11-9-2005

Factors affecting the alignment of grades and reading scores for third grade students on the Florida Comprehensive Assessment Test

Kristine L. Dittmar
Florida International University

DOI: 10.25148/etd.FI14062299

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

FACTORS AFFECTING THE ALIGNMENT OF GRADES AND READING SCORES FOR THIRD GRADE STUDENTS ON THE FLORIDA COMPREHENSIVE ASSESSMENT TEST

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF EDUCATION

in

EDUCATIONAL ADMINISTRATION AND SUPERVISION

by

Kristine L. Dittmar

2005

To: Dean Linda Blanton
College of Education

This dissertation, written by Kristine L. Dittmar, and entitled Factors Affecting the Alignment of Grades and Reading Scores for Third Grade Students on the Florida Comprehensive Assessment Test, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Anthony H. Normore

Sharon W. Kossack

Paulette Johnson

Peter J. Cistone, Major Professor

Date of Defense: November 9, 2005

The dissertation of Kristine L. Dittmar is approved.

Dean Linda Blanton
College of Education

Dean Douglas Wartzok
University Graduate School

Florida International University, 2005

DEDICATION

I dedicate this dissertation to my husband, David for his continual devotion, patience, understanding, and love throughout this process. Through his support the completion of this final product became a reality. To my daughters, Erika and Sara, their love and understanding was constant, even when they would have preferred spending more quality time together with me. Additionally, this dissertation is dedicated to my parents, Carlisle and Marie, who instilled in me the love of learning and the ability to achieve my goals.

ACKNOWLEDGMENTS

It is with gratitude and appreciation that I thank the members of my committee. Dr. Peter J. Cistone, my major professor, guided me through every phase on the road to earning this degree. Not only was he the "Godfather" of our Doctoral Program Cohort group, but he has been there for me through my collegiate years at Florida International University. Dr. Anthony H. Normore was highly supportive while guiding my way. I also thank Dr. Sharon W. Kossack for her enduring friendship and knowledge, which has been my inspiration in the area of reading and writing. She offered her expertise in developing the teacher questionnaire and facilitating the process. Dr. Paulette Johnson was particularly instrumental in guiding me through the quantitative methodology and giving me the confidence in my abilities. She is the one that got me through the rough road on this journey.

I have found working in a Cohort group in the Educational Leadership and Supervision program to be stimulating and thoughtful, providing me with new tools to apply to my professional career.

ABSTRACT OF THE DISSERTATION

FACTORS AFFECTING THE ALIGNMENT OF GRADES AND READING SCORES FOR THIRD GRADE STUDENTS ON THE FLORIDA COMPREHENSIVE ASSESSMENT TEST

by

Kristine L. Dittmar

Florida International University, 2005

Miami, Florida

Professor Peter J. Cistone, Major Professor

The purpose of this study was to aid in understanding the relationship between current Reading report card grading practices and standards-based state standardized testing results in Reading and factors associated with the alignment of this relationship. Report card and Florida Comprehensive Assessment Test (FCAT) data for 2004 were collected for 1064 third grade students in nine schools of one feeder pattern in Florida's Miami-Dade County Public Schools. A *Third Grade Teacher Questionnaire* was administered to 48 Reading teachers. The questionnaire contained items relating to teachers' education, teaching experience, grading practices, and beliefs about the FCAT, instructional Reading activities, methods, and materials.

Findings of this study support a strong relationship between report card grades and FCAT Reading achievement levels. However, individual school correlational analysis showed significant differences among schools' alignment measures. Higher teacher alignment between grades and FCAT levels was associated with teachers spending more time on individualized methods of Reading instruction and to teachers

feeling there was not enough time to teach and help individual students. Lower teacher alignment of grades and achievement levels was associated with teachers taking homework into account in the final Reading grade. Teacher alignment of grades and achievement levels was not associated with teacher beliefs concerning the FCAT, instructional activities in Reading and Language Arts, the Reading program used, the model of delivery of the Reading program, instruction or type of instructional planning done by the teachers.

This study highlights the need for further investigations related to determining additional teacher factors that may affect the alignment relationship between report card grades and standards-based state standardized testing results.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Report card grades have been used as a method of communicating student progress in the American educational system since the 1800s. While the actual methods and forms that report cards use can vary, the concept of reporting student progress has not changed significantly over the years (Marzano, 2000; Smith, 1999; Trumbull & Farr, 2000). Another aspect of the educational system that evolved along with report cards is assessing and evaluating students' learning. This evaluation takes on many forms, yet all fifty states have some type of mandatory standardized testing to assess legislated curriculum standards that are required for all students to learn. This testing has become an instrument of accountability for legislators, educators, parents and the community by measuring student achievement and progress towards meeting the required mandated state standards (Heubert, Hauser & Committee on Appropriate Test Use, 1999; Johnson, 2001; Trumbull & Farr, 2000).

As a result of this accountability, educators are continually being challenged to align the curriculum, instruction, and assessment with state standards. However, educators are still required to report student achievement and progress in the form of report cards (Johnson, 2001; Marzano, 2000). Consequently, it is a general practice for the State Department of Education to send results of a student's standardized testing to inform parents of their child's progress towards achieving the mandated standards. Schools in turn distribute these results to parents (Heubert, Hauser & Committee on Appropriate Test Use, 1999). Additionally, the school sends home report cards that assess many of the same standards by a wide variety of methods, generally not coordinated with

the state testing. As a result, the two assessment systems do not always agree. Herein lies the problem that has led to the investigation undertaken by this study.

Statement of Problem

When analyzing current report card grading practices and their relationship to standards-based state standardized (SBSS) testing, parents, students, and educators have reason to expect that students demonstrating high achievement levels on SBSS tests would have high grades on their report cards. Conversely, there is also an expectation that students demonstrating low achievement levels on SBSS tests would have low grades on their report cards. It is reasonable to suggest that if teachers are teaching the state standards, then the results of instruction guided by the state standards are what teachers use to determine report card grades. As a result, report card grades should demonstrate a relationship with SBSS achievement test levels. This, however, is not always the case (Brennan, Kim, Wenz-Gross & Siperstein, 2001). In some instances, parents find that children who have above average report card grades score in the lowest achievement level, while others find that children with below average report card grades score in the highest achievement level (Marzano, 2000; Trumbull & Farr, 2000).

Reading is the foundation for all subject areas and is considered a critical indicator of success in school (Anderson, Hiebert, Scott & Wilkinson, 1985; Leinhardt, Zigmond & Cooley, 1981; Tankersley, 2003). In elementary schools, students are continually assessed in Reading to determine their progress and achievement, and ultimately to receive a report card grade. SBSS tests also assess Reading as an indicator of academic achievement. These tests are intended to objectively evaluate the state standards and skills in Reading that each classroom teacher is required to teach. In

comparison, report card grades also use the outcome of the same Reading standards and skills, but these are often not the sole criteria in determining the grade. As a result, Reading report card grades do not always agree with the SBSS test results, and can ultimately contribute to a low alignment between Reading card grades and achievement levels.

Since discrepancies can be found between Reading achievement levels and Reading report card grades, questions arise about the factors that contribute significantly to these relationships. It is conceivable that a single factor may emerge as the cause, but it seems more likely that multiple factors are the case. Furthermore, factors may even vary among schools and within the district depending upon the instructional methods and materials used, grading criteria used, and teacher beliefs. Also, teacher belief systems concerning the SBSS testing and accountability measures may impact these assessment relationships.

Statement of the Purpose

The primary purpose of this research study was to examine the relationship between current Reading report card grading practices and SBSS testing results in Reading. To accomplish this, final Reading report card grades were correlated to Florida Comprehensive Assessment Test (FCAT) Reading achievement levels for third grade students. A second purpose was to ascertain which factors are associated with the alignment of these assessment relationships. Some of the factors investigated included teacher beliefs about teaching and the FCAT, report card beliefs and grading practices, instructional activities used, how the Reading program was planned and delivered, and the Reading methods that were used.

Research Questions and Hypotheses

There were two central research questions and hypotheses that were investigated:

Research Question 1

Do discrepancies exist between Final Reading Report Card Grades and Florida Comprehensive Achievement Test (FCAT) Reading Achievement Levels for third grade students?

Research Question 2

What factors contribute significantly to the relationship between third grade Final Reading Report Card Grades and FCAT Reading Achievement Levels due to membership in a particular teacher's class?

Hypothesis 1

Third grade students with a Final Reading Report Card Grade of "A" or "B" will attain a FCAT Reading Achievement Level of 4 or 5, and those with a Final Reading Report Card Grade of "D" or "F" will attain a FCAT Reading Achievement Level of 1 or 2.

Hypothesis 2

Differences in teacher beliefs about teaching and the FCAT, report card beliefs and grading practices, instructional activities, Reading program planning and delivery and Reading methods contribute to the alignment between Final Reading Report Card Grades and FCAT Reading Achievement Levels.

Significance of the Study

This study is significant for two reasons. First, if both report card grades and state standardized tests are indicators of achievement, then these measures should be consistent and aligned with each other. If a very high positive correlation (Davis, 1971) exists

between these two factors, Reading report card grades could be considered indicators of expected achievement levels on state standardized tests in Reading. However, if a low correlation (Davis, 1971) is found, this would be a reason for further in depth investigations to better understand this relationship. Secondly, determining the correlation of teacher's beliefs about teaching, the FCAT, report cards, as well as the grading practices and instructional activities they use, how they plan and deliver the Reading program, and the Reading methods that they use will allow for analysis of what factors contribute to the either high or low alignment of final Reading report card grades and FCAT Reading achievement levels.

Delimitations

The sample for this research was 1064 students and 48 teachers. This represented the third grade population of nine schools in one feeder pattern in Miami-Dade County Public Schools, a large urban school district in the State of Florida. These schools represented similar populations of students and teachers with the teachers having varying degrees of teaching and educational experience. Teacher survey data were collected in small group settings and contained items concerning teachers' emphasis on specific Reading state standards from the FCAT, the specific Reading materials and methods being used, the amount of time spent on the instruction of specific Reading state standards, what specific standards were graded, teacher grading practices, and the teachers' beliefs about the FCAT and its content.

Constraints of this sample were that only one urban school district in the State of Florida was involved and the school populations used were very similar. However, these schools represented a diverse, multi-ethnic group of students coming from different

socio-economic levels. It can be assumed that a cross-section of the district or other districts in the state may yield results that could be better generalized to the entire population in the State of Florida.

Definitions and Operational Terms

The following term definitions are given in alphabetical in order to clarify their use in this study.

Achievement. What a student learns as a result of instruction.

Alignment. The proper, logical, or expected relation of one thing to another. (Urdang & Flexner, 1968).

Assessment. The process of gathering, describing, or quantifying information about performance (CRESST Assessment Glossary, 1996).

Benchmark. A specific statement of a specific level that describes what students should know and be able to do. Benchmarks are part of the Sunshine State Standards and can be used as checkpoints to monitor progress toward meeting performance goals within and across grade levels (Florida Department of Education, 2004).

Classroom assessment. An assessment developed, administered, and scored by a teacher or set of teachers with the purpose of evaluating individual or classroom student performance on a topic. The results of classroom assessment are ideally used to inform and influence instruction that help students reach higher standards (CRESST Assessment Glossary, 1996).

Correlation. The nature, or extent, of the relationship between two variables (Hinkle, 1998).

Criterion-referenced assessment. An assessment where an individual's performance is compared to a specific learning objective or performance standard and not to the performance of other students. Criterion-referenced assessments tell how well students are performing on specific goals or standards rather than just telling how their performance compares to a norm group of students nationally or locally. In criterion-referenced assessments, it is possible that none or all of the examinees will reach a particular goal or performance standard (CRESST Assessment Glossary, 1996).

Curriculum. A planned course or skills that are deemed necessary for students to master in a particular grade or subject area (Burson, 2001).

Cut Score. Performance standards dividing acceptable levels of readiness from unacceptable levels (Heubert, Hauser, & Committee on Appropriate Test Use, 1999).

Evaluation. When used for most educational settings, evaluation means to measure, compare, and judge the quality of student work, schools, or a specific educational program (CRESST Assessment Glossary, 1996).

Florida Comprehensive Assessment Test (FCAT). The primary purpose of the Florida Comprehensive Assessment Test is to assess student achievement of the high-order cognitive skills represented in the *Sunshine State Standards* (SSS) in Reading, Writing, Mathematics, and Science. The SSS portion of FCAT is a criterion-referenced test. A secondary purpose is to compare the performance of Florida students to the Reading and Mathematics performance of students across the nation using a norm-referenced test (Florida Department of Education, 2004).

Florida Comprehensive Achievement Test (FCAT) Standardized achievement test level. Achievement levels describe the success a student has achieved on the Florida

Sunshine State Standards tested on the FCAT. Achievement levels range from 1 to 5, with Level 1 being the lowest and Level 5 being the highest. The achievement levels are helpful in interpreting what a student's scale score represents (Florida Department of Education, 2004).

