

THE IMPACT OF STANDARDS-BASED REPORT CARDS ON READING

ACHIEVEMENT IN THE THIRD GRADE

By

Terri L Wyrosdick

A Dissertation
Submitted to the Faculty of
Columbus State University
In Partial Fulfillment of the Requirements
For the degree of Doctor of Education
In Curriculum and Leadership

Columbus State University
Columbus, GA

April 2014

Copyright by
Terri L Wyrosdick
2014

THE IMPACT OF STANDARDS-BASED REPORT CARDS ON READING
ACHIEVEMENT IN THE THIRD GRADE

By

Terri L Wyrosdick

Approved:

Thomas J. McCormack
Professor of Educational Leadership
(Committee Chair)

Larry J. Beatty
Professor of Educational Leadership
(Committee Member)

Jennifer L. Brown
Assistant Professor of Educational Foundations
(Methodologist)

Name: Terri L Wyrosdick

Previous degrees: B.S. Middle Grades Education, 1988, Columbus College; M.S. Educational Leadership, 2008, Troy State University; Ed.S. Curriculum and Leadership, 2013, Columbus State University

Date of Current Degree: April 2014

Institution: Columbus State University

Major Field: Educational Leadership

Major Professor: Dr. Thomas McCormack

Title of Study: THE IMPACT OF STANDARDS-BASED REPORT CARDS ON
READING ACHIEVEMENT IN THE THIRD GRADE

Pages in Study: 83

Candidate for Degree of Doctor of Education

School districts across the nation are transitioning away from traditional A-F letter grade report cards in favor of standards-based report cards (SBRC). Previous studies have indicated that many parents were confused by SBRC. The purpose of this study was to determine if a difference exists between the reading achievement of third-grade students using traditional A-F letter grade report cards and those students using SBRC. Pre-existing CRCT data of the pass/fail percentage of third graders from five school districts and 118 schools in 2009 and 2010, the year prior to and the year of implementation of SBRC, were analyzed. A chi square test indicated that no statistically significant difference existed between report card type and student reading achievement among third grade students. Districts may want to reconsider the time and expense involved in adopting a report card that so many parents find difficult to understand.

DEDICATION

This paper is dedicated to my late husband, John Wyrosdick, Jr. We started this journey together. I wish you were here to celebrate the conclusion. It is also dedicated to our son, John Wyrosdick, III, and most importantly, to my Lord and Savior, Jesus Christ.

ACKNOWLEDGEMENTS

I would like to thank Dr. Thomas McCormack for guiding me through this process and agreeing to chair my committee. He has encouraged me throughout the duration of the doctoral program. I would also like to thank Dr. Larry Beatty, who has generously shared his wealth of knowledge with not only myself, but the many students he has guided over the years. I will forever be grateful for his support during a time of tragedy. To Dr. Jennifer Brown, I am greatly appreciative for all the assistance with the statistics, the APA formatting, and the numerous drafts she helped me revise. I would also like to thank Dr. Iris Saltiel, who initiated me in the dissertation process and has stood by to cheer me on through the end. Additionally, I would like to thank my family and friends for patiently enduring all the missed lunches, family gatherings, and other social occasions from which I have been absent over these past few years.

TABLE OF CONTENTS

DEDICATION	iv
ACKNOWLEDGEMENTS	v
LIST OF TABLES	viii
CHAPTER	
I. INTRODUCTION	2
Introduction.....	2
Statement of the Problem.....	6
Purpose.....	7
Research Question	8
Definitions.....	8
Assumptions of the Study	10
Significance of the Study	10
Limitations of the Study.....	11
Summary	11
II. REVIEW OF RESEARCH AND RELATED LITERATURE.....	12
Introduction.....	12
History of Grading	12
Standards-Based Movement	16
Problems with Grading	17
Purposes of Grading.....	20
Types of Grading	20
Norm-Referenced.....	20
Criterion-Referenced.....	21
Letter Grades.....	22
Percentage Grades.....	23
Standards-Based Grading.....	24
Motivational Effect of Grades	28
Positive Influence of Grades	28
Intrinsic versus Extrinsic Motivation.....	30
Cameron versus Deci	30
Other researchers	32
Negative Influence of Grades	34
Future of Grading.....	35
Summary of the Literature	35
III. METHODOLOGY	43

Introduction.....	43
Research Design.....	44
Population and Sampling.....	44
Instrumentation.....	46
Procedures.....	49
Data Analysis.....	50
Limitations.....	51
Assumptions.....	51
Data Interpretation.....	52
Implications.....	52
Summary and Expectations.....	53
 IV. RESULTS.....	 54
Introduction.....	54
Descriptive Data.....	55
Data Analysis.....	57
Results.....	57
Summary.....	60
 V. SUMMARY.....	 61
Summary.....	61
Interpretations.....	62
Conclusions.....	63
Practical Implications.....	64
Limitations.....	66
Recommendations for Future Study.....	67
 REFERENCES.....	 69
 APPENDIX	
A. COBB COUNTY THIRD GRADE REPORT CARD.....	80
B. MUSCOGEE COUNTY THIRD GRADE REPORT CARD.....	82

LIST OF TABLES

1	Concept Analysis for Significant Studies Related to the Problem of Grading	19
2	Norm Referenced and Criterion Referenced Report Card Legends	23
3	Sample Report Card Legend for Percentage Grades	23
4	Concept Analysis for Studies Related to Standards-Based Grading	26
5	Report Card Legends for Standards-Based Report Cards	27
6	Concept Analysis for Studies Related to Standards-Based Report Cards	29
7	Concept Analysis for Studies Related to Motivation.....	38
8	Enrollment Data of Georgia School Districts that Adopted Standards-Based Report Cards in 2010	47
9	Demographic Data of Georgia School Districts that Adopted Standards-Based Report Cards in 2010	47
10	Number of Schools Included in the Study.....	55
11	Subgroup Populations Examined in the Study	56
12	Subgroup Populations Disaggregated by County and Year.....	56

CHAPTER I

INTRODUCTION

Introduction

Society has long rewarded students for report card grades, from parents, grandparents, and other family members paying cash for A's and B's to establishments with video games passing out tokens for every A. Generations of students have answered the question, "Wad-ja-Get?" (Kirschenbaum, Simon, & Napier, 1971, p. 15). A current trend in education, standards-based-report cards, would eliminate this decades-old practice (Cherniss, 2008).

A standards-based report card typically contains a list of the state's or local school district's learning standards for a specified grade level, and gives information about students' achievement of those standards. Achievement is measured in relation to the standard as opposed to averaging grades or normative student comparisons (Bostic, 2012). Each standard is evaluated independently. Some report cards use numeric performance levels which correspond to a specified achievement level. The most commonly used set of descriptors matches performance levels of 1, 2, 3, and 4 with achievement labels *Beginning*, *Progressing*, *Proficient*, and *Exceptional* or with the behavioral labels *Seldom*, *Sometimes*, *Usually*, and *Consistently/Independently* (Guskey & Bailey, 2001). Other types of cards simply have spaces for marks to indicate the category most suitable for the student's skills, such as emerging, proficient, basic; or does not meet, meets exceeds; no letter grades. A standards-based report card provides more

detailed information about a student's achievement (Bostic, 2012). Does more detailed information translate into better information, though?

Assessment experts Guskey and Bailey (2001) have identified the following six major purposes of grading, but acknowledge that educators seldom agree on which purpose is most important:

1. Communicating student achievement to parents and others.
2. Providing information students can use for self-evaluation.
3. Selecting, identifying, or grouping students for certain educational plans or programs.
4. Providing student learning incentives.
5. Evaluating program effectiveness.
6. Providing evidence of student's lack of effort or irresponsibility.

Assessment expert Airasian (1994) asserts that many agree that the general purpose of a report card is to communicate information about a pupil's academic achievement, but within that general purpose he identifies four more specific purposes: administrative, informational, motivational, and guidance. Indeed, researchers have reported extensively on the multi-various purposes for grades and report cards (Munk & Bursuck, 2001; Wrinkle, 1947), additionally including the purposes of instructional planning (Marzano, 2000), sorting students (Resh, 2009), and communicating student behavior (Carlson, 2003; Jung & Guskey, 2010). In citing the purposes of grades and report cards, though, many researchers agree that grades provide motivation or incentive to learn (Airsasin, 1994; Guskey & Bailey, 2001), factor significantly in determining student effort (Cameron & Pierce, 1994), and tend to support student motivation and

success (Malone, Nelson, & Van Nelson, 2002). Could the transition away from letter grades and traditional A-F report cards diminish the motivational factor of grades and impact the academic achievement of students?

Though calls for reform in grading began over a century ago (Kirschenbaum et al., 1971), the current call for reform through standards-based report cards follows the call for standards-based curriculums, which many consider to have begun in 1983 with the U. S. Department of Education's publication *A Nation at Risk: The Imperative for Education Reform* (Cherniss, 2008; Paepflow, 2011). Recommendations in that report included more rigorous and measurable standards, higher expectations for academic achievement and student conduct, and grades that are accurate indicators of academic achievement and reliable for determining readiness for further study (The National Commission on Excellence in Education, 1983). The call intensified a decade later with the 1994 adoption of Goals 2000: Educate America Act and again in 2001 with the passage of the No Child Left Behind Act. States subsequently responded by developing content standards for every grade level and for every subject (Marzano, 1998). Common Core Standards, established in 2009 and currently adopted by 45 states, reflect a national alignment of standards-based education reform from kindergarten through high school (Rogers, 2013). Once those standards and assessments were in place, educators then faced the daunting challenge of determining best practices for grading and reporting student learning according to those standards (Guskey, 2001).

The changes in curriculum were not the only catalysts for changes in report card grading. Dating back to the early 1900s, researchers have reported on the inconsistencies

in grading and what grades actually mean (Starch & Elliott, 1912, 1913a, 1913b).

Whipple (1913) wrote,

When we consider the practically universal use in all educational institutions of a system of marks, whether numbers or letters, to indicate scholastic attainment of the pupils or students in these institutions, and when we remember how very great stress is laid by teachers and pupils alike upon these marks as real measures or indicator of attainment, we can but be astonished at the blind faith that has been felt in the reliability of the marking system. School administrators have been using with confidence an absolutely uncalibrated instrument...What faults appear in the marking systems that we are now using, and how can these be avoided or minimized? (p. 1)

In 1933, Middleton (1933) described the difficulties of chairing a committee tasked with revising his school's grading and reporting system:

The Committee on Grading was called upon to study grading procedures. At first, the task of investigating the literature seemed to be a rather hopeless one. What a mess it all was! Could order be brought out of such chaos? Could points of agreement among American educators concerning the perplexing grading problem actually be discovered? It was with considerable misgiving and trepidation that the work was finally begun (p. 5).

More recently, Marzano (2000) expressed his concern that grades were so imprecise that they were virtually meaningless. His views are echoed by many who express concerns about averaging percentage score grades, contending that averaging grades falsifies grade reports (Marzano, 2000; O'Connor, 2009, 2010; Reeves, 2010; Wormeli, 2006), that averaging grades fails to report student mastery at the end of the learning process (O'Connor & Wormeli, 2011), and that factoring in zeroes makes obtaining a passing grade almost impossible (O'Connor & Wormeli, 2011; Reeves, 2004). Moreover, the concerns extend beyond percentage score averaging into how grading is done. Grading practices lack uniformity across states, districts, and even within schools, resulting in vast

variations in student assessment from teacher to teacher (Carifo & Carey, 2009). Additionally, many teachers factor in a number of non-achievement measures, such as effort, ability, and improvement (Brookhart, 1991; Cross & Frary, 1999; Pilcher, 1994; Stiggins, Frisbie, & Griswold, 1989). O'Connor and Wormelli (2011) contend that any instructional decision based upon such fabricated grade reports are unreliable, as they offer imprecise documentation and are useless for descriptive feedback. Some educators have called for an end to grading altogether (Kohn, 2011).

Statement of the Problem

Standards-based report cards have replaced traditional report cards in many districts across the country. Standards-based report cards focus on the individual skills that students are expected to master and provide information about those skills through either a narrative or with number or symbols (Manzo, 2001; O'Connor, 2010). Many researchers argue that standards-based report cards are a more accurate and more objective measure of student knowledge than traditional A – F grades based upon percentage (Guskey, 2001; Marzano & Kendall, 1998; O'Connor, 2010) and are the next logical step in aligning state standards to student achievement (Cherniss, 2008). Given the motivational factor of grades (Malone et al., 2002), however, does the more detailed information provided on standards-based report cards actually translate into improved academic outcomes for students, or does it possibly do more harm than good for certain student populations?

Purpose

The researcher's purpose of this study was to determine if a difference exists between the reading achievement of third-grade students using traditional A-F letter grade report cards and those students using standards-based report cards. Though researchers and educators question the validity of the traditional grading system, the rewards-based nature of the traditional system has long been ingrained in American society, and research indicates that the use of grades encourages student motivation and success (Malone et al., 2002). Teachers have struggled for years with issues of student motivation. If A to F grades become obsolete, will student motivation, and ultimately student achievement, be affected?

When one school district switched to standards-based report cards, teachers met with parents to explain fully the rationale of standards-based report cards and how students would be assessed. In addition to the scoring rubric of one to four for each standard, one overall grade was given for each subject, based solely upon summative tests given for each standard. One parent asked why her child should do homework if it was not going to count towards that one grade. What slowly occurred over the course of the year was that students would complete homework to keep from suffering some type of consequence for not having it, such as completing it at the silent lunch table, but many students did not care if the answer was right or wrong. For every assignment given, a student would ask, "Does this count towards our summative score?" When grades were removed from the equation, some parents and students saw little point in assignments. The rubric score for each standard meant little to either the parents or the students. Teachers felt their hands were tied. Few students were intrinsically motivated to learn,

and the most powerful extrinsic motivation was gone. Teachers became very concerned about the academic success of the students, especially those students for whom their parents had the least understanding of the report card. The researcher explored if a relationship exists between the loss of letter grades on student-report cards and changes in the reading achievement of elementary students, as measured by the Georgia Criterion-Referenced Competency Test.

Research Question

The overall guiding question for this study was, “Is there a difference between reading achievement of third-grade students using traditional A-F letter grade report cards and those students using standards-based report cards?”

The following hypothesis guided this study:

H₁: A difference exists between the reading achievement of third-grade students using traditional A-F letter grade report cards and those students using standards-based report cards.

The null hypothesis is:

H₀: A difference does not exist between the reading achievement of third grade students using traditional A-F letter grade report cards and those students using standards-based report cards.

Definitions

Academic Achievement: student academic growth as evidenced by some qualitative or quantitative measure of learning (Bradbury-Bailey, 2011).

Extrinsic Motivation: behavior that is motivated by some external reward, such as grades, praise, fame, or money and that arises from outside a person as opposed to originating from inside the person (Cherry, n.d.).

