly, J., & Callahan, R. (2003). English learners in California schools: Unequal resources, unequal outcomes. *Education Policy Analysis Archives, 11 (36).* Retrieved from http://epaa.asu.edu/epaa/v11n36

Georgia Department of Education (2008). Retrieved from www.doe.k12.ga.us

Georgia Department of Education (2011). Retrieved from www.doe.k12.ga.us

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. *American Psychologist*. 18, 519-521.

Glaser, R. & Klaus, D.J. (1962). Proficiency measurement: Assessing human performance. In R. Gagne (Ed.), *Psychological principles in systems development.* New York: Holt, Renehart, & Winston.

Guskey, T.R. (1994). Making the grade; What benefits students. Educational Leadership, 52(2), 14-20.

Guskey, T. (2001). Helping standards make the grade. *Educational Leadership,* 59(1), 20-27. Retrieved from http://course1winona.edu/1gray/el626/articlesonline/guskey_he

Guskey, T. R. (2005). A historical perspective on closing achievement gaps. *NASSP Bulletin*, 89(644). 76- 89.

Guskey, T. R. (2007). Multiple sources of evidence: an analysis of stakeholders' perceptions of various indicators of student learning. *Educational Measurement: Issues and Practices*, 19-27.

Guskey, T.R. & Jung, L.A.(2006). The challenges of standards-based grading. *Leadership Compass*, 4(2), 6-10.

Guskey, T.R., Swan, G.M., & Jung, L.A. (2010). Developing a statewide, standards-based student report card: A review of the Kentucky initiative. Paper presented at

the annual meeting of the American Educational Research Association. May, 2010. Denver, CO.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. *American Psychologist*. 18, 519-521.

Hakuta, K., Goto, Y., & Witt, D. (2000). *How long does it take English learners to attain proficienc?* (Policy Rep. No. 200-1). Santa Barbara, CA: The University of California Liguistic Minority Research Institute.

Haney, W. (1993). Testing and minorities. In W. Weis and M. Fine (Eds.), *Beyond silenced voices: Class, race, and gender in United States schools* (45-73). Albany, NY SUNY Press.

Hatch, J.A. (2010). Rethinking the relationship between learning and development: Teaching for learning in early childhood classrooms. *The Educational Forum, 74, 258-268.*

Haertel, E.H, & Lourie, W.A. (2004). Validating standards-based test score interpretations. *Measurement: Interdisciplinary Research and Perspectives, 2,* 61-103.

Hanushek, E.A. & Raymond, M.E. (2004). The effect of school accountability systems on the level and distribution of student achievement. *Journal of the European Economic Association, 2,* 406-445.

Hill, H.C. (2001). Policy is not enough: Language and the interpretation of state standards. *American Educational Research Journal, 38,* 298-318.

Hiliard, A. (2000). Excellence in education versus high stakes standardized testing. *Journal of Teacher Education, 51,* 293-304.

Hood, S. (1998). Culturally responsive performance-based assessment: Conceptual and psychometric considerations, *Journal of Negro Education, 67,* 187-196.

Jencks, C. (1998). Racial bias in testing. In C. Jencks, & M. Phillips (Eds.), *The Black-White Test Score Gap (*55-85). Washington, D.C.: Brookings Institution Press.

Jones, M.G., Jones, B.D., Hardin, B., chapman, L., Yarbrough, T., & Davis, M. (1999). The impact of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan, 81,* 199-203.

Kafer, K. (2005). U.S. girl students outperform boys in most subjects, study finds. *School Reform News.*

Kifer, E. (2001). *Large-scale assessment: Dimentions, dilemmas, and policies.* Thousand Oaks, CA: Corwin Press.

Kohn, A. (2000). Burnt at the stakes. *Journal of Teacher Education, 51,* 315-327.

Kohn, A. (2000). *The case against standardized testing.* Portsmouth, NH: Heinemann.

Lay, J.C, Stokes-Brown, A.K. (2009). Put to the test: Understanding differences in support for high-stakes testing. *American Politics Research, 37(3),* 429-448. Retrieved from: http://apr.sagepub.com/content/37/3/429.refs.html

Lehrer, J. (2009). *How we decide*. New York: Houghton Mifflin.