Grade. A score, mark, or grade in school on a report card or student product and is a judgment of a student by a teacher. Generally, it is a relative judgment in that depending on the test, the class, or the instructor, the mark may vary (Johnson, 2001).

Grading. Primarily a communication system that informs decisions and actions (Trumbull & Farr, 2000).

High stakes tests. Those tests that have high stakes consequences for students, that is, when an individual student's score determines not just who needs help but whether a student is allowed to take a certain program or class or will be promoted to the next grade (Heubert, Hauser, & Committee on Appropriate Test Use, 1999).

Norm-referenced assessment. An assessment where student performance or performances are compared to a larger group. Usually the larger group or "norm group" is a national sample representing a wide and diverse cross-section of students. Students, schools, districts, and even states are compared or rank-ordered in relation to the norm group. The purpose of norm-referenced assessment test (NRT) is usually to sort students and not to measure achievement towards some criterion of performance (CRESST Assessment Glossary, 1996).

Performance assessment. An assessment that requires students to generate a response to a question rather than choose from a set of responses provided to them. Students would ideally be required to accomplish complex and significant tasks while

bringing to bear prior knowledge, recent learning, and relevant skills to solve realistic or authentic problems (CRESST Assessment Glossary, 1996).

Reliability. The degree to which the results of an assessment are dependable and consistently measure particular student knowledge and/or skills. It may be expressed as: (a) the relationship between test factors intended to measure the same skill or knowledge (item reliability), or (b) the relationship between two administrations of the same test to the same student or students (test/retest reliability) (CRESST Assessment Glossary, 1996).

Report cards. Teachers use multiple elements, a variety of measures and combine them in some way to arrive at a single grade to represent a student's accomplishment of how well a student is doing in a specific area. The grades are given on a report card that is distributed periodically as a means to communicate information about how well children are doing in school (Trumbull & Farr, 2000). Report cards take on a variety of forms across the nation and are the primary method of documenting and reporting student achievement.

Scale score. The score used to report test results on the entire FCAT test. Scale scores on the FCAT Sunshine State Standards tests range from 100 to 500 for each grade level and content area. A computer program is used to analyze student responses and to compute the scale score (Florida Department of Education, 2004).

Standards. The broadest of a family of terms referring to statements of expectations for student learning (CRESST Assessment Glossary, 1996).

Standards-based curricula. A process for adapting instruction, materials, and assessment to make sure all students can achieve the standards (CRESST Assessment Glossary, 1996).

Standards-based student achievement. Standards define the learning that is essential for students' success in schooling (Trumbull & Farr, 2000). Therefore, a student's achievement is measured by specific standards.

State standardized test. A consistent set of procedures for designing, administering, and scoring an assessment and is administered by a state. This type of test assures all students are assessed under the same conditions so that their scores have the same meaning and are not influenced by differing conditions.

Sunshine State Standards (SSS). The *Sunshine State Standards* were approved by the Florida's State Board of Education in 1996 to provide expectations for student achievement in seven subject areas. They identify what students should know and be able to do for the 21st century and include content areas, performance standards and benchmarks. The standards are benchmarked at developmental levels. The Sunshine State Standards provide guidelines for the educational curriculum in the State of Florida (Florida Department of Education, 2004).

Validity. The extent to which an assessment measures what it is supposed to measure and the extent to which inferences and actions made on the basis of test scores are appropriate and accurate. A valid standards-based assessment is aligned with the standards intended to be measured provides an accurate and reliable estimate of students' performance relative to the standard, and is fair. An assessment cannot be valid if it is not reliable (CRESST Assessment Glossary, 1996).

Summary

This chapter introduced this study with a statement of the problem under investigation and the reasons for studying the problem. The research questions and hypotheses are stated as well as the significance for the research study. The study sample is then discussed and delimited. Lastly, the terms essential for the understanding of this investigation are clarified. Chapter II reviews selected literature relating to this research. Chapter III gives a detailed explanation of how the research is conducted, including the procedures. Study results and findings are located in Chapter IV. Finally, Chapter V offers conclusions, discussion and recommendations for future research.

CHAPTER II

REVIEW OF THE LITERATURE

The purpose of this chapter is to present a review of literature that is relevant to this study. These areas of review include: (a) history of school report card grading practices, (b) purpose of report cards and grades, (c) problems and concerns of grading, (d) parent perceptions about grading, (e) growing emphasis on student performance, (f) outcomes of high stakes testing, (g) appropriate test use and alignment, (h) factors affecting test performance, (i) assessments and grades, and (j) report cards and high stakes testing.

History of School Report Card Grading Practices

The origin of report card grades in the United States was traced by Robert Marzano (2000) an educational scholar, back to the 1800s when Harvard University first used a numerical scale to assess students' learning. Prior to that teachers gave feedback to students with narrative comments. Thomas Guskey (1996) an educational researcher, reported that the increasing number of students in the late 1800s, especially at the high school level due to compulsory attendance laws, caused schools to begin grouping students in grade levels according to their age. At the same time, written progress evaluations of students' work began to appear. With the larger more diverse population of students, teachers also started using objective tests. As a result, high schools began using percentages to report student progress. At the elementary level, however, teachers continued to use narrative comments to document student learning.

Research done in the early 1900s concluded that there was a wide variation in teacher grading practices (Trumbull & Farr, 2000). In 1912, Starch and Elliott published

a study questioning the subjectivity and reliability of using percentages as grades (as cited in Johnson, 2001; Smith, 1999). Consequently, teachers turned to grading scales with fewer, broader categories. This was the inception of grades with: Excellent, Average, Poor; or a 5-point scale of: Excellent, Good, Average, Poor, and Failing. The 5-point scale was later translated into A, B, C, D, or F grades (Johnson, 2001). By the 1930s, most schools were using letter grades, but the issue of subjectivity still persisted.

Research available in the 1930s suggested that intelligence within a group approximated a normal probability curve, which later served as a basis for the practice of grading on a curve (Guskey, 1996; Trumbull & Farr, 2000). Grading on a curve meant that a group of students were first rank-ordered and then percentages of the group were assigned to each grade. It was felt that this grading practice was more fair and equitable, and less subjective. Since then, many additional variations of grade reporting have appeared. Some institutions eventually abolished grades while others returned to narrative reporting, the use of pass-fail systems or a mastery approach. The debate over grading practices has continued over the years.

A review of the Educational Resources Information Center (ERIC) System by Guskey (2001) indicated that in the years between 1960 and 2000 there were over 4,000 references to articles and reports on the topic of grading (Johnson, 2001). He also commented that while the topic of grading and reporting practices continues to be fodder for many researchers, a lack of consensus about improvements has made it difficult to change practices. It appears, therefore, that despite ongoing controversy over grading practices in the United States, these practices have nonetheless become a tradition that will be resistant to significant change.

Purpose of Report Cards and Grades

There appears to be a consensus that the primary purpose of report cards is to provide communication links about academic progress between teachers, students, and parents (Friedman & Frisbie, 2000; Howley, Kusimo, & Parrott, 1999; Marzano, 2000; Waltman & Frisbie, 1994). The form this communication took and how it was interpreted by the audience that received it was, however, very divergent and muddled (Waltman & Frisbie, 1994).

Robert Marzano (2000) categorized the use of grades primarily: (a) to give students feedback about their progress and achievement, (b) for administrative purposes, (c) to provide guidance to students about future course work, (d) to provide guidance to teachers for instructional planning, and (e) to motivate students. The most widely accepted purpose of grades was to provide feedback about student achievement. Administrative purposes included using grades to make decisions about student matriculation and retention, to place students transferring from one school to another, and to make decisions about entrance into college. Counselors used grades to recommend courses to individual students and to recommend suitable occupations for students to consider (Johnson, 2001). Sometimes teachers used grades to make decisions about a student's strengths and weaknesses and to be able to group students for instruction. Grades were also used to motivate students positively or negatively, both as rewards and punishments (Guskey, 1996; Guskey & Bailey, 2001).

Some experts contend that grading and reporting were not essential to instruction (Friedman & Frisbie, 2000; Guskey & Bailey, 2001; Waltman & Frisbie, 1994). In 1958, Ellis Page, an educational scholar and researcher investigated how teachers' grades and

comments affected student learning. He found students receiving standard comments with their grades achieved significantly higher scores than those who received only a score and grade (as cited in Guskey, 1996; Trumbull & Farr, 2000). Page also found that the beneficial effect of grades on student learning came only when teachers used specific or individualized comments to accompany them. His research validated that grades were not essential to teaching or learning, but that they could be used in positive ways to improve students' achievement and performance.

The purpose of grading is different for teachers at different grade levels. Teachers at the elementary level primarily use grade reporting as a method to communicate information about student achievement to parents and students (Howley et al., 1999). Their intention is for grading to help students self-evaluate, enlist parent support in helping their child learn, assist with identifying learning objectives and counseling students, provide incentives for learning and to provide documentation of progress or lack of progress (Guskey, 1994; Johnson, 2001). Secondary teachers view grades as necessary to inform students, other teachers, and colleges about performance (Trumbull & Farr, 2000). Grade reporting procedures generally fall into two categories: formative evaluation (informing students of progress during instruction); or summative evaluation (providing marks at periodic marking periods) (Johnson, 2001). Both of these procedures concern teachers reporting a student's achievement. No matter what the level, a grade, a score or a mark is a relative judgment of a student by a teacher. This relative judgment is usually dependent on an assignment, a test, the class, or the instructor. The grade, score, or mark may vary depending on the situation (Johnson, 2001).

Problems and Concerns of Grading

Teacher grading practices for report cards have been inconsistent. One letter grade on a report card often indicates the average of several graded items (Smith, 1999). These items may reflect classwork, class participation, homework, projects, progress, effort, discipline, extra work, and summative and/or formative assessments. Sometimes teachers assign percentages to the different variables they consider. These percentages may then vary according to the graded item or according to the subject being graded. If numerical testing results for a group are judged to be too low to fit within the predetermined grading scale, then teachers will often grade on a curve. Given the numerous assessment variables and the opportunity for teacher discretionary grading practices, there is the potential for grading to become further removed from what the student has learned. Findings on a national survey done by Bursuck, Polloway, Plante, and Epstein (1996) and other researchers indicate that teacher discretionary practices are common, with about 50% of teachers using specific grade adaptations such as basing grades on improvement, giving multiple grades for a test or assignment, and making individual adjustments (Howley et al., 1999). Furthermore, the wide array of methods available for teachers to use when grading also contribute to the great variability, unreliability; and ultimately, subjectivity in grading practices (Marzano, 2000; Smith, 1999; Trumbull & Farr, 2000).

A key factor with grading is that teachers most often have enough control over grading practices to make critical decisions on their own. It is not uncommon for many teachers to develop their own grading practices and systems of determining grades. These types of teacher grading practices sometimes are the result of their ignorance of district policies, ignorance of measurement issues, and/or a lack of training. Although many

school systems have procedures in place for determining a student's grades and what the grades mean (Smith, 1999), grading practices are still influenced by a teacher's own attitudes, values, and assumptions (Cizek, 1996).

Much of the subjective teacher influence on grading comes from the grading of nonachievement factors such as effort, attendance, and behavior (Marzano, 2000). The Miami-Dade County Public Schools in Miami, Florida, and many other school system grading policies, separate these factors from more objective assessments. A sample report card can be found in Appendix A. Ideally, with these factors filtered out, the grade reported should reflect the overall academic achievement (Marzano, 2000). Most secondary level grading practices often give this appearance because single grades dominate the reporting system (Guskey & Bailey, 2001). Unfortunately, that single grade often includes effort, attitude, and achievement (McMillan & Workman, 2002). Wiggins (1994) commented that the single reported grade raises questions about validity and value, and that it can hide more than it reveals. Likewise, Howley et al. (1999) reported that achievement is usually part of the grade, but often not the whole of it.

Teachers' grading practices can be influenced by each individual's perception of what a grade should mean. Many teachers view grades as reward structures in the classroom (Guskey, 1996; Howley et al., 1999; Marzano, 2000). Some also view grades as something students earn as compensation for their effort. Research indicates that grades have some value as rewards, but no research has been able to validate the use of grades as punishments (Guskey, 1996; Guskey & Bailey, 2001; Trumbull & Farr, 2000). Failing grades cause some students to have a poor self-concept and often cause them to withdraw from learning. Seeley (1994) describes the mismatch of grades and

achievement by questioning how grades can reflect student progress and still encourage students to continue learning.

It has been suggested by some researchers that teachers' grading practices differ by school and are shaped by school culture (Howley et al., 1999). Analyzing data from the National Education Longitudinal Study (NELS) of 1988, researchers at the Office of Educational Research and Improvement (OERI, 1994, as cited in Howley, et. al., 1999) indicated that students in high poverty schools who received high grades had lower achievement test scores than students in wealthy schools with the same grades. Specific findings indicated high poverty school students receiving A grades had lower achievement test scores than their counterparts in affluent schools. High poverty school students receiving B grades received achievement test scores that were the same as students from affluent schools that received D grades. Students from high poverty schools receiving C grades got about the same achievement test scores as students from affluent schools that received F grades. It was suggested that the shared ethos of grading in the schools contributed to the differences. Student attitudes and family variables contribute to behavior, and achievement grades are indirectly affected by student behavior (Willingham, et. al., 2002). This research adds another dimension to factors that contribute to variations in grading and supports the fact that teachers' grading practices are divergent. Teachers are an important factor in determining students' grades, and the grading differences that exist may not be based entirely on what the student has learned (Smith, 1999).