Grades: a summary statement of student evaluations for a specified time period, as reported by numbers or letters (Marzano, 2000).

Grading: a teacher's professional judgment of student achievement, based on the evaluation and collection of student achievement and performance evidence (Guskey, 2002).

Intrinsic Motivation: motivation that originates from inside a person rather than from an outside reward, such as grades or money, and is derived from the pleasure obtained from the task itself (Bainbridge, 2013).

Measurement: the assignment of marks as determined by explicitly set rules (Marzano, 2000).

Motivation: an internal condition, state, want, or desire that drives and directs goal-oriented behavior; the influence of one's needs and desires on behavioral direction and intensity (Huitt, 2001).

Reporting: the process by which teachers' judgments of student evaluation, as indicated by grades or marks of designated performance levels, are communicated to students, parents, or others (Guskey, 2002).

Standards-Based Grading: measuring students' proficiency on well-defined course objectives (Tomlinson & McTighe, 2006), based on the principle that grades are about what students have learned, not what they have earned, and should be accurate indicators of student achievement of standards (Brookhart, 2011).

Standards-Based Report Card: An alternate method of reporting student progress which involves assessing student proficiency on state and local standards and benchmarks (Craig, 2011), utilizing a rubric score or some other descriptive measure for each individual standard.

Traditional A-F Report Card: A report of student progress, provided at set intervals throughout the school year, which assigns a letter grade of A to F to indicate student performance in a given course of study.

Assumptions of the Study

The researcher in this study made the assumptions that teachers graded students without bias for gender, race, religion, or socio-economic status, and that students taking the Georgia Criterion Referenced Competency Test provided as much effort as possible to accurately demonstrate their level of knowledge.

Significance of the Study

School districts across the county are transitioning to standards-based grading and standards-based report cards (Marzano & Heflebower, 2011); however, research on their implementation and effectiveness is limited (Cherniss, 2008). Researchers and educators tout them as being less biased and subjective while being more valid and reliable and as providing more accurate information (Marzano. 2000). The researcher found little research, though, to determine what, if any, impact this transition to standards-based report cards has had on the reading achievement of elementary students, as determined by standardized test scores. A search of ProQuest databases in January, 2014, using the key words standards-based report cards, yielded only one study that examined the relationship

between student achievement and standards-based report cards. This study contributes to the limited amount of existing research on the academic impact of implementing standards-based report cards.

Limitations of the Study

The findings of this study are limited to the populations studied and not generalized to other populations. The study focused on Georgia school districts that have transitioned to standards-based report cards at the elementary school level, prior to the implementation of the Common Core Georgia Performance Standards, which ushered in a new curriculum. Relationships between the transition to standards-based report cards and any changes in student achievement would not be deemed as a causal relationship, as school districts use many varied strategies for improving student achievement.

Summary

Increasing numbers of school districts have transitioned towards standards-based report cards and away from traditional A-F letter grade report cards. Many standards-based report cards use a rubric score to represent either an achievement or behavior level. The trend towards standards-based reporting is a response to the trend towards a standards-based curriculum which began in the 1980s, as well as a response to questions of validity and reliability in common grading practices. Many educators and researchers agree that one purpose of grades is to motivate students. Could the absence of the potential motivating effect of grades impact student achievement?

CHAPTER II

REVIEW OF RESEARCH AND RELATED LITERATURE

Introduction

From the time grades became prevalent in the American education system, controversy has surrounded their use (Cross & Frary, 1999). Researchers report numerous problems with today's grading systems, including lack of reliability (Guskey, 2001), lack of validity (Brookhart, 1991), inconsistency amongst teachers (Guskey, 2001), and inclusion of non-academic factors (Cross & Frary, 1999). This review of literature highlights the history of American grading practices, beginning with the first known use of grades in a public school, and then discusses perceived problems with modern grading systems. Various methods of grading and their various shortcomings are summarized, including what current research says about standards-based reporting. Lastly, the issue of grades as a motivational factor in student achievement, a highly contentious topic, is explored.

History of Grading

The practice of assigning grades began at the college level, and archival evidence indicates that the first American educational institution to issue grades was Yale College in 1785 (Tocci, 2008). Prior to that time, students received verbal or narrative feedback (Marzano, 2000). In 1813 Yale modified its grading scale to a 1-4 numeric scale, with one corresponding to optima (Tocci, 2008). Other universities began to follow suit, and this four-point scale was the origin of the 4.0 system used by today's colleges and universities (Durm, 1993). In 1830, Harvard implemented a 20-point scale, and then in

1877 switched to a 100 point scale in which students were classified into divisions according to where they fell on the scale (Marzano, 2000). Most universities began moving to a 1-5 scale (Curreton, 1971), and in 1897, Mount Holyoke College initiated an A to E letter grade system (Marzano, 2000).

The Boston school system of 1845 has the first recorded use of grades in a public school in the United States. A “proto-standardized exam was given to students across the city and straight percentages of right and wrong were computed” (Tocci, 2008, p. 765). No known grading and reporting practices existed in public schools prior to this time (Guskey, 1994; Tocci, 2008). Instead, teachers gave oral reports of student progress to parents, usually during a visit to the home, and students of all ages and backgrounds were grouped together with one teacher. Few of these students were educated beyond the elementary level (Guskey & Bailey, 2001). When McGuffey readers became popular, many schools used them to classify children according to the grade number of the book from which they could read (Morris, 1952). After compulsory elementary attendance laws were passed in the late 1800’s, the number of students attending high school dramatically increased, and the number of students attending public high schools went from 110,000 in 1880 to over two million in 1920 (Guttek, 1986). This rapid expansion of the public school system in the early 1900’s initiated a myriad of grading practices. Schools began grouping students in grades according to their age and issuing formal progress evaluations in which teachers would write down the skills each student had mastered and which ones were yet to be mastered, prior to moving on to the next grade (Edwards & Richey, 1947). Grades became a matter of managerial efficiency for a growing student population (Tocci, 2008). During the early 1900’s, elementary schools

experimented with written descriptions and narrative reports (Guskey & Bailey, 2001), S and U for Satisfactory and Unsatisfactory (Tocci, 2008), and a 'passed', 'conditioned', and 'not passed' scale (Curreton, 1971), while percentage grades became customary in the high schools, aligning an A-F scale with the 1-100 scale (Kirschenbaun et al., 1971).

In 1912, a study by Starch and Elliott sparked debate about the reliability of percentage grades. In the study, two papers, written for a first-year high school English class, were given to 142 teachers for grading. On one paper, 15 percent of the teachers gave it a failing grade, while 12 percent scored it at 90 or higher. Grades on the other paper ranged from 50 to 97 (Starch & Elliott, 1912). Critics of the research contended that the large variance in scores was a natural result of the subjectivity involved in grading language work; therefore, Starch and Elliott conducted a follow-up study a year later, repeating the process with geometry papers. These studies received an even greater variance in scores (Starch & Elliott, 1913a). Yet another follow-up study conducted in the same manner with history papers yielded similar results of wide variance in teacher scoring (Starch & Elliott, 1913b). As a result of these studies, some educators were briefly prompted to eliminate percentage grades and return to grading scales which had fewer and larger categories, such as Excellent, Average, and Poor (Guskey, 1994). In 1918, categorical grading scales were replaced with the the letters A, B, C, D, and F (Chapman & Ashbaugh, 1925).

Based upon his ground-breaking research indicating the wide variance with which teachers scored student work, Starch (1913) proposed that distribution of grades of large groups of students should follow the probability curve, in which 3% of the students should receive an A+ (97-100), 7% should receive an A- (93-96), 16% should receive a

B+ (89-92), 23% should receive a B- (85-88), 23% should receive a C+ (81-84), 16% should receive a C- (77-80), 5% should receive a D+ (73-76), 3% should receive a D- (70-72), and 4% should fail. Though disagreement existed over the exact percentages in the distribution and the exact shape of the curve (Starch, 1913), the idea of grading on the curve emerged, and the University of Missouri became the first to initiate this grading method (Tocci, 2008). Grading on the curve became increasingly popular in the 1930s as educators sought to minimize subjectivity in grading (Guskey, 1994). Strong opposition to use of the normal curve for grade distribution quickly developed (Davis, 1931), and the debate over grades continued, leading some schools to forego grades altogether and return to verbal descriptors, pass fail systems, or mastery approaches (Guskey, 1994).

The idea of including narrative comments along with letter grades gained support after research by Page (1958) indicated that students achieved higher scores on classroom tests when grades were accompanied by positive teacher comments. His study included 74 secondary school teachers who administered a test to their students and scored it as they normally would. A letter grade of A, B, C, D, or F was assigned to each test in correspondence to the numeric score given by the teacher. The teachers then randomly divided each set of papers into three groups. In the first group, students received only the numeric score and letter grade. In the second group, students received the following standard comments, in addition the numeric score and letter grade:

- A: Excellent! Keep it up.
- B: Good work. Keep at it.
- C: Perhaps try to do still better?
- D: Let's bring this up.

F: Let's raise this grade!

In the third group, students also received the numeric score and letter grade, but the teacher made individualized comments on each, having been instructed to write whatever comments conformed to their own feelings and practices. The students who received the standard comments scored significantly higher on their next assessment in that class than those students who had no comments. The students who received individualized comments achieved even higher scores.

Throughout the ongoing controversies surrounding grades, letter grades became the most prominent means of reporting student achievement (Guskey, 2002); however, still not satisfied with the “hodgepodge” (Brookhart, 1991, p. 36) of grading and marking systems, many school reform efforts in the United States today have included modifying report cards to more effectively communicate student learning (Lake & Kafka, 1996), and new generations of educational researchers have called for yet another means of reporting student achievement – the standards-based-report card (Guskey & Bailey, 2001; Marzano, 2000).

Standards-Based Movement

Marzano and Kendall (1998) trace the beginnings of the standards based movement to the 1983 report, *A Nation at Risk*, and detail how it dramatically changed the rhetoric of educational reform, eventually leading to an education summit in 1987 with then President George Bush and the nation's governors. That summit led to the publication of *The National Education Goals Report: Building a Nation of Learners*

(National Education Goals Panel, 1991), which included six broad goals for American education, two of which were specifically related to academic standards:

Goal 3: By the year 2000, American students will leave grades four, eight, and twelve having demonstrated competency in challenging subject matter, including English, mathematics, science, history, and geography; and every school in America will ensure that all students learn to use their minds well, so they may be prepared for responsible citizenship, further learning, and productive employment in our modern economy.

Goal 4: By the year 2000, U. S. students will be first in the world in science and mathematics achievement (p. 4).

In 1996, President Bill Clinton convened a second education summit with the nation's governors, at which time they committed to designing standards for each state. The No Child Left Behind Act (NCLB) of 2001 required standardized testing of these standards to ensure that all students were, in fact, achieving the state's standard course of study (Paepflow, 2011). With the adoption of Common Core Standards established in 2009, states unified their standards for language arts and math. Measurement expert Susan Brookhart contends that the counterpart to the standards and accountability movement, through which schools are held responsible for ensuring that all student learn, is standards-based grading, which could also be referred to as learning-focused grading. (Brookhart, 2011).

Problems with Grading

Problems with how students should be graded have been a source of concern for over 100 years (Meyer, 1908). Rugg (1918) stated that the one point of absolute agreement over the previous fifteen year was that the methods by which instruction was measured in the public schools should be thoroughly overhauled. He then proceeded to list three "very apparent" reasons for his statement: "(1) the striking variability in

teachers' marks; (2) the unreliability, the lack of consistency, with which teachers mark; (3) the inconsistency in the way in which teachers distribute their marks" (p. 702). In a 1972 *Time* magazine article, education professor Simon proclaimed, "The grading system is the most destructive, demeaning, and pointless thing in American education" (1972, p. 61). More recently, Ebel and Frisbie (1986) identified three reasons for the controversies regarding grading: (1) measuring educational achievement is technically difficult, (2) educational philosophies differ widely, and (3) teachers are conflicted in their roles as both advocates and judges.

Most measurement specialists agree that grades in academic subjects should be based solely on achievement measures, exclusive of other non-achievement factors, such as conduct, effort, ability, or growth, (Cross & Frary, 1999; Gronlund, 2006); however, many classroom teachers fail to follow these recommended practices for grading (Brookhart, 1993; Stiggins, Frisbie, & Griswold, 1989), even though they are highly concerned with effective evaluation (Tyler, 1935). This failure to follow recommended practices has led to grades that are a hodgepodge of achievement, effort, and attitude. (Brookhart, 1991). Even when considering academic factors alone, the weight that different teachers place on different aspects of what is to be graded can yield such a wide variance in grades that the validity of those grades could be called into question (Starch & Elliott, 1912, 1913a, 1913b). Significant studies related to problems with grading are featured in Table 1.

Table 1 Content Analysis for Significant Studies Related to the Problem of Grading

Study	Purpose	Participants	Design/ Analysis	Outcomes
Brookhart (1993)	To determine teachers' interpretation of grades	84 classroom teachers enrolled in MSED classes at Duquesne University	quantitative survey	Teachers view grades as something students earn, compensation for their work. The emphasis is more on the activities that students perform and not about what students actually have learned.
Cross & Frary (1999)	To examine teachers' grading practices	307 middle and high school teachers and their students from a non-specified U.S. school district	quantitative survey	The majority of teachers considered numerous factors other than student achievement in their grades. Both teachers and students consider such hodgepodge grading to be fair.
Starch & Elliott (1912)	To examine the variability in the way teachers assess and mark student work	152 English teachers from different high schools	quantitative descriptive	The wide variance in grades on the papers suggested that grading was highly subjective and that grades were not valid measures of performance.
Starch & Elliott (1913)	To examine the variability in the way teachers assess and mark student work	140 math teachers from different high schools	quantitative descriptive	The wide variance in grades on the papers suggested that grading was highly subjective and that grades were not valid measures of performance.
Starch & Elliott (1913)	To examine the variability in the way teachers assess and mark student work	122 history teachers from different high schools	quantitative descriptive	The wide variance in grades on the papers suggested that grading was highly subjective and that grades were not valid measures of performance.

Purposes of Grading

Not only is how students should be graded a point of contention, but the exact purpose in grading is itself problematic. Numerous purposes of grading have been cited by researchers (Airasian, 1994; Guskey & Bailey, 2001), but those purposes often conflict with each other (Carifo & Carey, 2009), and educators are not in agreement on which purpose is the most important. They then try to address all of those purposes in a single reporting device – the report card – and usually end up achieving no purpose very well (Austin & McCann, 1992). Waltman and Frisbie (1994) assert that the main purpose of report card grades is to communicate to parents their students' achievement, but that when grades are not specifically related to learning, they do not inform on academic strengths and weaknesses and can actually be counterproductive (Winger, 2005).