Madaus, G. (1995). Technological and historic consideration of equity issues associated proposals to change our nation's testing policy. In M.T. Nettles & A.L. Nettles (Eds.). *Equity and excellence in educational testing and assessment* (23-68). Boston: Kluwer.

Madaus, G., & Clarke, M. (2001). The impact of high-stakes testing on minority students. In M. Kornhaber & G. Orfield (EDS.) *Raising standards or raising barriers: Inequality and high stakes testing in public education (*85-106). New York: Century Foundation.

Mager, R.F. (1962). *Preparing instructional objectives*. Palo Alto, CA: Reardon Publishers.

Monsaas, J. A., & Engelhard, G. J. (1991). Attitudes toward testing practices as cheating and teacher testing practices. Paper presented at the Annual Meeting of theAmerican Educational Research Association, Chicago, IL.

Marzano, R.J. (2000). Transforming classroom grading. Alexandria: VA: ASCD.

Marzano, R.J. (2010) Formative assessment &standards-based grading. Bloomington, IN: Marzano Research Laboratory.

McTighe, J &O'Connor, K. (2005). Seven practices for effective learning. *Annual Editions: Educational Psychology,* 209-213.

National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010

Newman, K. S. & Margaret, M.C. (2003). High stakes: Time poverty, testing, and the children of the working poor. *Qualitative Sociology, 26(1)* 3-34.

Nitko, A.J. (1980). Distinguishing the many varieties of criterion-referenced tests. *Review of Education Research, 50(3),* 461-485. Retrieved from://www.jstor.org/stable/1170441

Peterson, P and Xabel Lastra-Anadon, C. (2010). State standards rise in reading fall in math, *Education Next. 10(4), 12-16.*

Phelps, R.P. (Ed.). (2005). *Defending standardized testing.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Popham, W.J., & Husek, T.R. (1969). Implications of criterion-referenced measurement. *Journal of Educational Measurement, 6,* 1-9.

Popham, W.J., Cruse, K.L., Rankin, S.C., Sandifer, P.D., & Williams, R.L. (1985). Measurement-driven instruction: It's on the road. Phi Delta Kappan, 66, 628-

634.

Opfer, D, Henry, G., & Mashburn, A. (2008). The district effect: systemic responses to high stakes accountability policies in six southern states. *American Journal of Education, 299-332.*

Reeves, D.B. (2011). *Elements of grading: a guide to effective practice*. Bloomington, IN: Solution Tree Press.

Resnick, L.B. (2006). Making accountability really count. *Educational Measurement, 25 (1)* 33-37.

Ridley, M. (2003). *Nature via nurture: Genes, experience, and what makes us human.* New York: HarperCollins.

Romberg, T.A. Zarinna, E.A., Williams, S.R. (1989). The influence of mandated testing on mathematics instruction: Grade 8 teachers' perceptions. Madison, WI: Univ. Of Wisconsisn, Center for Educational Research.

Scriffiny, P. (2008). Seven reasons for standards-based grading. *Educational Leadership, 66(2),*70-74.

Shanker, A. (1995). The case for high stakes and real consequences. In D. Ravitch (Ed.), *Debating the future of American education: Do we need national standards and assessment?* (pp. 145-153). Washing, DC: Brookings Institute.

Sherer, M. (2001). How and why standards can improve student achievement. *Educational Leadership, 59(1),* 14-19.

Stiggins, R. (2001). Assessment through the student's eyes. *Annual Editions: Educational Psychology*, 199-203.

Stiggins. R.J., Frisbie, D.A., & Griwsold, P.A. (1989). Inside high school grading

practices: Building a research agenda. *Educational Measurement: Issues and Practice, 8* (2), 5-14.

Sunderman, G.L. & Kim, J. (2004). *Inspiring vision, disappointing results: Four studies on implementing the No Child Left Behind Act.* Cambridge, MA: Harvard University Press, The Civil Rights Project.