Parent Perceptions About Grading

Most people perceive that one of the primary purposes of grading students is to communicate student achievement. This communication, primarily for students and parents, is meant to indicate student progress. Just as teachers' grading practices are diverse, so are the parent perceptions concerning the meaning of report card grades. Waltman and Frisbie (1994) studied perceptions of fourth grade students' parents and those students' teachers to see how they interpreted report card grades. Both groups perceived report cards as a valuable communication tool, but substantial differences emerged between how parents perceived the meaning of report cards and the teachers' intended meaning. Parents' perceptions of grades were influenced by grades their child had received previously and by grading practices of previous teachers their child had encountered (Waltman & Frisbie, 1994).

Research on parent perception of grading secondary students with and without disabilities was done by Munk and Bursuck (2001). Results indicated both groups of parents did not feel report card grades were effective in conveying information that was meaningful to them. Parents of average achieving students in both groups felt grades were more effective in communicating the need for special help and programs than did parents of high and low achieving students. Additionally, parents of high and low achieving students, with and without disabilities, felt grades were best at communicating effort and work habits. The only significant difference found between the perceptions of both groups was that parents of high achieving students without disabilities ascribed more importance to grades as conveying information to postsecondary schools or employers than did other parents.

Just as parents' perceptions of report card grades are discrepant, teachers' grading practices can also vary according to their perception of what will please parents (Lentz, 1997). Conversely, the meaning parents conclude about report cards is not necessarily what the teacher intended (Friedman & Frisbie, 2000). Generally, however, parents believe grades reflect achievement even though teachers may be factoring in effort, attitude, and/or behavior (Trumbull & Farr, 2000).

Growing Emphasis on Student Performance

Grades are also a method of accounting for student achievement to the community. Some states are grading schools as a method of reporting to parents certain characteristics and criteria each school has met. These criteria can change from year-to-year and state-to-state, which only add to parents' confusion about what these grades mean. The grading of individual schools and the federal government's No Child Left Behind Act (NCLB) was the beginning of an era of standards and growing accountability (Johnson, 2001). With this comes a growing emphasis on student performance and performance-based forms of assessment.

Assessment is defined as the process of gathering, describing, or quantifying information about performance (CRESST Assessment Glossary. 1996). A performance assessment is an assessment that requires students to generate a response to a question rather than choose from a set of responses provided to them. Students would ideally be required to accomplish complex and significant tasks while bringing to bear prior knowledge, recent learning, and relevant skills to solve realistic or authentic problems (CRESST Assessment Glossary. 1996).

In order to assess legislatively mandated education achievement goals, states have defined learning standards and benchmarks for learning. Standards are the broadest of a family of terms referring to statements of expectations for student learning (CRESST Assessment Glossary. 1996). A benchmark is a specific statement of a specific level that describes what students should know and be able to do (Florida Department of Education, 2004). States generally developed mandatory standardized tests around those standards and benchmarks. A state standardized test is a consistent set of procedures for designing, administering, and scoring an assessment and is administered under directions from the state. This type of test assures all students are assessed under the same conditions so that their scores have the same meaning and are not influenced by differing conditions. These tests known as, "high stakes tests" have high stakes consequences for students. An individual student's score determines not just who needs help, but whether a student is allowed to take a certain program or class or will be promoted to the next grade (Heubert, et. al., 1999). With this shift in assessment, testing, and accountability, there has been a change in curriculum and teaching to insure what is being taught is standards-based.

A standards-based curriculum is a process for adapting instruction, materials, and assessment to make sure all students can achieve the standards (CRESST Assessment Glossary. 1996). With the standards-based curriculum have come assessments such as projects, portfolios, rubrics, self-evaluations, and performance demonstrations. Students are also being asked to do more problem-solving tasks. They must apply and integrate skills they have learned. They are to think, plan, analyze, and construct (Smith, 1999). New types of assessments have added new ways to document student achievement, and as a result, academic growth is being measured in different ways. Current academic

growth is being compared to the previous year's growth. Student writing portfolio assessments over several years are also maintained.

Establishing educational standards and benchmarks has been a starting point for measuring student performance. These standards can improve student achievement by defining clearly what is to be taught and what kind of performance should be expected (Boser, 2000; Burson, 2001). Not all researchers agree that standards and assessments are driving instruction (Hoff, 2001).

Outcomes of High Stakes Testing

Test-based reform strategies and the use of standards have gained widespread political acceptance across the country, especially as a means of accountability. The idea of accountability is also central to the theory of school reform (Heubert et al., 1999). There is a long history of using tests to change pedagogical priorities and practices (Abram & Madaus, 2003). In the United States, the use of testing as a tool for school reform dates back to1845 in Boston when Horace Mann, Secretary for the State of Massachusetts Board of Education, replaced oral exams with a standardized written exam. Internationally, testing use for school reform can be traced to the 15th century in Treviso, Italy, where teacher salaries were linked to student performance on an examination (Abram & Madaus, 2003).

In 1988, Madaus examined the effects of high stakes testing programs on teaching and learning in Europe and the United States. His findings, along with current research, confirmed seven principles that hold true for contemporary statewide testing efforts (Abram & Madaus, 2003). They are: (a) The power of tests to affect individuals, curriculum or instruction is a perceptual phenomenon that produces large scale effects

when it is believed the results are important; (b) The more tests are used for social decision-making, the more likely the social process it is intended to monitor will become distorted and corrupt; (c) If important decisions are attached to the test, teachers will teach them; (d) Where high stakes tests operate, the exam eventually defines the curriculum; (e) Teachers pay attention to the form of questions on high stakes tests and adjust their instruction accordingly, ultimately, "teaching to the test"; (f) When test results define future education or life choices, the test results become the goal of education rather than an indicator of achievement; (g) The control of the curriculum is transferred to the agency controlling the high stakes exams. Historically, tests without stakes or with low stakes seldom drive change or improvement (Reville, 2004).

Today, the consequences of student outcomes on high stakes tests determine who needs help, who will take a certain class or program, who will be promoted to the next grade, and who will graduate from high school. While many states and school districts base promotion and retention decisions on a combination of grades, test scores, attendance, and teacher recommendations, the trend is to base promotion mainly on test scores (Heubert et al., 1999). In the State of Florida, third grade students not meeting a required score on the Florida Comprehensive Assessment Test (FCAT) in Reading may not be promoted to the next grade (Florida Department of Education, 2004). These guidelines have been written into Section 1008.25 of the Florida Statutes. For those third grade students not passing the FCAT Reading Test, other assessment opportunities are also provided, although these are very difficult as well for the student. Nevertheless, students must still meet a specific cut score or the law requires that they must be retained in third grade. A cut score is a performance standard that divides the acceptable and

unacceptable levels of readiness (Heubert et al., 1999). Cut scores on the FCAT are divided into five different levels, with levels 2, 3, 4 and 5 being acceptable and level 1 being unacceptable. Likewise, high school students not meeting a proficiency score for the high school level FCAT may not graduate by receiving a regular diploma. Advocates feel that making the stakes high will cause teachers and students to take tests more seriously and work harder. The skeptics worry about the harmful consequences to individual students as a result of high stakes test decisions (Heubert et al., 1999). Minorities do worse if high stakes testing is used for promotion and graduation decisions (Brennan, Kim, Wenz-Gross, & Siperstein, 2001). When used properly, test results can be valuable in making informed decisions about student learning, instructional programs, and school performance (Plake, 2002).

Appropriate Test Use and Alignment

In 1982, the National Research Council adopted a framework for assessing whether a planned or actual test use is appropriate. An important consideration was whether the test had validity to the extent that it measured what it was intended to measure and to the extent that inferences and actions made on the basis of test scores were appropriate and accurate (CRESST Assessment Glossary, 1996). Furthermore, reliability, which is the degree that the results of an assessment are dependable and consistently measure particular student knowledge and/or skills, must also be considered (CRESST Assessment Glossary). In addition, it must be determined whether a student's performance on a test reflecting knowledge and skill is based on instruction or the result of poor instruction or factors such as language barriers and/or disabilities. Finally, it must

be considered whether the test scores lead to placements and other consequences that are educationally beneficial (Heubert et al., 1999).

In 2000, the American Educational Research Association (AERA) published an article: "Position Statement Concerning High Stakes Testing in Pre K-12 Education." This paper took the National Research Council's framework and a position from the American Psychological Association (APA) one step further. The document addressed concerns about a student's opportunity to learn the measured content, the need for validation of the test results for the intended purposes, the alignment of test content with both the curriculum and the content standards, and the validity of mastery levels for student classifications (Plake, 2002).

The issue of aligning the test content with the curriculum and content standards is significant for two reasons. First, it is important that what is being tested is consistent with what is being taught. Second, it is important to evaluate if tests are appropriately targeted to their goals. In a nationwide study of 47 states and 4,200 teachers, Pedulla (2003) found that state testing programs were affecting teachers and their instruction. The results were analyzed by a teacher's grade level and the types of stakes attached to the state tests. Pedulla found that 75% indicated that the district curriculum was aligned to the state-mandated test programs. Between 60% and 65% of teachers also agreed that the state-mandated test was compatible with daily instruction. Furthermore, between 55% and 65% indicated that their instructional textbooks and materials were compatible with the state-mandated test. In questions concerning test format and alignment, only one-half of the teachers indicated that they aligned their tests with the state tests, and less than one-half indicated the tests they used had the same format as the state tests.

Pedulla (2003) found a closer alignment of instructional materials and teacher tests where the stakes on state assessments were highest. He also found that differences existed between elementary and secondary teachers in the amount of time spent on tested areas. Elementary school teachers spent the most time on this instruction. About 75% of the teachers agreed that the state-mandated tests were causing them to teach in ways that contradicted their own ideas of good educational practices. Additionally, results indicated that teachers felt pressure from the district superintendent (more than 90% agreement) and the principal (more than 80% agreement) to raise test scores. Fewer than 30% of the teachers agreed that the state-mandated test measured high standards of achievement. These survey results are reflective of teachers across the United States and are relevant to this study.

Factors Affecting Test Performance

Research shows that teachers' attitudes about testing practices affect the way they prepare and administer standardized tests (Monsaas & Engelhard, 1994). Data suggest that test preparation is greater in the lower grades as opposed to the upper grades, and that teachers in schools with low socio-economic status engage in more test preparation than those in higher socio-economic schools (Trepanier-Street, McNair, & Donegan, 2001). Monsaas and Engelhard (1994) found the amount of pressure teachers felt to increase standardized scores was a predictor of testing practice behavior. This focus on testing practices can lead to "teaching to the test" (Jacobson, 2003). Barksdale-Ladd and Thomas (2000) found that teachers view high stakes tests as hurting their performance as good teachers and hurting the children because the instructional focus of "teaching to the test" causes anxiety and stress. In comparing predictors of testing practices, the teachers'

attitudes about testing practices were a better predictor than was the amount of pressure they felt to increase test scores. The lower the elementary school grade, the more teachers engaged in test preparation practices (Monsaas & Engelhard, 1994).

As a result of standards-based reform, curricula in the schools in the State of Massachusetts are changing to align with the state's standards (Burson, 2001; Reville, 2004). The Massachusetts Comprehensive Assessment System (MCAS) was part of an extensive school reform agenda in the early 1990s. There are no stakes attached to this assessment test until the tenth grade when students are required to pass a rigorous test for graduation. Comparisons with national standardized exams indicate improved student achievement (Burson, 2001; Reville, 2004) and strong correlations between student attendance and MCAS performance (Reville, 2004). It is logical to assume that high student attendance would mean high test results. The Massachusetts standards-based reform strategy encompasses high learning standards for all students, regular assessment to track progress, accountability, and consequences for both educators and students. The standards are the goals teachers strive to achieve, while the tests serve as the yardstick with which progress is measured (Reville, 2004).

Research asserts that the relationship between assessment and effective education should be revisited in the American schools. His research indicates that the most important factor for students' success in school is building self-confidence so students feel capable of success. Statewide assessments and report cards with an "F" do not allow students to believe in themselves as learners (Stiggins, 1999). He also believes that there is a need for professional development to allow educators to learn how to motivate students through the effective use of classroom assessment. The need for professional

development related to assessment is also echoed by Pearl Solomon (2002) in her book, *The assessment bridge: Positive ways to link tests to learning, standards, and curriculum improvement.*

Another area affecting test performance is the quality of the educational tests used for making high stakes decisions. Some experts contend that the standardized testing industry is unregulated and its products are of inferior quality (Jacobson, 2003). Standardized testing is also complicated for the elementary grades because new subjects are often introduced each new year. Concerns of equity and fairness between ethnic minority students and white students, and between female and male students, affect test performance on high stakes testing (Brennan et al., 2001; Jacobson, 2003).