Guskey (2002) cited five particularly difficult challenges of grading and reporting which teachers face:

- (1) limiting the negative aspects of subjectivity,
- (2) balancing instructional concerns with grading requirements,
- (3) establishing grading criteria,
- (4) deciding what sources of evidence to use, and
- (5) relating the evidence to their purpose in grading (p. 39).

Types of Grading

Norm-Referenced

Most all grading practices fall into one of two categories – norm-referenced or criterion-referenced. Norm-referenced grading, also known as grading on the curve,

assesses one student in relation to other students. Teachers rank students according to their performance or achievement on a given measure of assessment, and then assign grades according to set percentages that correspond to the bell-shaped, normal probability curve (Guskey, 2001). Exact percentages vary among educators, but essentially the top percentage group, usually 10 to 20 percent of the class, scores the highest grade, the next percentage group, perhaps 20 to 30 percent, scores the second highest grade, and so on. This method assigns grades based on one pupil's performance compared to other pupils' performance (Airasian, 2000; Guskey, 2001); hence, students achieve high grades by performing better than their classmates, not necessarily by performing well (Bostic, 2012). Normative grading communicates nothing about a student's learning (Guskey, 2002) and creates a game of losers and winners, with the majority of the students becoming the losers (Haladyna, 1999). Additionally, normative grading negatively impacts students' relationships with each other and with the teacher (Krumboltz & Yeh, 1996).

Criterion-Referenced

Criterion-referenced grading compares a student's performance to a specific learning criterion, or clearly stated performance objective, as opposed to comparing a student's performance to that of others in the group with norm-referenced grading. Students are judged according to their own performance, regardless of that of their classmates (Guskey, 2001). Criterion-referenced grading is intended to show how much of the taught curriculum a pupil has learned (Airasian, 2000), and is a reflection of the effectiveness of the instructional program (Denton & Henson, 1979). Strong research

evidence suggests that classroom grading and reporting should always be criterion-referenced (Guskey, 2002).

Letter Grades

Schools have used letter grades, the best known and most utilized of all grading methods, since the early 1900's (Guskey, 2002). Most letter grade scales range from A to either E or F, with A being the highest performance level and E or F being the lowest. Because teachers are concerned with student motivation and self-esteem, many base their grades on a combination of criteria that takes into account individual circumstances, including elements of achievement, effort, and improvement (Brookhart, 1991; Guskey, 2001). Interpreting those grades then becomes difficult for parents and students (Friedman & Frisbie, 2000), and what teachers are trying to communicate in the grade and what parents actually interpret may not necessarily be the same (Waltman & Frisbie, 1994). Wiggins (1996) contends that a single letter grade actually hides more than it shows, forcing teachers to use too few grades to report on too many - and too many types of - tasks, but that the problem is not the letter grade itself but the lack of clear reference points for what that letter grade means. An A may mean that the student already knew the material prior to instruction, did not learn all that should have been learned but put forth great effort, or made significant improvement. Even when teachers consider strictly academic achievement alone, research has shown wide discrepancies in grading practices based upon the manner in which teachers weigh various assignments (Marzano, 2000).

Table 2 displays examples of different reference scales for interpreting letter grades. The “Less Desirable” scale uses norm-referenced language, while the “More Appropriate” scale uses criterion-referenced language (Guskey, 2002).

Table 2 Norm Referenced and Criterion Referenced Report Card Legends

Less Desirable		More Appropriate	
A =	Outstanding	A =	Excellent
B =	Above Average	B =	Good
C =	Average	C =	Satisfactory
D =	Below Average	D =	Poor
F =	Failing	F =	Unacceptable

Percentage Grades

Percentage grades are the second most commonly used grading method after letter grades. In fact, they are usually paired with letter grades. Table 3 displays a common pairing of percentage grades with letter grades.

Table 3 Sample Report Card Legend for Percentage Grades

Grade	Percentage-Based Criteria
A	90 % to 100 %
B	80 % to 89 %
C	70 % to 79 %
D	60 % to 69 %
F	less than 60 %

Percentage grades use cut-off scores based on the percentage of corrects answers, or, in the case of the report card, based on an averaged percentage of mastery from multiple assessments (Airasian, 2000). Teachers and parents both seem to prefer percentage grades. Teachers like its convenience of use and air of precision (Friedman &

Frisbie, 2000), and parents like that they know this grading method and that it makes sense to them (Guskey, 2002). Like letter grades, though, percentage grades are subject to the same potential shortcomings in unreliability of grading practices. A percent score of 85 on a report card generally does not mean that a student knows 85% of the required content but that the student scored an average of 85% on the various assessments used by the teacher (Friedman & Frisbie, 2000). An additional shortcoming is in the use of zero in averaging grades. According to assessment expert Reeves (2004), the use of zeros in a 100 point scale creates a disproportionate ratio of grading from which students may not be able to recover; moreover, insisting on using zeroes on a 100-point scale is to deem that work that is not turned in is deserving of a penalty far more severe than work that is turned in but done wretchedly.

Standards-Based Grading

Standards-based grading is based on the principle that grades should convey how well students have achieved standards (Brookhart, 2011) and should always be criterion-referenced (Guskey, 2001). Students must work towards mastery of a particular standard, and teachers must plan for and assess student mastery of that standard, basing their grades solely on mastery and no other non-academic factors. (Bradbury-Bailey, 2011). The impetuses for transition to this grading method include: (1) the inconsistencies in grading policies and practices, (2) standards-based learning and performance assessments, (3) advancements in the use of technology for reporting detailed information on student learning, and (4) the gap between common grading practices and knowledge of grading and reporting methods (Guskey, 2002).

Though recent studies have indicated a correlation between standards-based grading and standardized achievement scores, as well as increased mean scores with standards-based grading (Bradbury-Bailey, 2011; Haptonstall, 2010), teachers have traditionally been very resistant to changing their grading practices (Cross & Frary, 1999). Possible explanations for this include teachers' ability to incorporate classroom management practices into points grading (Cross & Frary, 1999), the amount of time transferred from instruction of students to performance-based assessments of students when teachers are pressured to cover numerous standards (Cooney, Bell & Fisher-Cauble, 1996), the significantly increased workload of teachers in identifying and assessing student learning goals or performance standards and in determining which evidence best supports student attainment of or progress toward those goals (Guskey & Bailey, 2001), and the struggles school leaders experience in implementing any reform effort (Guskey & Jung, 2012). Furthermore, even with the emphasis on mastery of standards, some researchers have found that standards-based assessments do not adequately report student progress on certain diagnostic skills (Rupp, Lesaux, & Siegel, 2006), and that standards-based grading does not adequately reflect student growth (Paeplow, 2011). Table 4 features three significant studies conducted on standards-based grading.

Standards-based grading does not necessitate the use of standards-based report cards, though many researchers consider them to be essential in aligning student achievement to state standards (Cherniss, 2008). A standards-based report card typically lists the grade-level learning goals or performance standards to be mastered, and a scaled mark is assigned to each standard. Table 5 displays two potential scales for standards-

based assessment, one based on achievement descriptors, the other based on behavioral descriptors (Guskey & Bailey, 2001).

Table 4 Content Analysis for Studies Related to Standards-Based Grading

Study	Purpose	Participants	Design/ Analysis	Outcomes
Bradbury-Bailey (2011)	To examine the impact of standards-based grading on African-American students in science	386 high school science students in a pre-dominantly African-American school	quantitative causal comparative	African American students scored higher with a standards-based grading system, not a standards-based report card, than did African-American students with a traditional grading system.
Haptonstall (2010)	To examine the correlation between classroom grades and the Colorado Student Assessment Program	Students from 5 Colorado school districts in grades 6 -10	quantitative correlational	Schools that used a standards-based grading system had a higher level of correlation to the Colorado Student Assessment Program and had higher mean scores on the assessment.
Paeplow (2011)	To explore the implementation of standards-based grading in the Wake County Public School System	102 elementary schools in Wake County, North Carolina	mixed methods	Teachers believed that standards-based grading did not adequately reflect student growth and that the report card was not helpful to parents who could not read English. Student grades were strongly correlated with End of Grade exams.
Rupp, Lesaux, & Siegel (2006)	To examine the relationship between performance on a standards-based assessment and a diagnostic battery of reading skills assessments in 4th grade	1,111 4th grade students and a subsample of 818 students for whom data from kindergarten was also available	quantitative causal comparative	The proficiency classifications of a standards-based assessment in reading did not accurately reflect the diagnostic component skills of reading.

Table 5 Report Card Legends for Standards-Based Report Cards

Performance Level	Achievement Descriptors	Behavioral Descriptors
4	Exceptional	Consistently/Independently
3	Proficient	Usually
2	Progressing	Sometimes
1	Beginning	Seldom

Many districts that have transitioned to standards-based report cards have been met with community resistance; parents understood letter grades, but many found number scales to be confusing (Manzo, 2001). Anecdotal evidence indicated that parents were perplexed as to why numbers were low at the beginning of the year, and the idea of numbers representing stages in a process was not clear to them (Tuten, 2007). Some teachers considered the report to be more about tracking progress for administrative reasons than for informing parents of academic achievement (Grause, 2011). Additionally, standards-based reporting forms were often too lengthy and too complicated for parents to understand and may not have adequately communicated student achievement and performance (Guskey & Bailey, 2001).

The limited amount of research that has been published on standards-based report cards has mostly involved qualitative studies of their implementation (Bryant, 2012; Olson, 2005; Panchisin, 2004). One such study found that Title I parents were confused by the report card, that many parents lacked understanding of the scoring measurements, that all participants were confused by the vagueness of the grading symbols, and that the length of the card and wording of the standards were considered weaknesses (Mathura, 2008). One recent quantitative study of significance examined the academic achievement

of fourth grade students who had transitioned to standards-based report cards. That researcher hypothesized that students would show greater achievement gains as a possible result of the more detailed information provided by the card; however, no differences in achievement were found (Craig, 2011). Paepflow (2011) found in her study that student grades on standards-based report cards were strongly correlated with End of Grade exams. This finding is consistent with the previously discussed studies on standards-based grading and with previous research that has indicated that rubric scores, which are often used on standards-based cards, have a higher correlation to standardized assessments than percentage scores (Wright & Wiese, 1988). Table 6 features significant studies related to standards-based report cards.

Motivational Effect of Grades

Positive Influence of Grades

Various studies have indicated a positive motivational influence of grades. One of the earliest and most significant was that of Ellis Page. The results of his 1958 study indicated that achievement improved when students were given positive narrative comments in addition to their grades. Later, Terwilliger (1977) determined from research studies that differential grading tends to motivate students. More recent studies include a 2004 study of Norwegian students in grades 8, 9, and 10, in which the researcher concluded that effective teachers are able to manipulate student effort through their grading methods after students who were exposed to hard grading (given good grades for high achievement only) performed significantly better than other students (Bonesronning, 2004).

Table 6. Concept Analysis for Studies Related to Standards-Based Report Cards

Study	Purpose	Participants	Design/ Analysis	Outcomes
Cherniss (2008)	To investigate elementary public school teachers' perceptions of the effectiveness of a standards-based report card.	teachers from a K-5 elementary school in California	qualitative case study	The teachers were in support of standards-based report cards, believing them to be essential to aligning state standards to student achievement.
Craig (2011)	To examine the effect of standards-based report cards on 4th grade student achievement	4th grade students from 103 elementary schools in south-eastern Massachusetts	quantitative, causal-comparative	No significant differences in academic achievement were associated with type of report card.
Mathura, (2008)	To examine how parents and teachers feel about using standards-based report cards for kindergarten students	parents and teachers in 2 elementary schools in Coweta County, Georgia	qualitative	Title I parents were confused by the report card; many parents lacked understanding of the scoring measurements; all participants were confused by the vagueness of the grading symbols; wording of the standards and length of the card were considered weaknesses.
Paepflow (2011)	To explore the implementation of standards-based grading in the Wake County Public School System	102 elementary schools in Wake County, North Carolina	mixed methods	Teachers believed that standards-based grading did not adequately reflect student growth and that the report card was not helpful to parents who could not read English. Student grades were strongly correlated with End of Grade exams.

Another 2004 study involved community college students in which the researcher compared student performance on assessments not linked to course outcomes with student performance on assessments that were linked to course outcomes. Motivation was cited as a determinant in how students performed:

It is reasonable to conclude that when student performance on assessment measures is not linked to course outcomes (i.e., course GPA or pass-fail outcomes), due to a lack of motivation, their scores cannot serve as reliable indicators of their true learning or mastery of the curriculum. However, when scores on assessment measures are linked to course outcomes, students will be motivated to maximally perform (Napoli & Raymond, 2004, p. 926).

A 10-year study by Natriello and Dornbusch (1984) indicated that students worked harder when they knew that the results would be a significant part of their grade, and that students were motivated by the rewards and punishments they would receive as a consequence of their grades. Pilcher surmised in her 1994 study of high school students that the value students placed on grades was contingent upon the internal and external punishments or rewards they would receive. In a study of college students, the majority of students perceived grades as powerful tools for administering either reward or punishment. (Pulfrey, Buschs, & Butera, 2011).

Intrinsic versus Extrinsic Motivation

Cameron versus Deci. The use of grades as motivation, as well as motivation in general, is a highly contentious educational debate (Akin-Little, Eckert, Lovett, & Little, 2004; Pulfrey, Darnon, & Butera, 2013). The controversy involves theoretical applications of intrinsic and extrinsic motivation, and two opposing camps of debate which have garnered considerable literary review are those debates between Judy

Cameron and Edward Deci (Akin-Little et al., 2004; Cameron, 2001; Deci, Koestner, & Ryan, 2001a). Deci first reported in 1971 of his research conclusions that extrinsic rewards can undermine intrinsic motivation. For several years following, his continued research sustained his conclusions (Deci, 1972a, 1972b, 1975; Deci & Ryan, 1985, 1987; Deci, Koestner, & Ryan, 1999a). Cameron began in 1994 reporting her research conclusions that reward does not generally decrease intrinsic motivation and that verbal praise increases intrinsic motivation, and later determined that, under certain conditions, rewards can increase intrinsic motivation (Cameron & Pierce, 1996; Cameron, Pierce, Banko, & Gear, 2005; Pierce, Cameron, Banko, & So, 2003). After Deci released an article in which he and his colleagues concluded that tangible rewards tended to be especially detrimental to children (Deci et al., 1999a), Cameron and her colleagues responded specifically to his article to refute his research and conclusions, arguing that (1) depending upon the method of presentation, rewards can increase, decrease, or have no effect on intrinsic motivation; (2) rewards can increase perceived self-determination; (3) in applied studies featuring characteristics of everyday life, rewards have either positive or null effects on intrinsic motivation; (4) rewards that convey the personal or social significance of a task can increase intrinsic motivation, while rewards that convey the triviality of a task can decrease intrinsic motivation (Eisenberger, Pierce, & Cameron, 1999)

Deci, Koestner, and Ryan (1999b) replied back that all their findings were reliable and called into question the methodology and conclusions of Cameron's team. When Deci released further studies (Deci, Koestner, & Ryan, 2001a), Cameron (2001) again defended her research and again called into question Deci's (Cameron, Banko, & Pierce,

2001), who again responded with his rebuttal of Cameron's work and defense of his own (Deci, Koestner, & Ryan, 2001b).