Thernstrom, A., & Thernstrom, S. (2003). *No excuse: Closing the racial gap in learning.* New York: Simon & Schuster.

Tognolini, J., Stanley, G. (2007). Standards-based assessment: A tool and means to the development of human capital and capacity building in education. *Australian Journal of Education, 51(2),* 129-145.

U.S. Census Bureau, population division (2008). Retrieved from: www.google.com/publicdata

U.S. Department of Education (2010). Retrieved from: http://www2.ed.gov/programs/racetothetop-assessment/index.html

Valenzuela, A. (2000). The significance of the TAAS test for Mexican immigrant and Mexican American adolescents: A case study. *Hispanic Journal of Behavioral Sciences, 22*, 524-539.

Waber, D., Gerber, E. Turcios, V. Wagner, E., and Forbes, P. (2006). Executive functions and performance on high-stakes testing in children from urban schools. *Developmental Neuropsychology, 29(3),* 459-477.

Wang, L., Beckett, G., and Brown, L. (2006). Controversies of standardized assessment in school accountability reform: A critical synthesis of multidisciplinary research evidence. *Applied Measurement in Education, 19 (4),* 305-328.

Welsh, M.E., & D'Agostino, J.V. (2009). Fostering consistency between standards-based grades and large-scale assessment results. In T.R. Guskey (Ed.), *Practical solutions for serious problems ins tandards-based grading (pp. 75-104). Thousand Oaks, CA: Corwin Press.*

Wiggins, G. (1996). Honesty and fairness: toward better grading and reporting. In. T. R. Guskey (ed.). *Communicating student learning: the association for supervision and curriculum development yearbook,* Alexandria, VA: Association for Supervision and Curriculum Development, 141-176.

Wolf, M.K. Farnsworth, T., and Herman, J. Validity issues in assessing English language learners' language proficiency. *Educational Assessment, 13,* 80-107.

Yeh, S. S. (2005, October). Limiting the unintended consequences of high stakes testing. Education Policy Analysis Archives.

Young, J.W., Yeonsuk, C. Guangming, L, Cline, F. Steinberg, J., Stone, E. (2008). Validity and fairness of state standards-based assessments for English language learners. *Educational Assessment, 13(2/3),* 170-192.

Zane, T.W. (2009). Performance assessment design principles gleaned from constructivist learning theory. *Tech Trends, 53(1),* 81-88.

Zehr, M. (2008). Hurdles Remain High for English-Learners. (Cover story). Education Week, 27(39), 1-11. Retrieved from EBSCOhost.

Zill, N. (2006). Family change and student achievement: What we have learned, what it means for school. In A. Booth & J.P. Dunn (Eds.) *Family-school links: How do they affect educational outcomes? (*pp. 139-174). Hillsdale, NJ: Lawrence Erlabaum Axociates.

APPENDIX A

Revised 8/26/09

Jackson County School District
Grade 05 Report Card
2009-2010
Page 1 of 2

Student:
Student ID:
School:
Teacher:

Academic Performance Level for Learning Standards				
NA	DNM	IP	M	E
Not Applicable	Does Not Meet	In Progress	Meets the Standard	Exceeds the Standard

5TH GRADE	Term 1	Term 2	Term 3	Term 4
I. READING				
FLUENCY, VOCABULARY, AND COMPREHENSION				
Read orally with speed, accuracy, and expression on 5th grade text				
Read and comprehend texts from a variety of genres on grade level				
Acquire and use grade level words to communicate effectively				
Read independently from different genres and sources				
II. LANGUAGE ARTS				
LISTENING, SPEAKING AND VIEWING				
Use oral, visual, and auditory, strategies to communicate				
WRITING AND CONVENTIONS				
Demonstrate competency using the writing process in a variety of genres				
Develop IDEAS for writing				
Produce ORGANIZED writing				
Use elements of STYLE				
Use appropriate CONVENTIONS in writing				
Identify correct elements of grammar				
III. MATHEMATICS				
ALGEBRA				
Represent and interpret relationships algebraically using variables				
DATA ANALYSIS				
Gather, organize, interpret & display data using the most appropriate graph				
GEOMETRY				
Understand congruence of geometric figures				
MEASUREMENT				
Understand and compute area				
Understand and compute volume				
Understand PI and determine circumference				
Understand and measure capacity				
NUMBER AND OPERATIONS				
Understand place value of numbers				
Understand and apply characteristics of whole numbers				
Understand multiplication of decimal fractions				
Understand division of decimal fractions				
Understand the meaning of common fractions				