The performance of a student on a standardized test is based on the assumption that good testing practices were upheld by the test developer in its construction, and that the test user has appropriately selected, administered, and interpreted the test (Heubert et al., 1999). The technical quality of educational tests affects how students perform on them (Plake, 2002). As a result, the American Psychological Association (APA), the American Educational Research Association (AERA), and the National Council on Measurement in Education (NCME) jointly developed the *Code of fair testing practices in education* in an effort to make *Standards and manuals for educational and psychological tests* available to all test users. The principles in the Code have been distributed to all testing companies (Heubert et al., 1999), but their use is voluntary and dependent on self-regulation. It is of vital importance that test users research selected tests prior to their use and ascertain whether they have validity, whether they are aligned to both the *Standards for and manuals for educational psychological tests* and the *Code*

of fair testing practices in education, and if they are also aligned to the standards which they contend to be assessing.

Assessment and Grades

As discussed earlier, a disparity exists amongst teachers in the methods they use for assigning grades and how members of the school community interpret the grades. A disparity of grading policies and procedures also exists within schools, across school districts, and across the nation (Seeley, 1994). As a result of these disparities, a situation has been created where the assessments used for reporting could be considered as generally unfair. If this is the case, then how can the system be made more "fair"? According to Woodward (2001), this means that the assessment system must reflect: knowledge, values, and experiences equally familiar to all students; knowledge and skills all students have had adequate time to acquire; and be free of cultural, gender, ethnic, and age bias. Furthermore, by giving grades based on fixed standards, there would be no need to grade on a curve as every student would then have the ability to demonstrate proficiency on each standard and benchmark. Other researchers believe that grades should not be a blend of other factors such as attendance, effort, and behavior (Marzano, 2000; McMillan & Workman, 1999; Trumbull & Farr, 2000). There is also the argument that since grades represent what students have learned, they should not be used as a reward or punishment (Marzano, 2000; Trumbull & Farr, 2000). Finally, there is also support to not consider homework to be part of the grade (McMillan & Workman, 1999; Woodward, 2001). If all of these factors were taken into consideration, grades would become more valuable in assessing student performance (Woodward, 2001).

As a result of the shortcomings of traditional grading practices and a desire to provide better information to parents, many school districts have moved to a standards-based reporting system instead of grades (Johnson, 2001; Marzano, 2000; O'Connor, 2002; Trumbull & Farr, 2000). Such systems, however, when they are not properly explained and implemented often include detailed standards-based reports which have caused confusion for parents and the community (Guskey & Bailey, 2001; Trumbull & Farr, 2000).

Whether reporting is done on standards-based report cards or by traditional report card grades, the key issues are the assessment methods that are being used by teachers, how they are aligned to the standards, and what factors contribute to the assessment of student performance. Clearly, the focus on standards is posing challenges in grading and reporting (Colby, 1999; Guskey & Bailey, 2001). Assessment specialists recommend teachers assign grades strictly on performance on an assessment using clearly defined performance criteria (Burson, 2002; Guskey & Bailey, 2001).

Report Card Grades and High Stakes Testing

While there is a multitude of research concerning report card grading practices, testing, and assessment, the research concerning the relationship between report card grades and high stakes testing is extremely limited. Most people agree that regardless of their form, report cards are used for communication of a student's educational progress. Likewise, high stakes tests based on state standards are also used to communicate a student's educational progress. If both instruments are intended to document a student's overall progress, then it would be reasonable to expect that there would be a strong correlation between the two. The NELS longitudinal study (Willingham, Pollack &

Lewis, 2002) found that grades and standards-based test scores only moderately correlated because the two instruments only partly overlap. However, these researchers found that grades and test scores were strongly related for some individuals as well as some groups. The differential strengths of grades and test scores were attributed to a significant grade variation among schools due to grading variability from one teacher to another.

The NELS study analyzed five factors contributing to grade and test differences (Willingham, Pollack, & Lewis, 2002). These included: subjects covered by the teacher, grading variations, reliability, student characteristics, and teacher ratings. Also analyzed were four categories that could be the source of discrepancies between grades and test scores. These included: content differences between grades and test scores, individual differences that interact with content differences, situational differences, and errors in grades or test scores. Research results concerning discrepancies indicated that grades represent broader content and reflect unique accomplishments but that tests more easily emphasize important content. In other words, grades reflect what a student has been studying, but tests reflect progress on significant long-term educational objectives.

Scholastic engagement was another factor that attributed to the grade and test score discrepancy. Teacher ratings indicated that a major factor was that teachers often took student behavior directly into consideration when assigning grades. The different correlational strengths between grades and tests were attributed to the validity and fairness of each of the measures used. As a result of the NELS study, Willingham et al. (2002) felt strongly that grades and tests should be used together in making consequential decisions about individual students.

Agnew (1989) (as cited in Howley, et al., 1999) also looked at correlations between grades on report cards and standardized measures of achievements. His data reflected alignment between classroom learning and mandated assessments. The resulting correlations were only moderate (Howley, et al., 1999). Olson (1989) (as cited in Howley, et. al., 1999) also found that when comparing the effects of teacher-made tests and standardized achievement tests with students' grades that the correlations showed a moderate relationship between achievement on standardized tests and report card grades (Howley, et al., 1999). He found a stronger correlation, however, between teacher-made tests and report card grades.

Research by Johnson (2001) compared students' report card grades in Reading, writing, listening, and mathematics with subtest scores the students obtained on the Washington Assessment of Student Learning (WASL), the State of Washington's standardized test. The results were analyzed both quantitatively and qualitatively and indicated some correlation between report card grades and the state assessment. There was a 78% agreement for report card grades and WASL test scores for fourth graders in the subjects of Reading and mathematics. Johnson contended that standardized assessments were only a snapshot of student learning as opposed to report cards that measure daily learning of skills. She also raised the fear that report card grades and WASL scores need to match or there could be legal ramifications.

Burson (2001) researched the correlation between the Pennsylvania System of School Assessment (PSSA) Reading and math tests, the Cognitive Abilities Test (CogAT) and report card grades. She found that the most consistent predictors of the PSSA scores were the results of the CogAT subtests. Burson (2001), like others, found

the need to align curriculum with evaluation constructs and curriculum standards and to focus on staff development for student expectations, assessment, and remediation (Boser, 2000; Cizek, 1996; Colby, 1999; McMillan & Workman, 1999, 2002).

The relationship between report card grades and high stakes tests is riddled with many questions that have not been definitively answered. The multitude of factors affecting report card grading practices and the many issues related to standardized assessment and standards-based instruction all contribute to the relationship. Investigators need to continue to understand the relationship and find out what other factors could be associated with the possible alignment of report card grades and high stakes tests.

Summary

This chapter reviewed the relevant literature and examined research findings concerning report cards, grading practices and testing. The chapter began with an overview of the history of report card grading practices, followed by a discussion of report card purposes, problems, concerns and parent perceptions of grading. High stakes testing, its use and factors affecting testing were also examined. Finally, the relationship between assessments and grades was reviewed and discussed.

The review of literature indicated that while report card grading practices in the United States are controversial, they are a tradition that is resistant to change. There is consensus that the purpose of report cards is for communication about a student's academic progress, but that the grades within them are obtained by a variety of methods and are often influenced by subjective teacher assessments. This is further complicated by differences between the teacher's intended purpose of a grade and the parents' interpretation.

Mandatory standardized state testing is a method of accountability for student achievement being used to assess progress towards the state and federal governments' education goals. These high stakes assessments affect the lives of students, parents, teachers, and schools, determine who needs help, eligibility for classes and programs, who will be promoted to the next grade, and who will graduate from high school. Research indicates these state-mandated high stakes assessments have created a conflict with traditional report card grades because the report card grades are poorly standardized, frequently subjective, and are detrimental if the curriculum and testing is not aligned to both assessments and report card systems.

Research concerning the relationship between report card grades and high stakes testing is limited. While one would expect a correlation between report card grades and high stakes testing, results indicate that there are only strong relationships for some teachers and some groups. Differences in the strengths of the correlations between the two are attributed to the grading variability from one teacher to another. Beliefs about grading differ from school to school and those differences result in variations in grading practices from one teacher to another. Other factors contributing to the differences include: scholastic engagement, validity and fairness of the assessment measures used, and factors such as behavior which contribute to the overall report card grade.

CHAPTER III

METHOD

This chapter discusses the procedures and methods used in this research and include the design of the study, subject sample group, instruments, data collection, data analysis, and a summary of the chapter.

Research Design

The purpose of this research was to determine if there were discrepancies between the final Reading report card grades for third grade students and their respective Florida Comprehensive Achievement Test (FCAT) Reading achievement levels. The report card grading system used by teachers is mandated by the State of Florida and the local School Board. The individual teacher determines the basis for grades and how the grading scale is applied. The child's final Reading report card grade is a reflection of cumulative grades earned quarterly in Reading throughout the school year. A third grade report card sample is located in Appendix A. The FCAT Reading test yields an individual Reading achievement level of 1 to 5, with 5 being the highest. Each level is derived from cut scores for the overall points that are scored on the test. Based on the number of points earned on the FCAT, cut score values are associated with each of the five individual Reading achievement levels. The Florida legislature annually determines the cut scores that are used for each achievement level A FCAT Reading Student and Parent Report sample is located in Appendix B. In the current study, the child's individual FCAT Reading achievement level information was correlated to the final Reading report card grade that the teacher assigned each child based on his or her achievement in Reading in the classroom.

The research methodology applied to the data in this quantitative study was correlational. Correlation coefficients between the FCAT Reading test and Reading report card grade were calculated to determine the strength of the relationship for the sample as a whole, for each individual school, and for each individual classroom teacher. The correlations from each teacher's classroom were used as alignment measures and were related to behaviors and beliefs obtained from teacher surveys. In addition, teacher variables that contributed significantly to the relationship between third grade final Reading report card grades and FCAT Reading achievement levels' alignment were also determined. These variables included instructional methods, materials used, the subjects or topics the teacher taught, and what was being graded for report card calculations. To obtain this data, the third grade teachers in the selected sample were given a questionnaire survey instrument to determine the Reading materials and instructional methods that they used, their beliefs towards the FCAT, what specific Reading state standards they were teaching and grading, and the amount of time that they spent teaching the specific Reading standards. These teacher variables were compared to their alignment measures, and their correlation between third grade final Reading report card grades and FCAT Reading achievement test levels.

Subject Sample Group

Data for this study were collected subsequent to receiving the required approvals from the Institutional Review Board at Florida International University (approvals located in Appendix C) and the Miami-Dade County Public Schools' Office of Educational Research (approvals located in Appendix D). Student data were collected using the Integrated Student Information System (ISIS) of the school district and with the

help of the Miami-Dade County Public School's Office of Performance Improvement. Teacher and school data were obtained with the permission and cooperation of the school principals and teachers at the nine schools selected to participate in this study.

The student data sample for this investigation included 1064 students. This represented the third grade population in nine schools of one feeder pattern in the Miami-Dade County Public Schools, a large urban school district in the State of Florida. The data were collected for the 2003-2004 school year from those students who were enrolled, who were tested, and who completed the school year in the selected schools. These nine schools are located in fairly close proximity to each other, and all of the students eventually progress on to the same high school. The size of the third grade population at each school varied from 81 to 155 students. The students reflected the multi-cultural characteristics of Miami-Dade County, with a mix of mainly Hispanic (59%), White non-Hispanic (20%), African-American (14.6%), Asian (3.4%) and Other (3%). Approximately, 39.6% of the students were classified as economically disadvantaged, receiving either free lunch or a reduced price for lunch.

The teacher data sample for this research included questionnaire responses from 48 third grade teachers in the nine schools. This was a response rate of 92% of the third grade teachers teaching Reading in these schools. When analyzing the alignment variable with the *Third Grade Teacher Questionnaire* data, 34 teachers' questionnaires were used. Some third grade teachers of Exceptional Student Education self-contained classes had children with disabilities that did not participate in the FCAT Reading assessment test, so these were eliminated from the analysis.

Instrument

The *Third Grade Teacher Questionnaire* (sample located in Appendix E) that was administered to teachers in this study was developed by adapting two questionnaires that had been used in national studies. The background data questions, the questions relating to time spent on various teacher activities, and the questions concerning whether Reading/Language Arts was taught as a separate subject were all adapted from a survey developed for a project coordinated by Blank, Halbrook and DuBois (2003). The *Third Grade Teacher Questionnaire* also incorporated a questionnaire developed by Mary Foertsch (n.d.) which had been used in research about Reading practices and achievement. Other questions used in the first part of the *Third Grade Teacher Questionnaire* instrument were adapted to relate to the FCAT and concerned Miami-Dade County Public School's grading practices. Miami-Dade County Public Schools' Student Progression Plan is the document that indicates the criteria and parameters teachers use for implementing report card grading procedures. Questions on the *Third Grade Teacher Questionnaire* refer to this document and the district's grading system.