Other researchers. The debate over internal versus external motivation in education and grading began generations before Cameron and Deci. Colvin (1912) wrote that teachers could be divided into two classes – those teachers who favored marks and those teachers who opposed them. One objection of teachers that opposed them was that marks were external motivators and that pupils should not study for ulterior motives but for the sake of the subject being pursued. Colvin stated that the chief value of a marking system was in its effects on students, and that even a bad marking system was better than no marking system at all. He then recounted his experience performing tests in which students were learning to say non-sense syllables. At first, the students worked diligently because of the novelty of the exercise. Once the novelty wore off, however, student interest waned, and grades were introduced to ensure motivation. In all of Covin's subsequent experiments with school children, he found he had to use grades in order to maintain student motivation and attention. While it was hoped that at some point students develop an internal desire to study and attain knowledge, Colvin stated that at one stage of learning, if students had not studied for the sake of their grade, they would never have studied at all. Haladyna similarly stated that even though you eventually want students to develop a love for learning as their primary motivation, in the meantime, the idea of earning a grade can be a kind of carrot to keep students working hard to achieve some course goals (1999). Researchers Workman and Williams (1980) studied numerous published studies regarding extrinsic motivation and concluded the following:

- Many children who are capable of learning a skill might never acquire that skill without some extrinsic incentive.
- Many children will not engage in tasks which are academically appropriate to them without external incentives.
- Many children who had previously experienced little academic joy or success have made substantial gains through use of extrinsic rewards.
- External reinforcements can maintain and increase intrinsic interest over prolonged periods of time in on-task behavior.

Not all researchers consider intrinsic and extrinsic motivation to be opposing forces; indeed, some have found that, under certain conditions, externally motivating factors can lead to increased internal motivation, and that distinguishing between the two is not always easy. DeCharms (1968) defined the achievement motive as a competition in striving for success with a standard of excellence, but while that definition stresses intrinsic satisfaction, it can be difficult to distinguish the intrinsic aspects from the extrinsic aspects when the achievement motive is used in conjunction with incentives.

Guay and his colleagues found that students may be both intrinsically and extrinsically motivated at the same time. They may like a particular subject, but still be motivated to perform well for external reasons, such as a reward or to avoid a negative consequence (Guay et al., 2010). Lepper and his colleagues also found that students may be simultaneously internally and externally motivated, seeking out activities they naturally find enjoyable while at the same time considering closely the extrinsic consequences associated with those activities (Lepper, Corpus, & Iyengar, 2005).

Negative Influence of Grades

Though many researchers have found a positive motivational influence of grades, many have also found a negative motivational influence, especially among low performing students. Glaser (1971) determined that lack of success contributed to non-motivation more than anything else, and Stiggins (2001) found that grades held no motivational value whatsoever for student who have given up. Moreover, poor grades have been shown to lead students to discount the value of the grade (Stephan, Caudroit, Boiche, & Sarrazin, 2011). After pointing out the potential of grades to motivate students to perform, Haladyna (1999) also pointed out that low grades can effects students' self-esteem, causing them to feel stupid and experience other negative emotions. Shim and Ryan (2005) also found that while positive feedback generally increases student motivation, negative feedback generally decreases it. Ciani and Sheldon (2010) concurred, stating that it is reasonable to conclude that letter grades affect student effort and persistence, as students who earn F's are potentially more likely to disengage and to avoid similar tasks, and students who earn A's are more likely to vigorously approach similar tasks.

The negative impact of grades on self-esteem persists even at the college level (Crocker, Karpinski, Quinn, & Chase, 2003). Other potential negative motivational influences of grades reportedly include conformity, reduced teacher-student interaction, and encouragement to cheat in order to receive a passing grade (Evans, 1976). Additionally, other researchers have found that intrinsic motivation declines and positive academic beliefs and behaviors erode as students get older and progress through the

school system (Gottfried, Fleming, & Gottfried, 2001). Table 7 features significant studies related to motivation.

Future of Grading

In 2000, Marzano called for a future move to report cards with no grades, such as a standards-based report card. Today, Guskey (2013a) calls for the same. He has proposed replacing the percentage grading system with an integer grading system of 0 to 4, such as many colleges and high schools use in calculating grade-point averages (GPA). He contends that this would eliminate the problems associated with factoring in 0's and in trying to convert percentage grades to GPAs, would align with levels already often used to classify students, such as *Below Basic*, *Basic*, *Proficient*, and *Advanced*, and it would align with four-point rubrics also already often used. In conjunction with the integer grading system, Guskey (2013b) has also called for mastery learning, allowing students to practice skills repeatedly, without penalty, until they attain mastery.

Summary of the Literature

Grading first appeared in United States public schools in the mid 1800's and had become wide-spread by the early 1900's. Over the years, educators have experimented with a number of grading systems: narrative reports, letter grades, percentage grades, pass/fail or satisfactory (S)/unsatisfactory (U) conditions (Tocci, 2008), and grading on the curve (Starch, 1913). From the beginning, grading systems were wrought with controversy as researchers and educators began to closely scrutinize them (Wrinkle, 1935). Wrinkle became the first American educator to focus his career on the study of

grades and grading (Laska & Juarez, 1992). Many others have come along since and still express the same concerns as Wrinkle (Airasian, 1994; Brookhart, 1991), including their lack of validity and unreliability (Brookhart, 1993) and the different criteria teachers use when assigning them (Guskey, 2011). The concerns over grading systems and the move towards standards-based instruction have led to the implementation of standards-based report cards. Standards-based report cards come with their own set of concerns, though (Manzo, 2001).

Many educators and researchers acknowledge that grading can positively influence students' achievement and performance, and provide incentives for many students to learn (Guskey & Bailey, 2002; Hills, 1981). Frisbie and Waltman (1992) determined that most students will be motivated to achieve the highest grades, along with the accompanying recognition for such grades, and that students will be motivated to avoid the lowest grades, along with the possible accompanying negative outcomes. However, the use of grades as motivation presents an unresolved theoretical controversy (Pulfrey et al., 2013). Several studies have indicated that intrinsic motivation wanes as students progress from early elementary school through high school (Gottfried, Fleming, & Gottfried, 2001). Some researchers have found external motivators to be highly detrimental (Deci et al., 1999a), while others have found them to be essentially neutral (Dickinson, 1989), and still others have found them to be highly positive (Cameron, 2001). Most consider intrinsic motivation to be the most effective means of motivation, with intrinsic motivation being far more predictive of academic achievement than other forms of motivation (Gottfried, 1990; Hayenga & Corpus, 2010).

Guskey points out that while grades have some value as rewards, they have no value as punishments (1994) and that no research supports the idea that low grades prompt students to try harder (Guskey, 2011), though even that point is debated (Ebel, 1980). The move towards standards-based report cards is a move away from a reporting system that most parents know and understand and will probably be met with much resistance (Manzo, 2001). It is also a move away from the potentially motivating influence that grades can have.

Table 7 Concept Analysis for Studies Related to Motivation

Study	Purpose	Participants	Design/ Analysis	Outcomes
Bonesronning (2004)	To determine if there is an association between teachers who grade hard and the academic achievement of students	887 Norwegian 10th graders	quantitative causal-comparative	Students who are exposed to hard grading perform significantly better than those who are not. High achieving students are negatively impacted by easy grading. No student subgroups achieve higher when exposed to easy grading.
Ciani & Sheldon (2010)	To determine if exposure to either the letter A or the letter F prior to a task impacted student performance on the task	131 students in a large research university in the United States	quantitative quasi-experimental	Students who were exposed to the letter A prior to a task demonstrated enhanced performance, and students who were exposed to the letter F prior to a task demonstrated impaired performance.
Cameron, Pierce, Banko, & Gear (2005)	To explore how rewards for achievement during the learning process impact intrinsic motivation	119 university students in an introductory psychology class	quantitative quasi-experimental	Achievement based given rewards given during or after learning increased the intrinsic motivation in the students participating in the target activity.

Table 7 continued

Study	Purpose	Participants	Design/ Analysis	Outcomes
Cameron, Pierce, Banko, & Gear (2005)	To explore how rewards for achievement during the learning process impact intrinsic motivation	119 university students in an introductory psychology class	quantitative quasi-experimental	Achievement based given rewards given during or after learning increased the intrinsic motivation in the students participating in the target activity.
Deci (1971)	To investigate the effects of external rewards on intrinsic motivation to perform an activity	24 introductory psychology students	quantitative quasi-experimental	Intrinsic motivation tended to decrease when money was used as a reward but tended to increase when positive feedback and verbal praise were given as rewards.
Guay, Chanal, Ratelle, Marsh, Larose, & Boivin (2010)	To investigate the academic motivations of elementary students	425 French-Canadian children from three elementary schools	quantitative quasi-experimental	The self-determination continuum is supported in reading, but not in math or writing. Motivations within one subject are more closely related to other motivations within that subject than to motivations towards other subjects.
Gottfried, Fleming, & Gottfried (2001)	To investigate the continuity of academic intrinsic motivation through the use of a longitudinal study	107 students measured at ages 9, 10, 13, 16, and 17	quantitative causal-comparative	Academic intrinsic motivation remains stable from elementary through high school for both verbal and math areas.

Table 7 continued

Study	Purpose	Participants	Design/ Analysis	Outcomes
Hayenga & Corpus (2010)	To identify and evaluate combinations of extrinsic and intrinsic motivation and their stability over time	388 6th, 7th, and 8th grade students from a public middle school in Portland, Oregon	quantitative survey	Students with a combination of high intrinsic motivation and low extrinsic motivation received higher grades than students with any other combination and maintained more stability over the course of a year than any other group.
Lepper, Corpus, & Iyengar (2005)	To examine the relationship between intrinsic and extrinsic motivation and how they are related to academic outcomes	797 3rd through 8th graders from two California public school districts	quantitative quasi-experimental	Intrinsic and extrinsic motivation are separate constructs. Intrinsic motivation significantly decreased from 3rd to 8th grade and is positively correlated to academic achievement.
Malone, Nelson, & Van Nelson (2002)	To examine whether or not there were differences in grading patterns between the plus/minus grading system and the A-F grading system	8,088 master's level students	quantitative survey	Grade point averages declined in some academic areas. Faculty opinion was that the plus/minus system was more appropriate for graduate students.

Table 7 continued

Study	Purpose	Participants	Design/ Analysis	Outcomes
Napoli & Raymond (2004)	To evaluate whether or not an assessment is graded influences the outcome of the assessment	80 community college students enrolled in introductory psychology	quantitative quasi-experimental	When student assessments are not graded and not linked to pass/fail, they are not reliable indicators of student learning.
Natriello & Dorn-busch (1984)	To explore the impact of how teachers evaluate on student behavior and effort	35 schools; 2,559 students; 343 teachers; 109 classroom observations	Mixed Methods	Students put more effort into evaluations for which they receive sanctions - grades, rewards, future benefits, social acceptance
Page (1958)	To investigate if and when teacher comments cause a significant improvement in student performance	74 secondary classrooms in 2 school districts; 2139 students	quantitative causal-comparative	Students who received positive comments in addition to a letter grade on assessments scored higher on subsequent assessments than students who received a letter grade only
Pierce, Cameron, Banko, & So (2003)	To examine how rewards affect intrinsic motivation when they were tied to increasingly demanding performance standards	60 university undergraduate students	quantitative quasi-experimental	Students who received rewards while completing a progressively demanding performance task spent more time on the task in a free choice situation than those students who either received no reward or were rewarded for attaining a constant level.

Table 7 continued

Study	Purpose	Participants	Design/ Analysis	Outcomes
Pilcher (1994)	To investigate how grades were assigned by teachers and perceived by students and parents	Six cases consisting of a high school student, his/her parent, math teacher, and English teacher	qualitative case study	Grades represent a combination of achievement, ability, and effort. Parents interpreted grades differently than teachers intended. The internal and external rewards students received for grades determined the value they placed on grades.
Pulfrey, Darnon, & Butera (2013)	To assess the power of task performance and task autonomy on intrinsic motivation	90 students in 7th to 9th grade in a public secondary school	quantitative quasi-experimental	Perceived task autonomy significantly affected continued task motivation. High grades and no grades enhanced intrinsic motivation.

CHAPTER III

METHODOLOGY

Introduction

Districts across the country are transitioning to standards-based reporting, replacing the single letter grade for a given subject with rubrics or scaled scores for numerous standards within that subject. Concerns that a single letter grade cannot convey student achievement accurately, in addition to the movement to standards-based learning, have prompted many school districts to make this change. In the quest to provide more detailed information about student achievement, however, might a change in student achievement actually take place? When the potentially motivating factor of letter grades is taken away from students, might student achievement decline?

This study explored the relationship between reading achievement in the third grade and standards-based report cards. The research question was “Is there a difference between the reading achievement of third-grade students using traditional A-F letter grade report cards and those students using standards-based report cards?” The hypothesis that guided this study was:

H₁: A difference exists between the reading achievement of third-grade students using traditional A-F letter grade report cards and those students using standards-based report cards.

The null hypothesis was:

H₀: A difference does not exist between the reading achievement of third grade students using traditional A-F letter grade report cards and those students using standards-based report cards.

Research Design

This study was conducted using a quantitative approach with a causal-comparative research design. The research question, “Is there a difference between the reading achievement of third-grade students using traditional A-F letter grade report cards and those students using standards-based report cards?”, was best answered with a causal-comparative design because numeric data was used to determine if a relationship exists between student achievement and report card type and no variables were manipulated. The researcher utilized pre-existing data obtainable from the Georgia Department of Education website, www.gadoe.org. The dependent variable was the percentage of students who passed the CRCT, and the independent variable, report card type, was not manipulated.

Population and Sampling

Third grade was chosen as the target grade level for this study for three reasons: (1) third grade is the first grade at which students experience high-stakes testing and are required to pass the reading portion of the CRCT to move on to the next grade, (2) third grade is the first year that the CRCT is administered, and (3) third grade students have been exposed to fewer interventions and external factors influencing achievement compared to fourth and fifth students. The sample was convenience sample, determined

by the maximum number of Georgia schools that transitioned to standards-based report cards at the third grade level in a given year between 2001 and 2012.