5TH GRADE	Term 1	Term 2	Term 3	Term 4
Add & subtract common fractions & mixed numbers				
Model multiplication and division of common fractions				
Understand the meaning of percent				
IV. SCIENCE				
CHARACTERISTICS OF SCIENCE				
Apply the processes of scientific inquiry				
EARTH SCIENCE				
Identify surface features caused by constructive & destructive processes				
Relate how technology & human intervention can control Earth's processes				
LIFE SCIENCE				
Classify organisms into groups (plants & animals)				
Realize offspring resemble parents in inherited traits & learned behaviors				
Diagram, label and explain the function of each cell part				
Relate how microorganisms benefit or harm larger organisms				
PHYSICAL SCIENCE				
Explain how an object is the sum of its parts (mass and magnification)				
Explain the difference between physical and chemical changes				
Understand electricity, magnetism, & their relationship				
V. SOCIAL STUDIES				
ECONOMIC UNDERSTANDINGS				
Use basic economic concepts				
GEOGRAPHIC UNDERSTANDINGS				
Locate important places in the United States				
GOVERNMENTICIVIC UNDERSTANDINGS				
Explain how a citizen's rights are protected under the US Constitution				
Explain the purpose and the process for amendments to the US Constitution				
HISTORICAL UNDERSTANDINGS				
Explain the causes, major events, and consequences of the Civil War				
Analyze the effects of Reconstruction on American life				

5TH GRADE				
	Term			
	Term 1	Term 2	Term 3	Term 4
Describe how life changed in America at the turn of the century				
Describe US involvement in WWI and post WWI America				
Expl. how Great Depression & New Deal affected lives of mill. of Americans				
Explain the reasons for America's involvement in WWII				
Discuss the origins and consequences in the Cold War				
Describe importance of key people, events & developments b/w 1950 & 1975				
Trace important developments in America since 1975				

5TH GRADE SUMMARY REPORT				
	Term			
	Term 1	Term 2	Term 3	Term 4
*SEE ACADEMIC LEARNING STANDARDS RUBRIC				
Reading Summary Mark				
Language Arts Summary Mark				
Mathematics Summary Mark				
Science Summary Mark				
Social Studies Summary Mark				
S=SATISFACTORY N=NEED IMPROVEMENT U=UNSATISFACTORY				
Music				
Art				
P.E.				
Health				

GENERAL LEARNER OUTCOMES				
	Term			
	Term 1	Term 2	Term 3	Term 4
1=RARELY 2=OCCASIONALLY 3=FREQUENTLY 4=CONSISTENTLY				
GLO 1: SELF-DIRECTED LEARNER				
Demonstrate motivation and responsibility for learning				
GLO 2: TEAM PLAYER				
Work well with others				
GLO 3: QUALITY PRODUCER				
Recognize and produce quality products and performances				
Teacher Comments:				

STUDENT PROGRAM INFORMATION				
	Term			
	Term 1	Term 2	Term 3	Term 4
See attachment				

ATTENDANCE				
Term	Term 1	Term 2	Term 3	Term 4
Days Present	14	0	0	0
Days Absent	2.0	0	0	0
Periods Tardy	0	0	0	0

***Comments:**
Explanation of Academic Performance Level:

Exceeds= Student applies the skill or concept independently and correctly; shows higher level thinking.

Meets= Student demonstrates understanding by showing clear thought processes through demands of task independently; has met all indicators for meeting the standard during the current term.

In Progress= Student demonstrates some understanding but can not complete tasks independently. Student is making sufficient progress to meet the standard by the end of the year.

Does Not Meet= Student is not making sufficient progress to meet the standard by the end of the year.

Not Applicable= Standard is not yet addressed in the environment.