The second part of the *Third Grade Teacher Questionnaire* concerning instructional activities in Reading and Language Arts, instructional Reading materials, and Reading methods was adapted from the "English Language Arts and Reading Survey," copyrighted in 2003 by the Council of Chief State School Officers, Washington, DC; the Center for Educational Research, Madison, Wisconsin; and Learning Point Associates/NCREL, Naperville, Illinois. Permission granted to reproduce the survey was allowed for educational purposes. While the format of this survey was used, the instructional activities listed were taken from Florida's Sunshine State Standards. The

instructional Reading materials section was adapted to reflect those programs and methods used in the Miami-Dade County Public School System. Likewise, the Reading methods section of the questionnaire reflected those specified in the Miami-Dade County Public Schools' *Comprehensive Reading Plan Manual.*

Items on the questionnaire were grouped together for analysis purposes. Background data were collected to ascertain the teacher's educational level, teaching experience, ethnicity, and time spent on specific activities related to teaching. Other questionnaire items were grouped and analyzed. These were teacher beliefs about teaching and school (12 items), teacher beliefs about the FCAT (9 items), teacher beliefs about report cards and grading (9 items), Reading programs in use, planning and delivery of instruction (8 items), instructional methods and activities being used in Reading and Language Arts (22 items), and Reading methods or strategies that were in use (17 items). These variables were analyzed as predictors of alignment of final Reading report card grades and FCAT Reading achievement levels.

Data Collection

Data from the 2004 third grade FCAT Reading results and the June 2004 final Reading report card grades were obtained through the Miami-Dade County Public Schools' Integrated Student Information System (ISIS), the district's computer database. For the purpose of analysis, the data were grouped by school and by membership in each teacher's class. Teacher survey data were collected in small group settings with questions concerning the specific Reading materials and methods in use, the amount of time spent on the instruction of specific Reading state standards, what specific standards were graded, teacher grading practices, and teacher beliefs about the FCAT and its content.

Upon obtaining approval from Florida International University to use the *Third Grade Teacher Questionnaire*, the instrument was field tested on 23 teachers enrolled in a graduate Reading class. Teacher participants in this research were provided with the Consent to Participate in a Research Study Form (sample located in Appendix C), which explained the parameters for teacher participation. After obtaining the required signatures from participating teachers, the *Third Grade Teacher Questionnaire* was distributed for field testing on the graduate class. In addition to responding directly to the questionnaire, these participants were requested to respond with comments concerning clarity and ease of understanding the instrument. Their suggestions resulted in minor changes being made prior to administering the instrument to the target group.

This research study involved sampling all elementary schools with third grade students from one feeder pattern. All third grade teachers at nine elementary schools in the feeder pattern were eligible to participate in the research study. During the last month of the school year, the researcher met with third grade teachers at each of the nine schools. Meetings at each school were scheduled during the teacher's planning period. Teacher participants in this research were given the Consent to Participate in a Research Study Form located in Appendix C, which provided the parameters for teacher participation. Consent signatures were obtained from every teacher participant. The teachers were informed by the researcher that the questionnaires were confidential and would not be shared with their administrator or anyone else. School administrators would only receive results of the overall study. Then the *Third Grade Teacher Questionnaire* was distributed to the teachers in small group settings. The researcher was present during the completion of the questionnaire. There was no requirement that the school

administrator be present in the room when the questionnaire was being answered. The questionnaires were collected during that same planning period in which they were given since it took only 15 to 25 minutes to complete. Participants were not given any monetary remuneration for their involvement in the study. They were offered a candy bar in appreciation for their completion of the questionnaire. After all the questionnaires were completed, the data were compiled and analyzed. It is important to note that the data collected from the questionnaires was self-reported by the teacher and reflects that teacher's perceptions.

Confidentiality of Data

Confidentiality of student information was maintained by recording all data by individual student identification number for the purposes of input. Once the input was completed and reviewed for entry errors the data were analyzed without the identification numbers. Teacher information was coded by a four-digit school number and the last four digits of the teacher's social security number. This data were also reviewed for entry errors and matched to classroom identification codes to allow for matching students with their teacher. The data were then analyzed without teacher information. All written data associating names and participants were kept in a locked file drawer in the researcher's office.

Data Analysis

To analyze Research Question 1 and Hypothesis 1, students' characteristic data about school, gender, ethnicity, lunch status, language proficiency and exceptionality were summarized using frequencies and percentages. Teachers' characteristic data about gender, level of education, years of teaching experience, years teaching Reading/

Language Arts, and teaching assignment were also summarized using frequencies and percentages. Pearson Correlation Coefficients between the children's final Reading report card grades and the FCAT Reading achievement test levels were determined and tested for significance. Correlations were analyzed for the overall sample of third grade students, for each school's sample, and for each third grade teacher's class. Student data for special education students in self-contained low-functioning Exceptional Student Education classes were eliminated from the sample due to students' inability to participate in the FCAT Reading assessment. Davis (1976) describes the magnitude of correlations .50 to .69 as substantial, .70 to .99 as very high, .30 to .49 as moderate, and .10 to .29 in the low range. Davis' correlation ranges were used for this study.

For purposes of a closer detailed examination, final Reading report card grades were cross-tabulated to FCAT Reading achievement test levels for the overall sample and for each school's sample. To test whether alignments differed by school, Pearson correlations were tested for equality using a chi-square test on Fisher's r to z transformations. Post hoc pairwise comparison tests were carried out using z tests on the transformed correlations with Bonferroni's procedure. Tests were performed using SPSS Version 13 for Windows and considered significant at $p < .05$.

The *Third Grade Teacher Questionnaire* survey items were summarized using frequencies and percentages. The teacher correlations between Reading report card grades and FCAT levels were used as alignment measures. To analyze Research Question 2 and Hypothesis 2, these alignment measures were correlated to responses on the *Third Grade Teacher Questionnaire* survey items using Spearman's rho. This nonparametric correlation was used since teacher questionnaire items were dichotomous

or ordinal. Correlations were tested for significance at the $p < .10$ level. Individual questionnaire items significantly correlated to alignment were interpreted as possible predictors of alignment.

Summary

This study examined the alignment relationship between final Reading report card grades and FCAT Reading achievement levels. The research involved nine elementary schools in one feeder pattern in the Miami-Dade County Public School system in Miami-Dade County, Florida. The student population from which the data were analyzed included 1064 third grade students in these schools. The teacher data were collected through administering a *Third Grade Teacher Questionnaire* to the third grade teachers responsible for teaching Reading at these nine elementary schools. This data were collected from 48 teachers.

Correlation coefficients and cross tabulations were used to describe the alignment of final Reading report card grades and FCAT Reading achievement test levels overall, by school and by teacher. Teacher questionnaire items were grouped and analyzed as teacher beliefs about teaching and school, teacher beliefs about the FCAT, teacher beliefs about report cards and grading, instructional materials used, instructional methods and activities in Reading and Language Arts, and methods of Reading planning and strategies in use. Teacher alignments of final Reading report card grades and FCAT Reading achievement levels were correlated to items on the *Third Grade Teacher Questionnaire* using Spearman's rho correlation to determine possible predictors of alignment.

CHAPTER IV

RESULTS

This chapter presents a description of the results from the analysis of data as it relates to the research questions and hypotheses. Statistical analyses of the data collected are presented according to the procedures outlined in Chapter III. The primary purpose of this study was to examine the alignment between current Reading report card grading practices and standards-based state standardized testing results in Reading. A second purpose was to determine if factors such as teacher's beliefs about teaching and the FCAT, report card beliefs and grading practices, instructional activities, how the Reading program was planned and delivered, and Reading methods were associated with this alignment.

For this investigation, final Reading report card grades and Florida's FCAT Sunshine State Standards-based Reading Test Reading achievement levels for the third grade students were used. Possible factors associated with grade and test alignment were obtained from the teacher questionnaire items. First, student and teacher demographics are presented. Then Research Question 1 and Hypothesis 1 are analyzed with correlations of final Reading report card grades and FCAT Reading achievement levels and cross-tabulations overall, by school and by teacher. In addition, school correlations are tested for differences. Then results of the *Third Grade Teacher Questionnaire* are presented. Finally, Research Question 2 and Hypothesis 2 are addressed using Spearman's correlation.

Demographics of the Student Sample

The demographics of all student and teacher participants are discussed before presenting and discussing the results of this study. Presented in Table 1 are the frequencies and percentages of students by school, gender, ethnicity, free/reduced lunch status, limited English proficiency program information, and exceptional student program information for the 1064 students participating in this study. The number of students per school involved in this study ranged from the smallest having 7.6% (n = 81) to the largest having 14.6% (n = 155) of this third grade sample. The data set included all third grade students from these schools consisting of n = 1064 students, where 51.7% were male and 48.3% were female. The ethnic distribution showed that the highest percentage of students were Hispanic (59%), followed by White (20%), and Black (14.6%), with a very low percentage of students being Asian or Other (includes multi-cultural and Indian). Frequently, subsidized lunch status of a student indicates socio-economic status. In this study, students were classified as either qualifying for free or reduced lunch or not qualifying for any assistance (Pay Full Price). The majority (60.4%, n = 643) of the students did not qualify for financial assistance.

English for Speakers of Other Languages (ESOL) is the program provided for students whose first language is not English. Of the total, 58.2% were English speaking only, while 41.8% did not have English as their first language. Of the total ESOL sample, 90.3% were classified as ESOL 5 students (those already exited from the program who are considered proficient in English), only 9.7%, (n = 43) of the students were not yet considered proficient in English (ESOL Levels 1–4).

Table 1

Description of Third Grade Students

Student Characteristic	Frequency	Percent
	Enrollment	
School A	81	7.6
School B	142	13.3
School C	129	12.1
School D	95	8.9
School E	118	11.1
School F	119	11.2
School G	155	14.6
School H	111	10.4
School I	114	10.7
Total	1064	100.0
	Gender	
Male	550	51.7
Female	514	48.3
Total	1064	100.0
	Ethnicity	
Hispanic	627	59.0
White	213	20.0
Black	156	14.6
Other	36	3.4
Asian	32	3.0
Total	1064	100.0

(table continues)

Table 1 (continued)

Student Characteristic	Frequency	Percent
	Lunch Status	
Free/Reduced	421	39.6
Pay Full Price	643	60.4
Total	1064	100.0
	English for Speakers of Other Languages (ESOL)	
ESOL Total	445	41.8
Non-ESOL Total	619	58.2
Total	1064	100.0
Level 1	0	0.0
Level 2	4	0.9
Level 3	20	4.5
Level 4	19	4.3
Level 5	402	90.3
ESOL Total	445	100.0
	Exceptional Student Education (ESE)	
ESE Total	329	30.9
Non-ESE Total	735	69.1
Total	1064	100.0
	Primary Exceptionality	
Specific Learning Disabilities	113	34.3
Other Health Impaired	25	7.6
Gifted	155	47.1
Other Exceptionalities	36	11.0
ESE Total	329	100.0

In the State of Florida, the Exceptional Student Education (ESE) population includes exceptionalities such as: Specific Learning Disabilities, Other Health Impaired, Gifted, Speech Impaired, Language Impaired, Autistic, Hearing Impaired, Physically Impaired, Profoundly Mentally Handicapped, Autistic, Severely Emotionally Disturbed, Emotionally Handicapped, and Trainable and Educable Mentally Handicapped. Within this data set, 30.9% of the students were classified as ESE. Of this number, 47.1% (n = 155) of the students in the ESE population were classified as Gifted. Gifted students are not always considered with the ESE population since they do not have true disabilities. When Gifted was excluded from consideration as ESE, only 16.3% of the total sampled population would be identified as ESE, with the predominant percentage of students 34.3% (n = 113) classified with Specific Learning Disabilities.

Demographics of the Teacher Sample

The third grade teacher data set (n = 48) for the surveyed population as shown in Table 2 were extracted from the *Third Grade Teacher Questionnaire*. The first five items of the questionnaire asked for demographic information. The third grade teaching staff was predominantly female (93.8%), and the level of education reported was closely split with 43.8% having a Bachelor's degree and 54.2% with a Master's degree. With regard to experience, only 8.3% were first year classroom teachers, while 56% had been teaching at least 10 years. The distribution of years teaching Reading and Language Arts was analogous to the years of teaching experience.

The teaching assignments reported revealed that 29.2% of the teachers were teaching ESE or Gifted. These teachers were not used in the alignment analysis because they were not solely responsible for the Reading grade, or the ESE population they taught

Table 2

Description of Teacher Population Surveyed

Teacher Characteristic	Frequency	Percent
	Gender	
Male	3	6.3
Female	45	93.8
Total	48	100.0
	Level of Education Degree	
Bachelor's	21	43.8
Master's	26	54.2
Doctorate	1	2.1
Total	48	100.0
	Years of Teaching Experience	
First year	4	8.3
2 to 9 years	17	35.4
10 to 19 years	13	27.0
20 to 33 years	14	29.1
Total	48	100.0
	Years Teaching Reading/Language Arts	
First year	5	10.4
2 to 9 years	19	39.5
10 to 19 years	11	22.9
20 to 33 years	13	27.0
Total	48	100.0
	Teaching Assignment	
Regular Classroom	33	68.7
ESOL Classroom	1	2.1
ESE Classroom	8	16.7
Gifted Classroom	6	12.5
Total	48	100.0

was low functioning and did not participate in the FCAT assessment. The remaining 70.8% (n = 34) of teachers from the regular or ESOL classrooms were used in the further analysis of alignment of final Reading grades and FCAT Reading achievement levels and for associating teacher questionnaire items to the alignment.