To determine the sample population, testing years were narrowed to between 2003 and 2012. The first administration of the CRCT to Georgia third graders was in 2002. A comparison of achievement data from before and after the transition to standards-based report cards necessitates that the 2003 school year be the earliest possible year of transition. Moreover, the state of Georgia changed its curriculum in the 2013 school year with the adoption of the Common Core Georgia Performance Standards (CCGPS), necessitating that the latest possible year of transition be school year 2012. The year of transition to standards-based report cards was then found for each school. The year in which the maximum number of schools transitioned to standards-based report cards at the third grade level became the determinant for including those schools in the sample.

Several school districts implemented standards-based report cards in waves, beginning with lower grades and slowly progressing up to third grade. The implementation year for third grade was considered for this study. In the 2010 school year, five Georgia school districts with a total of 116 elementary schools transitioned to standards-based report cards – Cobb County, Haralson County, Muscogee County, Oconee County, and Rockdale County. Cobb County was the largest of the Georgia districts to transition to standards-based report cards in 2010 and is the second largest school system in the state of Georgia and the 24th largest in the country. With only six schools in the district, Haralson was the smallest of the systems that transitioned to

standards-based report cards in 2010. Tables 8 and 9 provide enrollment and demographic data of each of the five school systems.

Instrumentation

The Georgia CRCT is designed using the professional standards established by the American Psychological Association, the National Council of Measurement in Education, and the American Educational Research Association in a process that ensures both validity and reliability. The Georgia Department of Education has published their process for ensuring the validity and reliability of the CRCT in *An Assessment & Accountability Brief: 2013 CRCT Validity and Reliability (2013)*. Validity of the CRCT is evidenced through a multi-step process. First, there is a clear identification of the purpose of the test, which is to measure how well students have mastered the state's curriculum, to identify the areas where students need improvement, to inform various stakeholders of academic progress in meeting state standards, to meet the requirements of the No Child Left Behind Act, and to gauge the overall quality of education in the state of Georgia. Next, committees of educators review the curriculum and establish what will be assessed and how it will be assessed, generating a test blueprint and test specifications. From these, content domain specifications are produced and then converted into a document entitled *CRCT Content Descriptions*. That document, along with an additional document, *CRCT Content Weight*, which details the relative proportion of items that will be included on each content area test, are then made available on-line for all stakeholders.

Table 8 Enrollment Data of Georgia School Districts that Adopted Standards-Based Report Cards in 2010

County	Total Student Population	Total Number of Schools	Total Number of Elementary Schools
Cobb	106,000	112	67
Haralson	3700	6	4
Muscogee	32,000	62	34
Oconee	6680	10	5
Rockdale	16,200	23	11

Table 9 Demographic Data of Georgia School Districts that Adopted Standards-Based Report Cards in 2010

County	% African-American	% White	% Other Races	% Econ Disadv	% with Disabilities	% Male	% Female
Cobb	31.4	42.4	26.2	44	11.7	51	49
Haralson	3.2	92.4	4.4	62	16.6	48	52
Muscogee	58	29	13	63.8	15.1	49.6	50.4
Oconee	5	88.4	6.6	23	8	50.4	49.6
Rockdale	61.6	20.25	18.15	69	5	52	48

Following that process, professional assessment specialists write the test questions, which are then reviewed by committees of Georgia educators for curricular alignment, suitability, and potential bias. Items are field tested through embedding with operational tests, ensuring that the field test items are taken under standard test conditions by a representative group of motivated students. Once field tested, the items and their accompanying performance data were analyzed by another committee of Georgia educators. Accepted items are banked for inclusion on future operational tests.

The next stage in the process is to select items for a test from based on a blueprint developed by Georgia educators. Each form of a test assesses the same range of content and carries the same statistical attributes. The final stage is to score tests and distribute results. Raw scores are converted to scale scores and are reported as performance levels. The Georgia Department of Education ensures that validity of the CRCT by attending carefully to this test development process.

Various reliability indices for the CRCT have indicated that its results are consistent and can be generalized. Cronbach's alpha reliability coefficient, which measures internal consistency, indicated strong reliability ($\alpha = .90$), for the third grade reading test. Additionally, the standard error of measurement (SEM), an index of the random variability in test scores, also indicated strong reliability ($SEM = 2.3.7$). The strength of these indicators of reliability supports the claims of validity.

The reading portion of the CRCT is divided into three domains: (1) reading skills and vocabulary acquisition, (2) literary comprehension, and (3) reading for meaning. Previous tests have included two sections of reading, and each section contained 30 questions. The test is administered in April of each year over the course of a 2-week time

period by teachers with valid teaching certifications within the state of Georgia. Students are classified into two categories according to their scores, “does not meet” standards or “meets” standards. Those students who meet standards may also fall into an additional category of “exceeds” standards. The state of Georgia commissions a committee each year to set the “cut” scores for each assessment. These committees, which usually consist of educators, content area specialists and state administrators, examine the test items and field test data which have been matched to the state curriculum to determine if a minimally competent student would get those items correct. The committees’ recommendations regarding the questions are taken and used to create the cut scores. The cut scores may vary from year to year (What Do My Child’s Test Scores Mean, n.d.).

Procedures

A list of all districts that had transitioned to standards-based report cards was compiled, and the exact year at which standards-based report cards was implemented at the third grade level was obtained. The year in which the most number of schools implemented standards-based report cards at the third grade level was chosen as the pivotal year in which to compare the passing rate of third graders from within those schools to third graders within those schools from the previous year. More schools implemented standards-based report cards in the 2010 school year than in any other year.

The Georgia Department of Education maintains CRCT data for each school dating back to 2002, when the CRCT was first implemented. The data is disaggregated according to subject, grade, race, gender, socio-economic status, and disability status.

The available data from each selected school was compiled on an Excel spreadsheet and later imported to SPSS.

The rows of the Excel spreadsheet included the following categories: all, male, female, black, white, ED (economically disadvantaged), not ED (not economically disadvantaged), SWD (students with disabilities), and S w/o D (students without Disabilities). For each category, the columns of the spreadsheet included the following: district, school, 2009 report card type, 2009 % did not meet, 2009 % met, 2009% exceeded; 2010 report card type, 2010 % did not meet, 2010 % met, 2010% exceeded.

Data Analysis

The researcher used the chi-square test to examine differences in the reading achievement of third grade students using traditional A-F report cards and those students using standards-based report cards. Within the selected schools, the percentage of students meeting and exceeding standards on the Georgia CRCT at the third grade level prior to the implementation of standards-based report cards were compared with the percentage of students meeting and exceeding standards at the third grade level in the school year of implementation. Differences in the percentage of students meeting and exceeding standards beyond what is normally expected were examined. The data used in the study was categorical and dichotomous, thus requiring the use of nonparametric statistics (Cohen & Lea, 2004). The independent variable was report card type, with a classification of either traditional A-F letter grade or standards-based, and the dependent variable was the percentage of either passing or failing.

According to Lomax (2007), the chi-square statistic can be used to determine if the observed outcomes in more than one category of a categorical variable differ from what is expected *a priori*. Additionally, it can be used to determine the exact categories which account for the observed differences, making it one of the most useful tools of analysis when testing hypotheses of nominal data (McHugh, 2013). The effect size was measured by the phi coefficient since the variables are dichotomous. A phi-coefficient of .5 or greater would indicate a strong relationship, a phi-coefficient between .3 and .5 would indicate a moderate relationship, and a phi-coefficient between .1 and .3 would indicate a weak relationship (Cohen, 1988). SPSS was used to calculate the chi square test statistic and phi-coefficient.

Limitations

This study was limited by the use of convenience sampling in selecting schools that transitioned to standards-based report cards at a set time. The study was further limited by the use of a non-parametric statistic. The results of a parametric statistic are based on the mean. The results of the chi square are not based on the mean, which limits its robustness and increases the likelihood of Type I errors, falsely rejecting the null hypothesis. The chi square statistic simply allows the researcher to determine whether the observed data is different from the expected data (Siegal & Castellan, 1988). The chi square statistic is also sensitive to large sample sizes. For this reason, the effect size coefficient was used to determine if the significance was meaningful.

Assumptions

The researcher made certain assumptions regarding the data. One assumption was that the frequency data within each category was normally distributed. Another assumption was that the collected data were frequencies in discrete, nominal data. The researcher further assumed that the samples were independent and that the frequency counts in each cell was greater than 20 (Siegal & Castellan, 1988).

Data Interpretation

The chi square statistic was compared to the critical value from a chi square table. If the chi square statistic was equal to or greater than the critical value (Siegal & Castellan, 1988), then the null hypothesis was rejected, indicating that there was a statistically significant difference in the percentage of students passing the CRCT following the implementation of standards-based report cards than was expected, based upon scores from the previous year. In this case, the effect size using the phi-coefficient will be examined. If the chi square statistic is less than the critical value, then the null hypothesis will fail to be rejected, and no statistically significant difference will have been found between reading achievement scores among third grade students who receive traditional letter grade report cards and those students who receive standards-based report cards.

Implications

Many school districts in the state of Georgia, as well as other states across the nation, have transitioned to standards-based report cards. Some of the reasons for this transition include the national shift to standards-based instruction and the numerous purported problems with traditional grading methods. While researchers have examined the implementation and perceptions of standards-based report cards, few have yet to report possible relationships between standards-based report cards and academic achievement. This research will add to the limited number of published studies on standards-based report cards and student achievement. The results could guide districts in making more informed choices regarding best practices for reporting student achievement.

Summary and Expectations

Standards-based report cards are increasingly becoming the reporting method of choice in many districts across the county, yet limited studies have indicated whether or not this trend may actually impact student achievement. Grades are commonly agreed to be a motivational influence for many students; however, standards-based reporting changes the way in which students receive grades. This researcher proposes a causal-comparative study to determine if an association exists between the transition to standards-based report cards and student achievement in third-grade reading.

CHAPTER IV

RESULTS

Introduction

This study was conducted using a quantitative approach. A causal-comparative design was used to explore the relationship between the implementation of standards-based report cards and the academic achievement of third grade students in reading on the Georgia CRCT. The researcher examined the relationship between report card type and CRCT pass/fail rates for the school year prior to the implementation of standards-based report cards and the school year of implementation. The question guiding this research was, “Is there a difference between the reading achievement of third-grade students using traditional A-F letter grade report cards and those students using standards-based report cards?” Differences were further explored according to gender, race, disability status, and socio-economic status. The hypothesis guiding this study was:

H₁: A difference exists between the reading achievement of third-grade students using traditional A-F letter grade report cards and those using standards-based report cards.

The null hypothesis was:

H₀: A difference does not exist between the reading achievement of third grade students using traditional A-F letter grade report cards and those using standards-based report cards.

Descriptive Data

The research data for this study were the CRCT scores of third grade students from five different Georgia school districts during the 2009 school year and the 2010 school year. All data were obtained from the Georgia Department of Education website. The data included a total of 118 schools; 63 schools within the sample received Title I funding. Table 10 displays the breakdown of these schools by district.

Table 10 Number of Schools Included in the Study

District	Number of Schools	Number of Title I Schools
Cobb	66	27
Haralson	2	2
Muscogee	35	23
Oconee	4	2
Rockdale	11	9
Total	118	63

For the two testing years of the study, a total of 24,904 student test scores were considered. Those scores were disaggregated according to race, gender, disability status, and economic status. Table 11 displays the specific subgroups included in the study, as reported by the Georgia Department of Education, and the total number of test participants during the 2009 and 2010 school years. In Table 12, those data are further disaggregated by school year and school district. For the sake of student privacy, the state of Georgia does not release data on any subgroup within a school if that subgroup consists of less than 10 students.

Table 11 Subgroup Populations Examined in the Study

Subgroup	Total Number of Test Participants
All	24,904
Black	9,070
White	9,701
Male	12,748
Female	12,156
Students with Disabilities	2,166
Students without Disabilities	21,866
Economically Disadvantaged	12,146
Not Economically Disadvantaged	12,138

Table 12 Subgroup Populations Disaggregated by County and Year

	Cobb		Haralson		Muscogee		Oconee		Rockdale	
	2009	2010	2009	2010	2009	2010	2009	2010	2009	2010
Total	8086	8066	300	289	2442	2463	478	419	1161	1200
Black	2414	2419	0	0	1444	1409	0	0	681	703
White	3392	3226	269	260	612	669	378	360	269	266
Male	4209	4146	145	132	1235	1233	233	216	594	605
Female	3877	3920	155	157	1207	1230	245	203	567	595
SWD	861	857	43	40	144	153	0	0	38	30
Sw/oD	7070	7015	257	249	2164	2102	429	389	1068	1123
ED	3370	3588	171	195	1603	1553	108	99	691	768
not ED	4529	4282	129	94	730	782	370	320	470	432

Note: SWD represents students with disabilities; Sw/oD represents students without disabilities; ED represents economically disadvantaged; not ED represents not economically disadvantaged.

Data Analysis

To determine if a difference existed in the reading scores of third-grade students using traditional A-F letter grades and those using standards-based report cards, data from each of the five school districts was obtained from the website www.gadoe.org. That data was compiled into an Excel spreadsheet and then imported to SPSS. The independent variable was report card type, with a classification of either traditional A-F letter grade or standards based. The dependent variable was the percentage of students either passing or failing. The chi-square statistic was calculated to determine if observed outcomes from 2010 differed from what was expected *a priori* based upon the 2009 data. The effect size was measured by the phi-coefficient. Data were analyzed not only for the total number of students but also for the following sub-groups: male, female, black, white, students with disabilities, students without disabilities, economically disadvantaged, and not economically disadvantaged.

Results

Descriptive statistics were run for each school district. The mean passing rates for each school district varied little between the two testing years, with a difference of 1.00 in Oconee County being the greatest variance. The mean passing rate for Cobb County in 2010 ($M = 95.21$; $SD = 5.11$) was slightly higher than in 2009 ($M = 94.67$; $SD = 5.06$). In Haralson County, the 2009 passing rate ($M = 89.00$; $SD = 1.41$) was slightly higher than the 2010 passing rate ($M = 88.5$; $SD = 0.71$). In Muscogee County, the 2010 mean passing rate ($M = 90.57$; $SD = 7.96$) was slightly higher than in 2009 ($M = 89.94$; $SD = 8.21$). The 2010 mean passing rate in Oconee County ($M = 98.00$; $SD = 1.41$) was slightly higher than in 2009 ($M = 97$; $SD = 2.16$). Lastly, in Rockdale County, the 2009

mean passing rate ($M = 96.64$; $SD = 1.96$) was slightly higher than the 2010 passing rate ($M = 96.09$; $SD = 2.43$).