Results for Hypothesis 1 for the Student Sample

The first hypothesis states that third grade students with final Reading report card grades of A or B will attain FCAT Reading achievement levels of 4 or 5, and those with a grade of D or F will attain FCAT Reading achievement levels of 1 or 2. The distributions shown in Table 3 are evidence that some discrepancies between grades and FCAT levels exist. While 50.2% of the students received A or B final Reading grades, only 38.5% of the students obtained FCAT levels 4 and 5. In contrast, only 12.3% received D or F grades on their final Reading report card, but 27.1% of students scored at FCAT levels 1 or 2. A graphic representation showing the distribution of Final Reading report card grades and the distribution of FCAT Reading achievement levels can be seen in Figure 1. The Pearson correlation between final Reading grade and FCAT Reading achievement level was $r = .63$, $p < .001$.

To further examine how the FCAT Reading achievement levels were distributed for each final Reading grade, a cross-tabulation of individual student data are presented in Table 4. Of the 50.2% (n= 534) of students receiving an A or B, 65.5% (n = 350) scored at FCAT Reading achievement levels 4 or 5. Only 6.9% (n = 37) of those receiving an A or B, scored at the low FCAT levels of 1 or 2. Of the 12.3% students (n = 131) receiving a final grade of D or F, 70.9% (n = 93) scored at FCAT Reading achievement levels 1 or 2. Of those students receiving D or F grades, only 3.8% (n = 5) scored at the higher

Table 3

Distribution of Final Reading Grades and Distribution of FCAT Reading Levels

Final Reading Grade	Frequency	Percent	FCAT Reading Level	Frequency	Percent
A	132	12.4	5	83	7.8
B	402	37.8	4	327	30.7
C	399	37.5	3	366	34.4
D	101	9.5	2	107	10.1
F	30	2.8	1	181	17.0
Total	1064	100.0	Total	1064	100.0

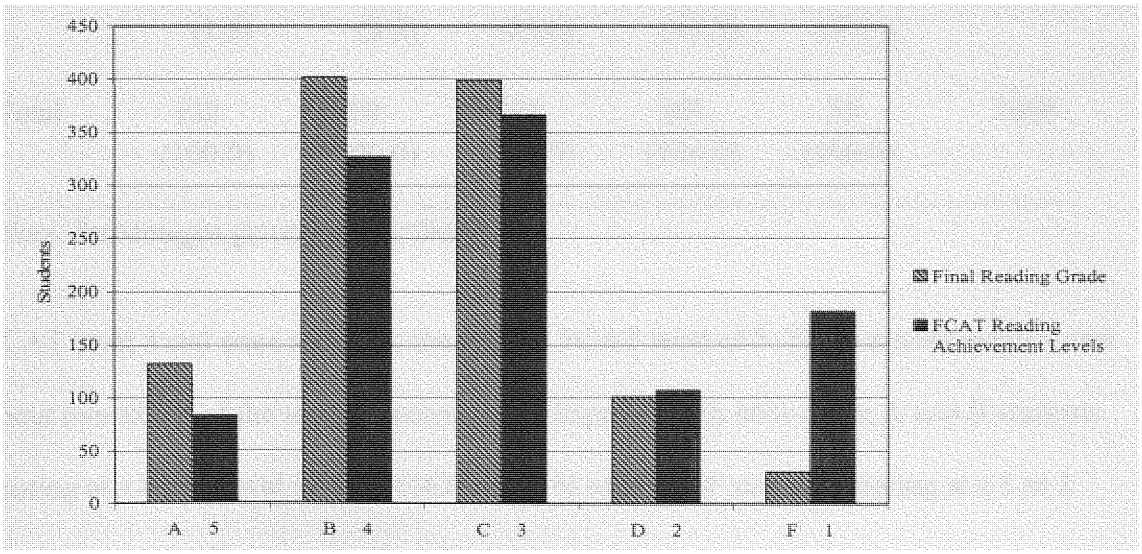


Figure 1. Distribution of final Reading grades and FCAT levels

Table 4

Distribution of FCAT Reading Levels by Final Reading Grades

	Final Reading Grade					
FCAT Level	A	B	C	D	F	Total
5	43 (32.6)	35 (8.7)	5 (1.3)	0 (0.0)	0 (0.0)	83
4	73 (55.3)	199 (49.5)	50 (12.5)	5 (5.0)	0 (0.0)	327
3	13 (9.8)	134 (33.3)	186 (46.6)	29 (28.7)	4 (13.3)	366
2	1 (0.8)	21 (5.2)	94 (16.0)	19 (18.8)	2 (6.7)	107
1	2 (1.5)	13 (3.2)	94 (23.6)	48 (47.5)	24 (80.0)	181
Total	132 (100.0)	402 (100.0)	399 (100.0)	101 (100.0)	30 (100.0)	1064

Note. Percentages given in parentheses

FCAT level 4 and none scored at the highest FCAT level of 5. These results indicate that about two-thirds of the sample scored, as anticipated, with 66% of the A and B students scoring at a 4 or 5 level on the FCAT and 71% of D and F students scoring at a 1 or 2 level. This partially supports Hypothesis 1.

Results for Hypothesis 1 by School

The relationship between Reading report card grades and FCAT Reading achievement levels was examined for individual schools and teachers. Alignment of final Reading grades and FCAT Reading levels data for each individual school are reflected in Tables 5 to 13. Data on the nine individual schools showed that for the majority (n = 7), 60% or more of the students who received a final Reading report card grade of A or B scored at FCAT Reading achievement levels 4 or 5. At the other two schools (A and H), the agreement at the upper range was distinctly lower at 54% and 41%, respectively. In the lower range at the majority of schools (n = 7), 70% or more of the students who received a final Reading report card grade of D or F scored at FCAT Reading achievement levels 1 or 2. This did not hold true for School B (35%) or D (55%) where discrepancies existed between receiving a D or F grade and scoring at a FCAT achievement level of 1 or 2.

The biggest discrepancy for grades as a predictor of FCAT achievement levels occurred at School B and H. At School B, 75% or more of the students receiving an A or B grade scored at FCAT levels 4 or 5, while only 35% or more of the students receiving D or F grades scored at FCAT levels 1 or 2. Conversely, at School H, only 41% of the students receiving A or B grades scored at FCAT levels 4 or 5, while 87% of the students receiving a final Reading grade of D or F scored at FCAT levels 1 or 2. Students at Schools B, D and F scored better than expected on FCAT while students at Schools H, A and B scored worse than expected.

School F showed the least discrepancies between final Reading report card grades and FCAT Reading achievement levels. Eighty-six percent of the students receiving

Table 5

Distribution of FCAT Reading Level by Final Reading Grade for School A

FCAT Level	Final Reading Grade			Total
	A & B	C	D & F	
5 & 4	18 (54.5)	0 (0.0)	0 (0.0)	18
3	13 (39.4)	11 (39.3)	4 (20.0)	28
2 & 1	2 (6.1)	17 (60.7)	16 (80.0)	35
Total	33 (100.0)	28 (100.0)	20 (100.0)	81

Note. Percentages given in parentheses.

Table 6

Distribution of FCAT Reading Level by Final Reading Grade for School B

FCAT Level	Final Reading Grade			Total
	A & B	C	D & F	
5 & 4	48 (75.0)	17 (32.7)	2 (10.0)	67
3	13 (20.3)	28 (53.8)	11 (55.0)	52
2 & 1	3 (4.7)	13 (25.0)	7 (35.0)	23
Total	64 (100.0)	52 (100.0)	20 (100.0)	142

Note. Percentages given in parentheses.

Table 7

Distribution of FCAT Reading Level by Final Reading Grade for School C

FCAT Level	Final Reading Grade			Total
	A & B	C	D & F	
5 & 4	39 (68.4)	5 (8.6)	0 (0.0)	44
3	14 (24.6)	32 (55.2)	2 (14.3)	48
2 & 1	4 (7.0)	21 (36.2)	12 (85.7)	37
Total	57 (100.0)	58 (100.0)	14 (100.0)	129

Note. Percentages given in parentheses.

Table 8

Distribution of FCAT Reading Level by Final Reading Grade for School D

FCAT Level	Final Reading Grade			Total
	A & B	C	D & F	
5 & 4	32 (72.7)	8 (19.0)	0 (0.0)	40
3	11 (25.0)	21 (50.0)	4 (44.4)	36
2 & 1	1 (2.3)	13 (31.0)	5 (55.6)	19
Total	44 (100.0)	42 (100.0)	9 (100.0)	95

Note. Percentages given in parentheses.

Table 9

Distribution of FCAT Reading Level by Final Reading Grade for School E

FCAT Level	Final Reading Grade			Total
	A & B	C	D & F	
5 & 4	36 (65.5)	10 (18.5)	0 (0.0)	46
3	14 (25.5)	18 (33.3)	2 (22.2)	34
2 & 1	5 (9.1)	26 (48.1)	7 (77.8)	38
Total	55 (100.0)	54 (100.0)	9 (100.0)	118

Note. Percentages given in parentheses.

Table 10

Distribution of FCAT Reading Level by Final Reading Grade for School F

FCAT Level	Final Reading Grade			Total
	A & B	C	D & F	
5 & 4	61 (85.9)	3 (7.9)	0 (0.0)	64
3	10 (14.1)	25 (65.8)	1 (10.0)	36
2 & 1	0 (0.0)	10 (26.3)	9 (90.0)	19
Total	71 (100.0)	38 (100.0)	10 (100.0)	119

Note. Percentages given in parentheses.

Table 11

Distribution of FCAT Reading Level by Final Reading Grade for School G

FCAT Level	Final Reading Grade			Total
	A & B	C	D & F	
5 & 4	54 (63.5)	3 (6.0)	1 (5.0)	58
3	31 (36.5)	21 (42.0)	4 (20.0)	56
2 & 1	0 (0.0)	26 (52.0)	15 (75.0)	41
Total	85 (100.0)	50 (100.0)	20 (100.0)	155

Note. Percentages given in parentheses.

Table 12

Distribution of FCAT Reading Level by Final Reading Grade for School H

FCAT Level	Final Reading Grade			Total
	A & B	C	D & F	
5 & 4	33 (41.3)	2 (8.7)	0 (0.0)	35
3	28 (35.0)	5 (21.7)	1 (12.5)	34
2 & 1	19 (23.8)	16 (69.6)	7 (87.5)	42
Total	80 (100.0)	23 (100.0)	8 (100.0)	111

Note. Percentages given in parentheses.

Table 13

Distribution of FCAT Reading Level by Final Reading Grade for School I

FCAT Level	Final Reading Grade			Total
	A & B	C	D & F	
5 & 4	29 (64.4)	7 (14.6)	2 (9.5)	38
3	13 (28.9)	25 (52.1)	4 (19.0)	42
2 & 1	3 (6.7)	16 (33.3)	15 (71.4)	34
Total	45 (100.0)	48 (100.0)	21 (100.0)	114

Note. Percentages given in parentheses.

A or B grades that scored at FCAT levels 4 or 5. Additionally, 90% of the students receiving D or F grades scored at FCAT levels 1 or 2. No other school was as predictable for both conditions.

School Alignment Correlations

The data in Table 14 indicates that all schools showed significant correlations at the $p < .01$ level, but in varying ranges. There was no evidence of correlations in Davis's moderate (.30 to .49) or low (.10 to .29) range. Correlations ranged from substantial, $r = .50$, at School B to very high, $r = .78$, at School F. There was a significant difference among the schools' correlations, $\chi^2 (8, N=1064) = 85.98, p < .001$. Post hoc pairwise comparisons using Bonferroni's procedure indicated several school differences, $p < .05$.

Table 14

Correlations of Final Reading Grade and FCAT Reading Level by School

School	*n*	*r*
School F	119	$.78_{a}$
School C	129	$.72_{ab}$
School G	155	$.72_{ab}$
School A	81	$.69_{bc}$
School I	114	$.64_{bc}$
School D	95	$.59_{bc}$
School H	111	$.56_{c}$
School E	118	$.55_{c}$
School B	142	$.50_{c}$

Note. All correlations are significant, $p < .001$. Correlations with different subscripts are significantly different using Bonferroni's procedure, $p < .05$.

The alignment correlation coefficient at School F ($r = .78$), was significantly higher than at Schools A ($r = .69$), I ($r = .64$), D ($r = .59$), H ($r = .55$), E ($r = .55$) and B ($r = .50$). The alignment correlation coefficient at Schools C ($r = .72$) and G ($r = .72$) were significantly higher than at Schools H ($r = .55$), E ($r = .55$), and B ($r = .50$). Since all school alignment correlations were significant, overall, this suggests that Reading report card grades and FCAT Reading achievement levels were aligned. However, because school alignment correlations varied from substantial ($r = .50$) to very high ($r = .78$), significant variability was observed. Schools F, C and G had higher correlations, even though they had characteristics much like the other schools. The schools and student populations had

similar demographics, yet some variability was found. Probable factors that affect these variations and contribute to the higher alignment will be discussed further in Chapter V.