A chi-square was conducted to determine if there was a difference between the reading achievement of third-grade students using traditional A-F letter grade report cards and those students using standards-based report cards. Phi coefficient was calculated to determine the effect size. Based on the data analysis, there was not a statistically significant difference between third grade reading achievement in 2009 ($M = 93.43$; $SD = 6.37$) with traditional A-F letter grade report cards and in 2010 ($M = 93.90$; $SD = 6.28$) with standards-based report cards ($\chi^2 = .03$; $p > .05$; $\phi = .01$). The mean percentage of passing scores from 2009 to 2010 increased by 0.47, and the standard deviation decreased by 0.09.

Descriptive statistics, as well as chi square and phi coefficient, were also calculated for each subgroup. Subgroup data were not reported in schools if less than 10 students were in the subgroup. Males had the least change in mean percentage of passing scores between 2009 ($M = 91.88$; $SD = 8.01$) and 2010 ($M = 91.85$; $SD = 8.63$) with only a 0.03 decrease. The standard deviation varied by only 0.62. With a chi square statistic of 0.0004 and phi coefficient of .01, they also had the weakest relationship between report card type and reading achievement ($\chi^2 = .0004$; $p > .05$; $\phi = .01$). The mean percentage passing rate for females increased from 2009 ($M = 95.25$; $SD = 5.71$) to 2010 ($M = 96.00$; $SD = 4.92$) by 0.75. Chi square indicated no statistically significant difference between their reading achievement and their report card type ($\chi^2 = .13$; $p > .05$; $\phi = .05$).

Both subgroups of race had slight increases in passing rates. Students categorized as black had a 0.13 increase in mean passing rate from 2009 ($M = 92.12$; $SD = 7.54$) to 2010 ($M = 92.25$; $SD = 7.59$). The chi square value of 0.01 and phi coefficient of .01 indicated no statistically significant difference between report card type and reading achievement ($\chi^2 = 0.01$; $p > .05$; $\phi = .01$). Students categorized as white increased by .4 their mean percentage of passing the CRCT from 2009 ($M = 96.78$; $SD = 5.13$) to 2010 ($M = 97.18$; $SD = 3.75$). A chi square of 1.49 and phi coefficient of .06 indicated no statistically significant relationship between reading achievement and report card type ($\chi^2 = 1.49$; $p > .05$; $\phi = .06$).

Both groups of students classified according to disability status also had slight increases in mean percentage rates. Students with disabilities increased by 0.65 from 2009 ($M = 81.62$; $SD = 17.15$) to 2010 ($M = 82.27$; $SD = 17.94$). Standard deviation increased slightly from 17.15 to 17.94. Chi Square indicated no statistically significant difference ($\chi^2 = .32$; $p > .05$; $\phi = .03$). Students without disabilities' passing rate increased by 0.62 from 2009 ($M = 95.07$; $SD = 5.90$) to 2010 ($M = 95.69$; $SD = 4.99$). Standard deviation decreased slightly from 5.90 to 4.99. Report card type was not statistically significantly related to student achievement in reading ($\chi^2 = .32$; $p > .05$; $\phi = .03$).

Economically disadvantaged students had the greatest mean increase from 2009 ($M = 90.35$; $SD = 7.27$) to 2010 ($M = 91.64$; $SD = 7.08$) at 1.29 percentage points. Standard deviation declined slightly by 0.19. Despite having the greatest increase, the chi square statistic still indicated no statistically significant difference between academic achievement in reading and report card type ($\chi^2 = .98$; $p > .05$; $\phi = .05$). Students who

were not economically disadvantaged had a mean decline of 0.20 from 2009 ($M = 97.60$; $SD = 3.05$) to 2010 ($M = 97.40$; $SD = 4.04$), and a .99 increase in standard deviation. As in all the other subgroups, academic achievement in reading was not statistically related to report card type, with a chi square of 3.65 ($\chi^2 = 3.65$; $p > .05$; $\phi = .10$).

Summary

To answer the research question, “Is there a difference in the reading scores of third-grade students using traditional A-F letter grades and those students using standards-based report cards?” a chi square test was conducted for the total sample population as well as for reported subgroups within the population. For the total sample population, the mean percentage passing rate varied by less than one-half of a percentage point. All subgroups had less than one percentage point variance in mean passing rates with the exception of economically disadvantaged students, who had a 1.29 increase. Males and not economically disadvantaged students had slight decreases in mean passing rates, while all other subgroups had slight increases. Neither for the total sample population nor for any subgroup was there a statistically significant relationship between report card type and academic achievement in third grade reading.

CHAPTER V

SUMMARY

Summary

A Nation at Risk, the 1983 report of the status of education in America, initiated a new era of educational reform and marked the beginnings of the standards-based movement (Marzano & Kendall, 1988). As the standards-based movement grew, the call for a grading system to be more closely aligned to those newly developing standards also grew (Guskey, 2001). Standards-based grading, a grading practice based solely on evaluation of standards' mastery, and standards-based report cards, a reporting practice whereby a scaled or rubric score is assigned to each standard individually, were subsequent outcomes. While standards-based report cards may provide more detailed information about student performance on specific tasks (Bostic, 2012), they eliminate the potentially motivating factor of grades, which many assessment experts have acknowledged as one purpose of grading (Airasian, 1994).

A quantitative study with a causal-comparative design was undertaken to answer the research question, "Is there a difference between reading achievement of third grade students using traditional A-F letter grade report cards and those students using standards-based report cards?" In 2010, five Georgia school districts with a total of 118 elementary schools transitioned to standards-based report cards. The chi square statistic was calculated to determine if a relationship existed between the percentages of third-grade students passing the reading portion of the Georgia CRCT in 2010 with standards-based report cards compared to 2009 with traditional A-F letter grade report cards. The

phi coefficient was also calculated to determine the effect size. Over the course of the 2-year time period, a total of 24,904 student test scores were considered. In addition to analyzing the total number of third-grade reading scores, the scores of the following subgroups were also analyzed: black, white, male, female, students with disabilities, students without disabilities, economically disadvantaged, and not economically disadvantaged.

Interpretations

The question guiding this research study was, “Is there a difference between the reading achievement of third-grade students using traditional A-F letter grade report cards and those students using standards-based report cards?” A chi square test statistic was calculated to determine if such a relationship existed. A phi coefficient was also calculated to determine the effect size. The hypothesis guiding this study was:

H₁: A difference exists between the reading achievement of third-grade students using traditional A-F letter grade report cards and those using standards-based report cards.

The null hypothesis was:

H₀: A difference does not exist between the reading achievement of third grade students using traditional A-F letter grade report cards and those using standards-based report cards.

The significance of the chi square ($\chi^2 = .03$; $p > .05$; $\phi = .01$) was greater than .05, leading the researcher to reject the hypothesis and accept the null hypothesis that a difference does not exist between the reading achievement of third grade students using traditional A-F letter grade report cards and those using standards-based report cards.

The null hypothesis was also accepted for all subgroups. Economically disadvantaged students, however, did have the greatest difference in mean passing rates from 2009 to 2010 with an overall increase of 1.29 percentage points. Despite the lack of statistical significance, these results are consistent with other studies that have found that the elimination of failing grades is beneficial for certain at-risk populations (Craig, 2011). Many educators have reported on the negative consequences of low grades, including a loss of self-esteem which causes students to feel stupid and experience other negative emotions (Haladyna, 1999); a decrease in student motivation (Shim and Ryan, 2005); and student disengagement from tasks similar to ones in which they have previously failed (Ciani & Sheldon, 2010). Glaser (1971) determined that lack of success contributed to non-motivation more than anything else. A standards-based report card would reflect that a student had not attained a standard, as opposed to having failed a standard or subject. Craig (2011) stated that because traditional grades tend to be more representative of conformity and work habits than of concept mastery, at-risk students may be more harmed by traditional grading methods than are other students. Students at-risk of learning, such as economically disadvantaged students, may respond more favorably to a lower score on the continuum of progress on a standards-based report card than a failing grade on a traditional report card.

Conclusions

The chi square statistic indicated that no statistically significant relationship existed between report card type and reading achievement in the third grade. Further analysis of the subgroups indicated no statistically significant relationships for them as

well. These results are in keeping with a previous study in which a causal-comparative design was used to examine the impact of report card type on the academic achievement of fourth grade students in math. That study found that report card types of standards-based, traditional A-F, or mixed had no impact on academic growth in math for the sample population (Craig, 2011).

Practical Implications

Dissatisfaction with common grading practices has been a controversial issue in education for over a hundred years (Meyer, 1908), as have calls for overhauling the methods by which teachers measure instruction (Rugg, 1918). Standards-based grading has evolved as a solution to the hodgepodge of grading practices that teachers commonly employ (Cross & Frary, 1999). Prior studies have shown standards-based grading to be more closely correlated to standardized test scores and to an increase in mean test scores (Haptonstall, 2010; Bradbury-Bailey, 2011). Tomlinson and McTighe (2006) have identified six principles of effective standards-based grading and reporting: (1) Grading and reporting should be based on learning goals and performance standards which have been clearly specified, (2) Only valid evidence should be used for grading, (3) Established criteria, and not arbitrary norms, should be the basis for grading, (4) Not all assessments should be included in grades, (5) Grading should not be based on averages, and (6) Factors other than achievement should be reported separately. Standard-based grading, however, does not necessitate the use of a standards-based report card, and these principles can be followed even with traditional reporting forms.

That empirical evidence has not shown the type of report card to significantly impact student achievement may give school districts pause in choosing to develop and implement standards-based report cards. Developing standards-based report cards is a multi-step process and requires a considerable amount of time and effort from teams of educators and other stakeholders. Guskey (2004) has described the process. The standards, or major learning goals, must first be identified. Then the specific performance criteria necessary to show mastery of the standard must be established. Benchmarks for achieving each standard must also be established. Labels that are meaningful to parents, students, and other stakeholders must then be attached to the benchmarks.

These labels, however, rarely hold the same meaning for parents as they do for educators, and even amongst educators there is sometimes confusion. Guskey (2004) goes on to say that parents tend to interpret the labels according to their own experiences with grades, which usually are traditional A-F letter grades. The label that corresponds to the highest level of attainment of the standard is interpreted as an “A”, the next level as a “B”, and so forth. Grading and reporting become more about challenges in effective communication than in quantifying student achievement.

Other studies and anecdotal evidence have expounded on the challenges of parents to make meaning out of standards-based report cards. Tuten (2007) found that parents were perplexed as to why numbers were low at the beginning of the year and that the idea of numbers representing stages in a process was not clear to them. Manzo (2001) also reported that number scales are confusing to parents. Guskey and Bailey (2001) have reported that standards-based report cards are often too lengthy and too

complicated for parents to understand and therefore may not adequately communicate student achievement and performance. Mathura (2008) also found that many parents were confused by the card and lacked understanding of the scoring measurements. Moreover, teachers, students, and parents alike were confused by the vagueness of the grading symbols and considered the length of the card and wording of the standards to be weaknesses. Grause (2011) additionally reported that teachers considered the report to be more about tracking progress for administrative reasons than for informing parents of academic progress.

Over the years, many researchers have detailed multiple purposes for grades and for report cards (Munk & Bursuck, 2001; Marzano, 2000; Resh, 2009; Wrinkle, 1947). Assessment expert Airasian (1994) contends that many agree that the general purpose of a report card is to communicate information about a pupil's academic achievement. If parents, and even some teachers, find standards-based report cards to be so confusing, are they actually serving the purpose of communicating a student's performance? In the absence of data to indicate that they impact student achievement either positively or negatively, school districts seeking to improve with their current reporting methods should consider the time, expense, and communication challenges of standards-based report cards.

Limitations

Numerous factors influence students' achievement, including family dynamics, socio-economic status, school climate, teacher effectiveness, curriculum, intervention programs, and others. To narrow potential changes in academic achievement to only one

source, standards-based report cards, would not be realistic, which is why this researcher explored relationships and not causes. Though similar results were obtained with a different population in a different state in a study of fourth grade mathematics achievement and report-card type (Craig, 2011), the lack of relationship between report card type and student achievement in third grade reading is limited to the population sampled. More studies with increased population samples would need to be conducted before generalizing these results. Additionally, the grading practices behind the reporting practices would need to be examined as well, since some research has shown that certain grading practices are associated with academic growth and achievement.

Recommendations for Future Study

This study examined the relationship between academic achievement in reading in the third grade with report card type, but only looked at the relationship in the year prior to and the year of implementation of standards-based report cards. No relationship was found between report card type and reading achievement. Further studies are needed to examine whether or not these results would hold true over a multi-year period. As students are further disassociated with A-F grades, does the loss of grades as motivation have a compounding effect that is manifested in later years? Conversely, as teachers become more adept at the standards-based grading practices that should be incorporated into standards-based report cards, does student achievement subsequently begin to increase?

Additional studies should also explore the long-term effect of a no-fail policy on economically disadvantaged students, as well as other student populations. Some short-

term studies have found increased mean scores on overall grades and on standardized tests on certain populations when failing grades are eliminated. Is this improvement sustained over a multi-year period? Ebel (1980) has reported that the removal of the threat of failure removes the incentive to work to avoid failure. Does a practice that produces a short-term gain ultimately produce a long-term loss, or does it, too, have a compounding positive impact over time? A comparison of student achievement from one year to the next provides only a portion of the full amount of data to be explored to better determine causal relationships involving student achievement.

REFERENCES

- Airasian, P. W. (2000). *Assessment in the classroom: A concise approach* (2nd ed.). Boston: McGraw Hill.
- Airasian, P. W. (1994). *Classroom assessment* (2nd ed.). Boston, MA: McGraw Hill.
- Akin-Litte, K. A., Eckert, T. L., Lovett, B. J., & Little, S. G. (2004). Extrinsic Reinforcement in the Classroom: Bribery or Best Practice. *School Psychology Review*, 33(3), 344-362.
- Austin, S., & McCann, R. (1992). "Here's another arbitrary grade for your collection": A statewide study of grading policies. (Paper presented at the annual meeting of the American Educational Research Association, San Francisco). Retrieved from <http://eric.ed.gov/>
- Bainbridge, C. (2013) Intrinsic Motivation. Retrieved from: <http://giftedkids.about.com/od/glossary/g/intrinsic.htm>
- Bonesronning, H. (2004). Do the teachers' grading practices affect student achievement? *Education Economics*, 12(2), 151-167.
- Bostic, K. (2012). Standards based grading in the middle school. *International Review of Social Sciences and Humanities*, 3(1), 17-23.
- Bradbury-Bailey, M. (2011). *A preliminary investigation into the effect of standards-based grading on the academic performance of African-American students* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3511593)
- 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.
- Brookhart, S. M. (2011). Starting the conversation about grading. *Educational Leadership*, 69(3), 10-14.
- Bryant, M. G. (2012). *Case study: Development of a standards-based report card for lower elementary grades (first to third)* (Master's thesis). Available from ProQuest Dissertations and Theses database. (UMI No. 1520559)
- Cameron, J. (2001). Negative effects of reward on intrinsic motivation: A limited phenomenon: Comment on Deci, Koestner, and Ryan (2001). *Review of Educational Research*, 71(1), 29-42.