Teacher Alignment Correlations

The correlations of final Reading grades and FCAT Reading achievement levels for individual teachers at each school are presented in Table 15. The grades of all 38 teachers had a significant alignment correlation with high stakes test results, $p < .05$. Thirty-four of the 38 teachers had a significant alignment correlation, $p < .01$. The highest correlations between final Reading grades and FCAT Reading levels were in the very high .80 to .90 range, while the lowest were in the moderate .35 to .49 range. Correlations between final Reading report card grades and FCAT Reading achievement levels for Teacher 13 (r = .87) at School C and Teachers 22 (r=.80) and 25 (r=.81) at School F were the highest, falling in the very high range. Teacher 6 (r=.47) and Teacher 7 (r = .37) at School B and Teacher 15 (r = .42) and Teacher 17 (r = .45) at School D had significant correlations at $p < .05$, but fell into the moderate range.

Third Grade Teacher Questionnaire Results

For this study *Third Grade Teacher Questionnaire* responses were analyzed. The first five items in the questionnaire requested demographic information about the respondents. This data has been discussed and can be found in Table 2. The remaining items involved questions about teacher beliefs about teaching and school, teacher beliefs about the FCAT, report card beliefs and grading practices, instructional activities, Reading programs in use, delivery of instruction, methods of planning, and Reading methods in practice.

Table 15

Correlations of Final Reading Grade and FCAT Reading Level by Teacher

Teacher	*n*	*r*	Teacher	*n*	*r*
School A			School F		
1	29	.74**	22	32	.80**
2	25	.69**	23	29	.78**
3	27	.68**	24	26	.73**
			25	32	.80**
School B			School G		
4	28	.63**	26	31	.58**
5	29	.64**	27	32	.68**
6	29	.46*	28	31	.77**
7	29	.37*	29	30	.57**
8	27	.49**	30	31	.68**
School C			School H		
9	25	.62**	31	28	.64**
10	24	.60**	32	26	.61**
11	24	.78**	33	29	.69**
12	27	.78**	34	28	.71**
13	24	.87**			
School D			School I		
14	23	.69**	35	27	.64**
15	26	.42*	36	29	.60**
16	22	.65**	37	29	.58**
17	24	.45*	38	29	.71**
School E					
18	29	.67**			
19	29	.56**			
20	24	.74**			
21	30	.59**			

*$p < .05$, **$p < .01$

Teacher Beliefs about Teaching and School

Teacher responses concerning the teaching profession, the teacher's school, teacher preparation time, and the overall school climate are presented in Table 16. More than 80% (n = 48) of teachers felt their workload was too high, and 58% (n = 48) felt they did not have enough time to teach and help individual students. Interestingly, more than 83% (n=48) responded that they had enough time to deliver quality lessons, while 31% (n=48) of the group did not feel that they had enough time to prepare these quality lessons. Approximately 88% (n=48) of the teachers understand what high performance means and an overwhelming 98% (n=48) are proud of their school. These factors are related to the school climate and can indirectly affect student learning.

Teacher Beliefs about the FCAT

Teachers' responses to the nine items measuring their beliefs about the FCAT are presented in Table 17. Almost 98% (n=48) of teachers felt there was too much emphasis placed on tests like the FCAT. Slightly more than half felt the FCAT did not measure what students had learned, what was taught, and they felt that the FCAT should be abolished altogether. More than 70% (n=48) of teachers agreed that the FCAT measured the Sunshine State Standards they were required to teach. It was also evident that teachers felt pressure related to FCAT performance from parents of students as well as the school leadership, and as a result, approximately 89% (n=48) felt they were teaching to the test.

Table 16

Teacher Beliefs about Teaching and School (n=48)

Questions	Percentage		
	Yes, Very Much	Yes	No
6. More enthused about teaching than when you started?	21.3	53.2	25.5
7. Recommend your school to a friend looking for job?	53.2	38.3	8.5
8. Proud of your school?	55.3	42.6	2.1
9. Enjoy your work?	66.0	29.6	4.3
10. Excellence recognized/rewarded at school?	28.9	60.0	11.1
11. Workload too high?	31.3	50.0	18.8
12. Enough time to teach and help individual students?	0.0	41.7	58.3
13. Enough time to prepare quality lessons?	8.3	60.4	31.3
14. Enough time to deliver quality lessons?	14.6	68.8	16.7
15. Leadership encourages risk taking and experimentation?	20.0	53.3	26.7
16. Discussions of educational issues with leadership?	25.0	62.5	12.5
17. Informed about what high performance means?	31.3	56.3	12.5

Table 17

Teacher Beliefs about the Florida Comprehensive Assessment Test (FCAT) (n=48)

Questions	Percentage		
	Yes, Very Much	Yes	No
18. Too much emphasis placed on tests such as FCAT?	62.5	35.4	2.1
19. Feel that FCAT is a good test?	4.2	56.3	39.6
20. Feel that FCAT should be abolished?	16.7	25.0	58.3
21. Feel FCAT accurately measures what students have learned?	2.1	41.7	56.3
22. Feel FCAT accurately measures what you taught your students?	2.1	41.7	56.3
23. Feel you are 'teaching to' the FCAT?	21.3	68.1	10.6
24. FCAT measures Sunshine State Standards in Reading?	6.4	66.0	27.7
25. Leadership pressures you for students to do well on FCAT?	17.0	53.2	29.8
26. Parents pressure you for their students to do well on FCAT?	14.6	50.0	35.4

Report Card Beliefs and Grading Practices

Questionnaire items 27 through 35 as presented in Table 18 relate to report cards, grading practices and teacher's beliefs about grading. One-third of the teachers did not feel the current report card grading system was adequate and more than 70% (n=48) felt the district's grading system was not fair and equitable. Approximately 62% (n=48) agreed that Reading report card grades reflected how well their students would do on the FCAT, while little more than 40% (n=48) felt Language Arts grades were reflective of how well students perform on the FCAT. Most teachers graded students by achievement on specific skills (98%) and on the Reading Sunshine State Standards (96%). Although homework is mandatory, about 43% (n=48) of the teachers did not count homework grades in the student's final grade. When asked whether test grades counted more than homework and daily assignments, 31.3% (n=48) responded they did not.

Instructional Time and Activities

The impact of teaching time and instructional activities in Reading and Language Arts is also a factor in Hypothesis 2. Question 36 was "Is Reading/ Language Arts taught mainly as a separate subject to your class?" Data indicated that two-thirds of teachers taught Reading and Language Arts as a separate subject, rather than combining the two subjects. The amount of time spent teaching Reading and Language Arts (also from Question 36) ranged from 90 minutes to 1120 minutes per week. Reflecting the school district mandate, slightly more than half (53.3%) of the teachers taught Reading and Language Arts for 600 minutes per week.

Table 18

Report Card Beliefs and Grading Practices (n=48)

Questions	Percentage		
	Yes, Very Much	Yes	No
27. Parents of students pressure their child to earn high grades?	23.4	57.4	19.1
28. Feel the current report card grading system is adequate?	0.0	66.7	33.3
29. Feel the district's grading system is fair?	4.3	67.4	28.3
30. Reading report card grades reflect performance on FCAT?	8.3	54.2	37.5
31. Language Arts grades reflect how he/she will do on FCAT?	4.2	39.6	56.3
32. Students graded by achievement on specific skills taught?	27.1	70.8	2.1
33. Students graded on the Reading Sunshine State Standards?	21.3	74.5	4.3
34. Homework grades count in students' final grades?	4.2	52.1	43.8
35. Test grades count more than assignments?	25.0	43.8	31.3

Questionnaire responses about instructional activities in Reading and Language Arts and how much time is spent on each of the skills are presented in Table 19. Skills fell into four major categories that included: Words and Phrases in Context; Main Idea, Plot and Purpose; Comparisons and Cause/Effect; and Reference and Research. In the Words and Phrases in Context category, more than 50% (n=48) of teachers spent more

Table 19

Instructional Activities in Reading and Language Arts (n=48)

Instructional Activity	Percentage of Time Taught				
	<10	11-25	26-50	51-75	>75
Words and Phrases in Context					
37. Predict content and purpose	8.3	6.3	16.7	18.8	50.0
38. Identify words and construct meanings	4.2	8.3	6.3	20.8	60.4
39. Determine meaning and increase vocabulary	2.1	8.3	4.2	27.1	58.3
40. Clarify understanding	2.1	4.2	8.3	14.6	70.8
41. Recognize effects of language	12.5	8.3	18.8	29.2	31.3
Main Idea, Plot and Purpose					
42. Determine main idea	2.1	10.4	4.2	6.3	77.1
43. Identify author's purpose	6.3	8.3	8.3	18.8	58.3
44. Recognize persuasive text	19.2	19.1	29.8	26.0	6.4
45. Personal preferences fiction or non-fiction	18.8	14.6	33.3	16.7	16.7
46. Recognize fact and opinion	4.2	12.5	18.8	29.2	35.4
47. Identify non-fiction, fiction, poetry, drama	12.5	10.4	25.0	22.9	29.2
48. Plot development and conflict resolution	6.3	6.3	14.6	22.9	50.0
49. Identify theme in story or non-fiction text	4.2	8.3	27.1	18.8	41.7
50. Form ideas from text, support ideas	2.1	10.4	10.4	22.9	54.2
Comparisons and Cause/Effect					
51. Recognize comparison and contrast	2.1	6.3	14.6	27.1	50.0
52 Similarities, differences characters, settings	4.2	4.2	16.7	20.8	54.2
53. Identify attitudes/values of time period	23.0	14.6	31.3	10.4	20.8
54. Identify and use literature terminology	8.3	16.7	29.2	22.9	22.9
55. Recognize cause and effect relationships	0.0	10.4	18.8	29.2	14.7
56. Explain motives/causes, compare own life	10.4	12.5	18.8	27.1	31.3
Reference and Research					
57. Use reference materials (maps, charts)	4.2	16.7	16.7	29.2	33.3
58. Organize information (reports, interviews)	6.3	25.0	31.3	20.8	16.7

than 75% of instructional time on the skills of: predict content and purpose, identify words and construct meanings, determine meaning and increase vocabulary, and clarify understanding (reread, summarize, etc.).

In the Main Idea, Plot and Purpose category, at least half of the teachers spent more than 75% of instructional time on: determine the main idea, identify author's purpose, understand plot development and conflict resolution, and form ideas from text by using information to support ideas. "Determine the main idea" was a skill that 77.1% (n=48) of the teachers spent more than 75% of their instructional time teaching. In the category of Comparisons and Cause/Effect, at least 50% or more of the teachers spent more than 75% of instructional time on: recognize comparison and contrast; and recognize similarities and differences in characters, settings, and events.

Reading Programs in Use, Delivery of Instruction, and Method of Planning

Questionnaire items 59 to 66 are related to the Reading programs and materials used, instructional planning, and how the Reading program is delivered to students. Results of these questionnaire items are summarized in Table 20. Generally, similarities existed amongst all schools in regards to their Reading program, textbook series and supplemental materials used. The Comprehensive Reading Program, a district mandated plan, was used by almost 98% (n=48) of respondents. More than 91% (n=48) of those using a basal Reading program used the *Scott Foresman Reading Textbook* series. All teachers used supplemental materials when teaching Reading. Additionally, almost everyone (98%) felt that the materials that were in use fit their philosophy.

While almost every third grade teacher (98%) surveyed taught Reading, there was a variation in how planning for Reading instruction was done. There were 42.6% of the

Table 20

Reading Programs in Use, Delivery of Instruction, and Method of Planning (n=48)

Question	Response	Percentage
59. Which Reading program is used?	Comprehensive Reading Program	97.9
	Direct Instruction	2.1
60a. Is a basal Reading program used?	Yes	45.8
	No	37.5
	Don't Know	16.7
61. Which basal Reading program?	Scott Foresman	91.2
	Houghton Mifflin	4.4
	Other	4.4
62. Are supplemental materials used?	Yes	100.0
	No	0.0
63. Who plans Reading instruction?	Myself	42.6
	With Other Teachers	57.4
64. Materials fit philosophy?	Yes	97.9
	Don't Know	2.1
65. Each teacher teaches Reading	Yes	97.9
	No	2.1
66. Switch teachers for Reading?	Yes	41.7
	No	58.3

teachers that planned individually, while 57.4% (n=48) did their Reading planning as a group with other teachers. Often times teachers group students and switch with other teachers to teach Reading. In this sample, it was discovered that only 41.7% (n=48) of the teachers switched children for Reading instruction.

Reading Methods in Practice

Other items considered were the methods used to teach Reading. Teacher responses concerning specific Reading methods that were used are presented in Table 21. The Reading methods surveyed were taken from the district mandated Comprehensive Reading Program. The Reading methods used by most teachers on a daily basis were: teacher directed whole group (83%), vocabulary development (64.6%), and questions and text discussion (62.5%).