- Cameron, J., Banko, K. M., & Pierce, W. D. (2001). Pervasive negative effects of rewards on intrinsic motivation: The myth continues. *The Behavior Analyst*, 24(1), 1-44.
- Cameron, J., & Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational Research*, 64(3), 363-423.
- Cameron, J., & Pierce, W. D. (1996). The debate about rewards and intrinsic motivation: Protests and accusations do not alter the results. *Review of Educational Research*, 66(1), 39-51.
- Cameron, J., Pierce, W. D., Banko, K. M., & Gear, A. (2005). Achievement-based rewards and intrinsic motivation: A test of cognitive mediators. *Journal of Educational Psychology*, 97(4), 641-655.
- Carifo, J., & Carey, T. (2009). A Critical examination of current minimum grading policy recommendations. *The High School Journal*, 93(1), 23-37.
- Carlson, L. A. (2003). *Beyond assessment to best grading practice: Practical guidelines*. Retrieved from ERIC database. (ED480071)
- Chapman, H. B., & Ashbaugh, E. J. (1925). Report cards in American cities. *Educational Research Bulletin*, 4(14), 289-293.
- Cherniss, A. (2008). *Standards-based report card: Teachers' perception on the development, transition, and implementation* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3311106)
- Cherry, K. (n.d.). What is extrinsic motivation? Retrieved from <http://psychology.about.com/od/eindex/f/extrinsic-motivation.htm>
- Ciani, K. D., & Sheldon, K. M. (2010). A versus F: The effects of implicit letter priming on cognitive performance. *British Journal of Educational Psychology*, 80, 99-119.
- Cohen, B. H., & Lea, R. B. (2004). *Essential of statistics for the social and behavioral sciences*. Hoboken, NJ: Wiley.
- Cohen, J. (1988). *Statistical power analysis for the behavior sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Colvin, S. S. (1912). Marks and the marking system as an incentive to study. In J. A. Laska & T. Juarez (Eds.), *Grading and marking in American schools* (pp. 85-92). Springfield, IL: Charles C Thomas.
- Cooney, T., Bell, K., & Fisher-Cauble, D. (1996). The demands of alternate assessment: What teachers say. *Mathematics Teacher*, 89, 484-488.

- Craig, T. A. (2011). *Effects of standards-based report cards on student learning* (Doctoral dissertation). Available from ProQuest Dissertation and Theses database. (UMI No. 3498282).
- Crocker, J., Karpinski, A., Quinn, D. M., & Chase, S. K. (2003). When grades determine self-worth for male and female engineering and psychology majors. *Journal of Personality and Social Psychology*, 85(3), 507-516.
- Cross, L. H., & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurements in Education*, 12(1), 53-72.
- Curreton, L. W. (1971). The history of grading practices. *NCME: Measurement News*, 2(4), 1-8.
- Davis, J. D. (1931). The Effect of the 6-22-44-22-6 normal curve system on failures and grade values. *The Journal of Educational Psychology*, 22(8), 636-640.
- De Charms, R. (1968). *Personal Causation*. New York, NY: Academic Press Inc.
- Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 18, 105-115.
- Deci, E. L. (1972a). Intrinsic motivation, extrinsic reinforcement, and inequity. *Journal of Personality and Social Psychology*, 22, 113-120.
- Deci, E. L. (1972b). The effects of contingent and noncontingent rewards and controls on intrinsic motivation. *Organizational Behavior and Human Performance*, 8, 217-229.
- Deci, E. L. (1975). *Intrinsic motivation*. New York, NY: Plenum.
- Deci, E. L., Koestner, R. & Ryan, R. M. (1999a). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6), 627-668.
- Deci, E. L., Koestner, R. & Ryan, R. M. (1999b). The undermining effect is a reality after all – extrinsic rewards, task interest, and self-determination, reply to Eisenberger, Pierce, and Cameron (1999) and Lepper, Henderlong, and Gingras (1999). *Psychological Bulletin*, 125(6), 692-700.
- Deci, E. L., Koestner, R. & Ryan, R. M. (2001a). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research*, 71(1), 1-27.
- Deci, E. L., Koestner, R. & Ryan, R. M. (2001b). The pervasive negative effects of rewards on intrinsic motivation: Response to Cameron. *Review of Educational Research*, 71(1), 43-51.

- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.
- Deci, E. L., & Ryan, R. M. (1987). The support of autonomy and the control of behavior. *Journal of Personality and Social Psychology*, 53, 1024-1037.
- Denton, J. J., & Henson, K. T. (1979). Mastery learning and grade inflation. *Educational Leadership*, 37(2), 150-152.
- Dickinson, A. M. (1989). The detrimental effects of extrinsic reinforcement on “Intrinsic Motivation”. *The Behavior Analyst*, 12(1), 1-15.
- Durm, M. W. (1993). An ‘A’ is not an ‘A’ is not an ‘A’: A history of grading. *The Educational Forum*, 57(3), 294-297.
- Ebel, R. L. (1980). The failure of schools without failure. In J. A. Laska & T. Juarez (Eds.), *Grading and marking in American schools* (pp. 105-112). Springfield, IL: Charles C Thomas.
- Ebel, R. L., & Frisbie, D. A. (1986). *Essentials of educational measurement* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Edwards, N., & Richey, H. G. (1947). *The school in the American social order*. Cambridge, MA: Houghton-Mifflin.
- Eisenberger, R., Pierce, W. D., Cameron, J. (1999). Effects of reward on intrinsic motivation – negative, neutral, and positive: Comment on Deci, Koestner, and Ryan (1999). *American Psychology Association*, 125(6), 677-691.
- Evans, F. B. (1976). What research says about grading. In S. Simon & A. Bellanca (Eds.), *Ungrading the grading myths: A primer of alternatives to grades and marks* (pp. 30-50). Alexandria, VA: Association for Supervision and Curriculum Development.
- Friedman, S. J., & Frisbie, D. A. (2000). Making report cards measure up. *Education Digest*, 65(5), 45-50.
- Frisbie, D. A., & Waltman, K. K. (1992). Developing a personal grading plan. *Educational Measurement: Issues and Practice*, 11(3), 35-42.
- Georgia Department of Education. (2013, September). *An assessment & accountability brief: 2013 CRCT validity and reliability*.
- Glaser, W. (1971). Reaching the unmotivated. *The Science Teacher*, 38(3), 18-22.
- Gottfried, A. E. (1990). Academic intrinsic motivation in young elementary school children. *Journal of Educational Psychology*, 82(3), 525 – 538.

- Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: a longitudinal study. *Journal of Educational Psychology, 93*(1), 3-13.
- Grause, E. (2011). Are we paving paradise? *Educational Leadership, 66*(7), 12-17.
- Gronlund, N. (2006). *Assessment of student achievement* (8th ed.). Boston: Pearson Education, Inc.
- Guay, F., Chanal, J., Ratelle, C., Marsh, H., Larose, S., & Bovin, M. (2010). Intrinsic, identified, and controlled types of motivation for school subjects in young elementary school children. *British Journal of Educational Psychology, 80*, 711-735.
- Guskey, T. R. (1994). Making the grade: What benefits students? *Educational Leadership, 52*(2), 14-20.
- Guskey, T. R. (2001). Helping students make the grade. *Educational Leadership, 59*(1), 20-27.
- Guskey, T. R. (2002). *How's my kid doing? A parent's guide to grades, marks, and report cards*. San Francisco, CA: Josey-Bass.
- Guskey, T. R. (2004). The communication challenge of standards-based reporting. *Phi Delta Kappan, 86*(1), 326-329.
- Guskey, T. R. (2011). Five obstacles to grading reform. *Educational Leadership, 69*(3), 16-21.
- Guskey, T. R. (2013a). The case against percentage grades. *Educational Leadership, 71*(1), 68-72.
- Guskey, T. R. (2013b). In search of a useful definition of mastery. *Educational Leadership, 71*(4), 18-23.
- Guskey, T. R., & Bailey, J. M. (2001). *Developing grading and reporting systems for student learning*. Thousand Oaks, CA: Corwin Press, Inc.
- Guskey, T. R., & Jung, L. (2012). Four steps in grading reform. *Principal Leadership, 13*(4), 22-28.
- Gutek, G. L. (1986). *Education in the United States: An historical perspective*. Englewood Cliffs, NJ: Prentice Hall.
- Haladyna, T. M. (1999). *A complete guide to student grading*. Needham Heights, MS: Allyn & Bacon.

- Haptonstall, K. G. (2010). *An analysis of the correlation between standards-based, non-standards-based grading systems and achievement as measured by the Colorado Student Assessment Program (CSAP)* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3397987)
- Hayenga, A. O., & Corpus, J. H. (2010). Profiles of intrinsic and extrinsic motivations: A person-centered approach to motivation and achievement in middle school. *Motivation & Emotion, 34*(4), 371-383.
- Hills, J. R. (1981). *Measurement and evaluation in the classroom* (2nd ed.). Columbus, OH: Charles E. Merrill Publishing Co.
- Huitt, W. (2001). Motivation to learn: An overview. *Educational Psychology Interactive*. Valdosta, GA: Valdosta State University. Retrieved from: <http://www.edpsycinteractive.org/topics/motivation/motivate.html>
- Jung, L., & Guskey, T. (2010). Grading exceptional learners. *Educational Leadership, 67*(5), 31-35.
- Kirschenbaum, H., Simon, S. B., & Napier, R. W. (1971). *Wad-ja-get? The grading game in American education*. New York: Hart Publishing Co.
- Kohn, A. (2011). The case against grades. *Educational Leadership, 69*(3), 28-33.
- Krumboltz, J. D., & Yeh, C. J. (1996). Competitive grading sabotages good teaching. *Phi Delta Kappan, 78*(4), 324-326.
- Lake, K., & Kafka, K. (1996). Reporting methods in grade K-8. In T. R. Guskey (Ed.), *Communicating student learning. 1996 Yearbook of the Association for Supervision and Curriculum Development* (pp. 90-188). Alexandria, VA: Association for Supervision and Curriculum Development.
- Laska, J. A., & Juarez, T. (1992). *Grading and marking in American schools*. Springfield, IL: Charles C Thomas.
- Lepper, M. R., Corpus, J. H., & Iyengar, S. S. (2005). Intrinsic and extrinsic motivational orientations in the classroom: Age differences and academic correlates. *Journal of Educational Psychology, 97*(2), 184-196.
- Lomax, R. G. (2007). *An introduction to statistical concepts* (2nd ed.). New York, NY: Psychology Press.
- Malone, B. G., Nelson, J. S., & Van Nelson, C. (2002). A study of the effect of the implementation of the plus/minus grading system on graduate student grades. *College and University, 77*(3), 11-22.

- Manzo, K. K. (2001). Districts tinker with report cards to make better sense of standards. *Education Week*, 21(4), 1-3.
- Marzano, R. (1998). *Models of standards implementation: Implication for the classroom*. Aurora, CO: Mid-continent Regional Educational Laboratory.
- Marzano, R. (2000). *Transforming classroom grading*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Marzano, R. J., & Heflebower, T. (2011). Grades that show what students know. *Educational Leadership*, 69(3), 34-39.
- Marzano, R. J., & Kendall, J. S. (1998). *Implementing standards-based education*. Washington, DC: National Education Association.
- Mathura, V. G. (2008). *A changing paradigm: A case study of a district's shift in communicating with standards-based report cards* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3324729)
- McHugh, M. L. (2013). The Chi-square test of independence. *Biochemia Medica*, 23(2), 143-149.
- Meyer, M. (1908). The grading of students. *Science*, 28(712), 243-250.
- Middleton, W. (1933). Some general trends in grading procedure. *Education*, 54(1), 5-10.
- Morris, L. (1952). Evaluating and reporting pupil progress. *The Elementary School Journal*, 53(3), 144-149.
- Munk, D. D., & Bursuck, W. D. (2001). What report cards should and do communicate: Perceptions of parents of secondary students with and without disabilities. *Remedial and Special Education*, 22(5), 280-287.
- Napoli, A. R., & Raymond, L. A. (2004). How reliable are our assessment data?: A comparison of the reliability of data produced in graded and ungraded conditions. *Research in Higher Education*, 45(8), 921- 929.
- Natriello, G., & Dornsbush, S. M. (1984). *Teacher evaluative standards and student effort*. New York, NY: Longman.
- National Education Goals Panel. (1991). *The National Education Goals Report: Building a nation of learners*. Washington, DC: Author.
- O'Connor, K. (2009). *How to grade for learning: Linking grades to standards*. Thousand Oaks, CA: Corwin.

- O'Connor, K. (2010). *A repair kit for grading: 15 fixes for broken grades*. Boston: Pearson.
- O'Connor, K., & Wormeli, R. (2011). Reporting student learning. *Educational Leadership*, 69(3), 40-44.
- Olson, K. E. (2005). *Standards-based report cards: Accountability at the classroom level* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3196865)
- Paepflow, C. G. (2011). *Easy as 1, 2, 3: Exploring the implementation of standards-based grading in Wake County elementary school* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3497209)
- Page, E. B. (1958). Teacher comments and student performance: A seventy-four classroom experiment in school motivation. *The Journal of Educational Psychology*, 49(4), 173-181.
- Panchisin, D. L. (2004). *A case study of parents' and teachers' perceptions of the usefulness of a newly developed elementary school report card in the Appoquinimink School District* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3156271)
- Pierce, W. D., Cameron, J., Banko, K. M., & So, S. (2003). Positive effects of rewards and performance standards on intrinsic motivation. *The Psychological Record*, 53, 561-579.
- Pilcher, J. K. (1994). The value-driven meaning of grades. *Educational Assessment*, 2(1), 69-88.
- Pulfrey, C., Buschs, C. & Butera, F. (2011). Why grades engender performance-avoidance goals: The mediating role of autonomous motivation. *Journal of Educational Psychology*, 103(3), 683-700.
- Pulfrey, C., Darnon, C. & Butera, F. (2013). Autonomy and task performance: Explaining the impact of grades on intrinsic motivation. *Journal of Educational Psychology*, 105(1), 39-57.
- Reeves, D. B. (2004). The case against the zero. *Phi Delta Kappan*, 86(4), 324-325.
- Reeves, D. B. (2010). *Elements of grading*. Bloomington, IN: Solution Tree.
- Resh, N. (2009). Justice in grades allocation: teacher's perspective. *Social Psychology in Education*, 12(3), 315-325.
- Rogers, K. (2013, November 21). Common core divides private schools. Fox Business. Retrieved from <http://www.foxbusiness.com/personal->

finance/2013/11/21/common-core-divides-private-schools-will-it-help-or-hurt-business/

- Rugg, H. O. (1918). Teachers' marks and the reconstruction of the marking system. *The Elementary School Journal*, 18(9), 701-719.
- Rupp, A. A., Lesaux, N. K., & Siegel, L. S. (2006). Meeting expectations? An empirical investigation of a standards-based assessment of reading comprehension. *Educational Evaluation and Policy Analysis*, 28(4), 315-333.
- Siegel, S., & Castellan, Jr., N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). Columbus, OH: McGraw Hill.
- Simon, S. (1972). Whadjaget? *Time*, 100(22), 61.
- Shim, S., & Ryan, A. (2005). Changes in self-efficacy, challenge avoidance, and intrinsic value in response to grades: The role of achievement goals. *The Journal of Experimental Education*, 73(4), 2005.
- Starch, D. (1913). Reliability and distribution of grades. *Science*, 38(983), 630-636.
- Starch, D., & Elliott, E. C. (1912). Reliability of the grading of high-school work in English. *The School Review*, 20(7), 442-457.
- Starch, D., & Elliott, E. C. (1913a). Reliability of grading work in mathematics. *The School Review*, 21(4), 254-259.
- Starch, D., & Elliott, E. C. (1913b). Reliability of grading work in history. *The School Review*, 21(10), 676-681.
- Stephan, Y., Caudroit, J., Boiche, J., & Sarrazin, P. (2011) Predictors of situational disengagement in the academic setting: The contribution of grades, perceived competence, and academic motivation. *British Journal of Educational Psychology*, 81, 441-455.
- Stiggins, R. J. (2001). *Student involved classroom assessment* (3rd ed.). Upper Saddle River, NJ: Prentice Hall, Inc.
- Stiggins, R. J., Frisbie, D. A., & Griswold, P.A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice*, 8(2), 5-14.
- Terwilliger, J. S. (1977). Assigning grades - philosophical issues and practical recommendations. *Journal of Research and Development in Education*, 10(3), 21-39.

- The National Commission on Excellence in Education (1983). *A nation at risk: The imperative for educational reform*. Retrieved November 23, 2012 from http://datacenter.spps.org/uploads/SOTW_A_Nation_at_Risk_1983.pdf.
- Tocci, C. (2008). An immanent machine: Reconsidering grades, historical and present. *Educational Philosophy and Theory*, 42(7), 762-778.
- Tomlinson, C., & McTighe, J. (2006). *Integrating differentiated instruction and understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Tuten, J. (2007). "There's two sides to every story": How parents negotiate report card discourse. *Language Arts*, 84(4), 314-324.
- Tyler, R. W. (1935). Evaluation: A challenge to progressive education. *Educational Research Bulletin*, 14(1), 9-16.
- Waltman, K. K., & Frisbie, D. A. (1994). Parents' understanding of their children's report card grades. *Applied Measurement in Education*, 7(3), 223-240.
- What do my child's test scores mean?* (n.d.). Retrieved from http://www.barrow.k12.ga.us/Testing/What_Do_My_Child.pdf
- Whipple, G. M. (1913). Editor's Preface. In I. E. Finkelstein, *The marking system in theory and practice* (pp. 1-2). Baltimore, MD: Warwick & York, Inc.
- Wiggins, G. (1996). Honesty and fairness: Toward better grading and reporting. In T. R. Guskey (Ed.), *Communicating student learning. 1996 Yearbook of the Association for Supervision and Curriculum Development* (pp. 90-188). Alexandria, VA: Association for Supervision and Curriculum Development.
- Winger, T. (2005). Grading to communicate. *Educational Leadership*, 63(3), 61-65.
- Workman, E. A., & Williams, R. L. (1980). Effects of extrinsic rewards on intrinsic motivation in the classroom. *Journal of School Psychology*, 18(2), 141-147.
- Wormelli, R. (2006). *Fair isn't always equal: Assessing and grading in the differentiated classroom*. Portland, ME: Stenhouse Publishers.
- Wright, D., & Wiese, M. J. (1988). Teacher judgment in student evaluation: A comparison of grading methods. *Journal of Educational Research*, 82(1), 10-14.
- Wrinkle, W. L. (1935). School marks – why, what, and how? In J. A. Laska & T. Juarez (Eds.), *Grading and marking in American schools* (pp. 93-100). Springfield, IL: Charles C Thomas.

Wrinkle, W. L. (1947). *Improving marking and reporting practices in elementary and secondary schools*. New York: Holt, Rinehart, and Winston.

APPENDIX A

COBB COUNTY THIRD GRADE REPORT CARD



Third Grade Report Card

Student: _____
Teacher: _____

School Year : _____
School: _____

PERFORMANCE INDICATORS
The purpose of this report card is to communicate to students, parents, and staff the progress each student is making toward accomplishing performance-based standards.

- 3+ = Exceeds Standards - In addition to the 3, makes applications and inferences beyond expectations
- 3 = Meets Standards - Consistently and independently
- 2 = Progressing toward meeting standards
- 1 = Limited progress or does not meet standards

or NA = Not assessed at this time

LANGUAGE ARTS	Q1	Q2	Q3	Q4
READING				
Foundational Skills				
Phonics and Word Recognition				
Fluency to Support Comprehension				
Reasons and Proof Reading Level	A B C D E F G H J K L M N O P			
	K 1st 2nd 3rd			
Library and Informational Text				
Key Ideas and Details				
Craft and Structure of Text				
Integration of Knowledge and Ideas				
Range of Reading and Level of Text Complexity				
WRITING				
Text Type and Purpose				
Narrative				
Research to Build and Present Knowledge				
LANGUAGE				
Conventions of Standard English				
Vocabulary Acquisition and Use				
SPEAKING AND LISTENING				
Participation and Collaboration				
Presentation of Knowledge				
SOCIAL STUDIES				
Explains the political roots and basic principles of our modern democracy				
Discusses Americans who expanded people's rights and freedoms				
Locates major geographical features of the United States				
Describes two four types of productive resources				

MATHEMATICS				
NUMBERS AND ALGEBRAIC THINKING				
Multiplication and division strategies				
Reason and solve problems				
MEASUREMENT				
Concepts of area				
FRACTIONS				
Understand fractions as numbers				
GEOMETRY				
Analyze two-dimensional shapes				
SCIENCE				
Investigate Earth, Life, and Physical Science Concepts				
Apply and transfer scientific knowledge				
Utilize scientific inquiry				
HEALTH				
Comprehends concepts of health promotion & disease prevention				
PHYSICAL EDUCATION				
Demonstrates motor skills and movement patterns				
Learning Skills & Behaviors				
ART				
Creates art based on observation, using the elements and principles of design				
Analyzes and discusses functions of art in historic periods and cultures				
Learning Skills & Behaviors				
MUSIC				
Applies knowledge of music concepts through singing and playing instruments				
Reads and notates music				
Describe and analyze music				
Learning Skills & Behaviors				
TECHNOLOGY				
Comprehends and applies third grade concepts and skills related to technology				
ATTENDANCE				
Absences				
Tardies				
PLACEMENT INFORMATION				
<input type="checkbox"/> Promoted to 4th grade <input type="checkbox"/> Placed in 4th grade				
<input type="checkbox"/> Additional Year in 3rd Grade				

LEARNING SKILLS & BEHAVIORS	Q1	Q2	Q3	Q4
S = SUCCESSFUL - Student displays appropriate learning skills and behaviors				
P = PROGRESSING - Student is progressing; displays appropriate learning skills and behaviors most of the time				
N = NEEDS IMPROVEMENT - Student does not display learning skills and behaviors that lead to success				
LEARNING SKILLS				
1- Organizes seat and materials				
2- Works independently				
3- Asks questions/Seeks help when needed				
4- Completes tasks in a timely manner				
LEARNING BEHAVIORS				
5- Uses seat control				
6- Exhibits good listening skills				
7- Accepts responsibility for behavior				
8- Works cooperatively with others				
9- Controls talking				
RULES AND PROCEDURES				
10- Follows directions				
11- Shows respect for peers				
12- Shows respect for authority				
13- Shows respect for property				

APPENDIX B

MUSCOGEE COUNTY THIRD GRADE REPORT CARD

Student:
Teacher:

G.T.I.D. #
School:

Key to Academic Achievement and Special Areas Standards	Standards	Performance
3	2	1
Exceeds and Exceeds Expects	Exceeds Expects	Meets Expects
Meets Expects	Meets Expects	Meets Expects
Does not meet Standard	Does not meet Standard	Does not meet Standard

Language Arts Standards	Q1	Q2	Q3	Q4
Reading				
Recognizes and understands vocabulary in context				
Comprehends the elements of various genres				
Recognizes author's purpose				
Demonstrates comprehension of grade-level text				
Identifies the main idea and supporting details				
Distinguishes fact from opinion				
Analyses story elements from various texts				
Identifies the cause and effect				
Reads fluently and applies oral reading strategies				
Writing/grammar	Q1	Q2	Q3	Q4
Writes legibly in production				
Produces writing that has a clear focus				
Organizes writing				
Uses style of writing appropriate for audience				
Uses conventions of capitalization and punctuation				
Identifies parts of speech and their purposes				
Applies the rules of sentence structure				
Applies common rules of spelling				
Research	Q1	Q2	Q3	Q4
Develops research skills				
Absences	Q1	Q2	Q3	Q4
Tardies				

Mathematics Standards	Q1	Q2	Q3	Q4
Number and Operations				
Understands and uses their value				
Solves addition and subtraction problems				
Knows multiplication facts				
Solves problems using multiplication				
Solves problems using division				
Understands fractions				
Understands and compares decimals				
Applies concepts of money				
Estimates to solve problems				
Algebra	Q1	Q2	Q3	Q4
Uses symbols and function rules				
Recognizes, describes, and extends patterns				
Geometry	Q1	Q2	Q3	Q4
Classifies, draws, and explains geometric figures and their properties				
Identifies parts of a circle				
Classifies triangles				
Measurement	Q1	Q2	Q3	Q4
Understands and measures elapsed time				
Measures, estimates, and compares length using appropriate units				
Understands and measures perimeter				
Understands and measures area				
Data Analysis and Probability	Q1	Q2	Q3	Q4
Understands, organizes, displays, and interprets graphs and tables				
Process Skills	Q1	Q2	Q3	Q4
Solves word problems				
Uses models to represent, communicate, and connect mathematics in multiple ways				
Student Assessed Using:	Q1	Q2	Q3	Q4
Modification				
Accommodations				
Student Served By:	Q1	Q2	Q3	Q4
IEP				
ESOL				

Social Studies Standards	Q1	Q2	Q3	Q4
Historical Understanding				
Identifies political roots of our modern democracy				
Describes historical figures	Q1	Q2	Q3	Q4
Geographic Understanding				
Locates topographic features				
Identifies cultural and geographical systems for historical figures				
Civic Understanding	Q1	Q2	Q3	Q4
Describes positive character traits				
Explains foundations of a republican form of government				
Economic Understanding	Q1	Q2	Q3	Q4
Explains government provided goods and services				
Explains interdependence and trade				
Describes producer resources				
Describes personal spending and saving choices				
Science Standards	Q1	Q2	Q3	Q4
Name of Science				
Uses scientific method and process				
Physical Science	Q1	Q2	Q3	Q4
Demonstrates knowledge of their energy				
Demonstrates knowledge of magnets				
Earth Science	Q1	Q2	Q3	Q4
Explains and recognizes the differences of rocks and minerals				
Compares the similarities and differences of soils				
Demonstrates knowledge of fossils				
Life Science	Q1	Q2	Q3	Q4
Demonstrates knowledge of habitats				
Recognizes the effects of pollution				

Key to Initiative and Work Standards	Q1	Q2	Q3	Q4
1 - SUCCESSFUL				
2 - PROGRESSING				
3 - NEEDS SUPPORT				
4 - INITIATIVE				
5 - INITIATIVE				
6 - INITIATIVE				
7 - INITIATIVE				
8 - INITIATIVE				
9 - INITIATIVE				
10 - INITIATIVE				
11 - INITIATIVE				
12 - INITIATIVE				
13 - INITIATIVE				
14 - INITIATIVE				
15 - INITIATIVE				
16 - INITIATIVE				
17 - INITIATIVE				
18 - INITIATIVE				
19 - INITIATIVE				
20 - INITIATIVE				
21 - INITIATIVE				
22 - INITIATIVE				
23 - INITIATIVE				
24 - INITIATIVE				
25 - INITIATIVE				
26 - INITIATIVE				
27 - INITIATIVE				
28 - INITIATIVE				
29 - INITIATIVE				
30 - INITIATIVE				
31 - INITIATIVE				
32 - INITIATIVE				
33 - INITIATIVE				
34 - INITIATIVE				
35 - INITIATIVE				
36 - INITIATIVE				
37 - INITIATIVE				
38 - INITIATIVE				
39 - INITIATIVE				
40 - INITIATIVE				
41 - INITIATIVE				
42 - INITIATIVE				
43 - INITIATIVE				
44 - INITIATIVE				
45 - INITIATIVE				
46 - INITIATIVE				
47 - INITIATIVE				
48 - INITIATIVE				
49 - INITIATIVE				
50 - INITIATIVE				
51 - INITIATIVE				
52 - INITIATIVE				
53 - INITIATIVE				
54 - INITIATIVE				
55 - INITIATIVE				
56 - INITIATIVE				
57 - INITIATIVE				
58 - INITIATIVE				
59 - INITIATIVE				
60 - INITIATIVE				
61 - INITIATIVE				
62 - INITIATIVE				
63 - INITIATIVE				
64 - INITIATIVE				
65 - INITIATIVE				
66 - INITIATIVE				
67 - INITIATIVE				
68 - INITIATIVE				
69 - INITIATIVE				
70 - INITIATIVE				
71 - INITIATIVE				
72 - INITIATIVE				
73 - INITIATIVE				
74 - INITIATIVE				
75 - INITIATIVE				
76 - INITIATIVE				
77 - INITIATIVE				
78 - INITIATIVE				
79 - INITIATIVE				
80 - INITIATIVE				
81 - INITIATIVE				
82 - INITIATIVE				
83 - INITIATIVE				
84 - INITIATIVE				
85 - INITIATIVE				
86 - INITIATIVE				
87 - INITIATIVE				
88 - INITIATIVE				
89 - INITIATIVE				
90 - INITIATIVE				
91 - INITIATIVE				
92 - INITIATIVE				
93 - INITIATIVE				
94 - INITIATIVE				
95 - INITIATIVE				
96 - INITIATIVE				
97 - INITIATIVE				
98 - INITIATIVE				
99 - INITIATIVE				
100 - INITIATIVE				

Placement Information	Q1	Q2	Q3	Q4
Placed in 4th grade				
Placed in 3rd grade				