Further analysis indicates that some teachers never used the Reading methods of: marginal note-taking (39.6%), dramatization such as role play, music/dance, poetry, puppets (16.7%), CRISS strategies (9.1%), reciprocal teaching where small groups practice critical Reading strategies (8.7%), *Accelerated Reader* software program (4.2%), and graphic organizers (2.1%). The Reading methods of teacher directed whole group instruction, question and text discussion, and vocabulary development were used consistently by most teachers on a daily basis.

Results for Hypothesis 2

The second hypothesis states that alignment between final Reading grades and FCAT Reading levels is related to: (a) teacher beliefs about teaching and school; (b) teacher beliefs about the FCAT; (c) report card beliefs and grading practices; (d) instructional activities in Reading and Language Arts; (e) Reading programs, delivery, and instruction; and (f) Reading methods in practice. In order to test this hypothesis, the teachers' correlations between the final Reading grades and FCAT achievement levels were used as the alignment measures. These were correlated to the

Table 21

Reading Methods In Practice (n=48)

Reading Methods	Response Percentage			
	Never	Some	Often	Daily
67. Teacher Directed Whole Group	0.0	4.3	12.8	83.0
68. Individualized (one to one)	0.0	38.3	34.0	27.7
69. Independent Reading (instructional level)	0.0	8.3	41.7	50.0
70. Questions and text discussion	0.0	2.1	35.4	62.5
71. CRISS (Strategies)	9.1	22.7	52.3	15.9
72. Accelerated Reader software program	4.2	20.8	29.2	45.8
73. Marginal note-taking	39.6	43.8	12.5	4.2
74. Multi-sensory activities	0.0	29.2	37.5	33.3
75. Brainstorming	0.0	18.8	43.8	37.5
76. Teacher Directed Small Group Guided	0.0	22.9	41.7	35.4
77. Vocabulary Development	0.0	2.1	33.3	64.6
78. Reciprocal Teaching (Reading strategies)	8.7	28.3	50.0	13.0
79. Read and Retell	0.0	10.6	48.9	40.4
80. Read Alouds (teacher reads material)	0.0	14.6	29.2	56.3
81. Buddy Reading (read to each other)	0.0	31.3	64.6	4.2
82. Graphic Organizers	2.1	12.8	55.3	29.8
83. Dramatization (role play, music, poetry)	16.7	50.0	25.0	8.3

Third Grade Teacher Questionnaire items in each of the six areas. Since the items on the questionnaire were dichotomous or ordinal, the nonparametric correlation, Spearman's rho, was used in all cases. There were 34 teachers with questionnaire and alignment data that were available for analysis. The significance level or this analysis was set at $p < .10$ to investigate all possible relationships between the teacher variable items and the alignment measure.

Items related to teachers' beliefs about teaching and school were specified in Table 16. Teachers' alignment was correlated to Question 12 "Is there enough time to teach and help individual students?", r_s=.31, $p < .07$. Higher alignment of Reading grades with FCAT Reading achievement levels was associated with teachers responding that they did not have enough time to teach and help individual students.

Questions concerning teacher beliefs about the FCAT were summarized in Table 17. None of the teacher beliefs were correlated to alignment. As a result, teacher beliefs concerning the FCAT were not a consideration in alignment between Reading grades and FCAT Reading achievement levels.

Items related to teacher beliefs concerning report cards and grading practices were presented in Table 18. Teacher's alignment was significantly correlated to Question 34, "Do homework grades count in students' final grades?", r_s=.31, $p < .08$. Lower alignment between final Reading grades and FCAT Reading achievement levels was associated with teachers responding that they counted homework in the final Reading grade.

Instructional activities in Reading and Language Arts were another factor under consideration for analysis. Questionnaire results about instructional activities were

displayed in Table 19. None of these items were significantly correlated with teacher alignment. No specific instructional activities supported closer teacher alignment of Reading report card grades and FCAT Reading achievement levels.

Reading programs used, the delivery of Reading instruction or how it was taught, and the way in which instructional planning was done were factors considered and reported in Table 20. Teachers' alignment was not significantly correlated with the Reading program used, with the model of delivery of Reading instruction or type of instructional planning. Thus, these factors were not possible predictors of teacher alignment of grades and FCAT Reading levels.

Various types of Reading methods used by teachers were also studied and were presented in Table 21. Teacher alignment was significantly correlated r_s=.42, $p < .01$, to Question 68, the amount of time used for individualized instruction. This indicates that teachers having higher alignment between Reading grades and FCAT Reading achievement levels spent more time using individualized instruction methods for teaching Reading.

Summary

The first research question asked whether discrepancies exist between final Reading report card grades and Florida Comprehensive Achievement Test (FCAT) Reading achievement levels for third grade students. The results for the total sample indicated that two-thirds of the sample scored in the upper ranges with Reading grades of A or B and FCAT Reading achievement levels of 4 or 5 and in the lower ranges with Reading grades of D or F and FCAT Reading levels of 1 or 2. Further investigation by school and by teacher revealed correlations between final Reading report card grades and

FCAT Reading achievement levels were all significant. Overall, a strong relationship exists between final Reading report card grades and FCAT Reading achievement levels for the total sample, each school and all teachers, although based on correlation coefficients, the strength of the relationship varies.

Hypothesis 2 stated that differences in teacher beliefs about teaching and the FCAT, report card beliefs and grading practices, instructional activities, Reading program planning and delivery and Reading methods contribute to correlations between final Reading report card grades and FCAT Reading achievement levels. Alignment was related to some teacher beliefs, grading practices, and Reading methods. Higher alignment of grades with FCAT levels was associated with teachers who felt they did not have sufficient time to teach and help individual students and teachers spending more time using individualized methods for teaching Reading. Lower alignment was associated with teachers who took homework into account in the final Reading grade. These results are discussed further in Chapter V.

CHAPTER V

CONCLUSIONS, DISCUSSION, AND RECOMMENDATIONS

This chapter begins with an overall summary of the study, followed by a discussion of the research findings and results. The literature reviewed will be related to the results of this investigation. Recommendations for further research will also be suggested along with implications for policy and practice.

Summary of the Study

The primary purpose of this research was to aid in understanding relationships between current Reading report card grading practices and standards-based state standardized (SBSS) testing results in Reading and the factors that affect the alignment of these relationships. In the study, the final Reading report card grades of third grade students were compared to Florida Comprehensive Assessment Test Reading achievement levels. Factors that might affect the degree of this alignment were also investigated.

By analyzing current report card grading practices and their relationship to SBSS testing, parents, students, and educators have reason to expect that students demonstrating high achievement levels on SBSS tests would have high grades on their report cards. Conversely, there is also an expectation that students demonstrating low achievement levels on SBSS tests would have low grades on their report cards. It is also reasonable to expect that if teachers' classroom instruction includes a focus on the state standards, then what teachers use to determine report card grades are the results of this instruction. As a result, report card grades would then be expected to demonstrate a corresponding relationship with SBSS achievement test levels. In practice, however, discrepancies can

be found between Reading report card grades and Reading achievement levels, which leads to questions about the factors that might contribute significantly to this discordance.

There were two central research questions and hypotheses that were investigated:

Research Question 1

Do discrepancies exist between final Reading Report Card Grades and Florida Comprehensive Achievement Test (FCAT) Reading Achievement Levels for third grade students?

Research Question 2

What factors contribute significantly to the relationship between third grade Final Reading Report Card Grades and FCAT Reading Achievement Levels due to membership in a particular teacher's class?

Hypothesis 1

Third grade students with a Final Reading Report Card Grade of "A" or "B" will attain a FCAT Reading Achievement Level of 4 or 5, and those with a Final Reading Report Card Grade of "D" or "F" will attain a FCAT Reading Achievement Level of 1 or 2.

Hypothesis 2

Differences in teacher beliefs about teaching and the FCAT, report card beliefs and grading practices, instructional activities, Reading program planning and delivery and Reading methods contribute to the alignment between Final Reading Report Card Grades and FCAT Reading Achievement Levels.

The research methodology applied to the data in this study used correlation coefficients to test the strength of the relationships for final Reading report card grades and FCAT Reading achievement levels for 1064 third grade students. The significance of

the correlation coefficient was used to measure the strength of the alignment between final Reading report card grades and FCAT Reading achievement test levels. School correlations were tested for differences using a chi-square test. A questionnaire survey instrument was administered to the students' third grade teachers to determine variables that may affect alignment between final Reading report card grades and FCAT Reading achievement levels. Spearman's rho correlations were used to describe the relationship of the teacher alignment measure with teacher beliefs about teaching and school, beliefs concerning the FCAT, report card beliefs and grading practices, instructional activities in Reading and Language Arts, the Reading program used, the model of instructional delivery and the type of instructional planning. This included correlating: the teacher's belief that there was not sufficient time to teach and help individual students; the amount of time teachers spent on individualized Reading instruction; and whether homework was accounted for in the final Reading grade.

Compilation of Findings

Here is what was found:

1. For the total sample, 66% of students whose final Reading report card grades were in the high range of A and B, scored in the high FCAT Reading achievement levels of 4 or 5.
2. For the total sample, 71% of students whose final Reading grades were in the low range of D and F, performed at low FCAT Reading achievement levels of 1 or 2.

3. While it was evident that for the overall teacher sample, a strong relationship, $r = .63$, existed between final Reading grades and FCAT Reading achievement levels, individual school correlations ranged from .50 to .78.
4. The majority (7 out of 9) of individual schools had 60% or more students who received final A or B Reading grades on the report card while receiving FCAT Reading achievement levels of 4 or 5, and 70% or more students who received D or F final Reading grades on the report card while receiving FCAT Reading achievement levels of 1 or 2.
5. Individual teacher alignments (correlations) between final Reading grades and FCAT Reading achievement levels were all significant and ranged from .37 to .87.
6. Higher teacher alignment of final Reading grades with FCAT Reading achievement levels was associated with those teachers responding that they did not have enough time to teach, $r_s = .31$.
7. Teacher beliefs about the FCAT did not correlate with the alignment of final Reading grades with FCAT Reading achievement levels
8. Higher teacher alignment of final Reading grades and FCAT Reading achievement levels was associated with teachers spending more time on individualized methods of Reading instruction, $r_s = .42$.
9. Lower teacher alignment of final Reading grades and FCAT Reading achievement levels was associated with those teachers that took homework into account in the final Reading grade, $r_s = .31$.

10. Instructional activities in Reading and Language Arts, the Reading program used, the model of delivery the Reading program delivery, and the type of instructional planning did not correlate with the teacher's alignment of final Reading grades with FCAT Reading achievement levels.

Discussion

This study examined the alignment of third grade students' final Reading report card grades with their Reading achievement levels attained on the Florida Comprehensive Assessment Test (FCAT). Both report cards and high stakes tests based on state standards are intended to communicate a student's educational progress. The expectation of alignment was based on the assumption that both instruments are intended to document and measure overall progress of a student. For students scoring in the upper range, this alignment assumption held for about two-thirds (65.5%) of the total sample where students performing with final Reading grades of A and B received FCAT Reading achievement levels of 4 or 5. For students performing in the lower range with D and F grades and receiving Reading achievement levels of 1 or 2, this assumption held for 70.9% of the cases. These results appear to be consistent with the Willingham et al. (2002) findings in the NELS longitudinal study where grades and standards-based test scores correlated moderately. Additionally, Johnson (2001) found that fourth grade students had a 78% agreement for Reading and Mathematics report card grades with the Washington Assessment of Student Learning (WASL) results, the high stakes assessment test for the State of Washington.

The NELS (2002) longitudinal study attributed the differential strengths of grades and test scores to significant grade variation among schools. This variation was further

traced to grading variability from teacher to teacher. The current investigation suggests the results of Willingham et al. (2002) findings. In this study, individual school correlations of alignment also showed discrepancies for grades as a predictor of FCAT Reading achievement levels. For seven of the nine schools evaluated, 60% or more of the students received a final Reading report card grade of A or B and scored FCAT Reading achievement level 4 or 5. In the lower range, seven of the nine schools had more than 70% of the students receiving a final Reading report card grade of D or F and FCAT Reading achievement level of 1 or 2. While all schools showed correlations significant at $p < .001$, the correlations ranged from .78 at School F to .50 at School B. Post hoc pairwise comparisons showed some significant differences among school's correlations. The correlation coefficient at School F was significantly higher than Schools A, I, D, H, E, and B, and correlation coefficients at Schools C and G were significantly higher than Schools H, E, and B. These findings suggest that teachers at School F have a closer understanding of students' Reading achievement as related to the FCAT and are grading items that are in congruence to what is being tested on the FCAT. Likewise, Schools C and D were also including tested FCAT skills in what was being counted for grades. What is being taught in Reading and what teachers use to count towards a Reading grade is key to uncovering factors that contribute to the relationship between grades and FCAT achievement. Differences in teacher's grading practices are evident here and can also be supported by research done by Howley et al. (1999).

When individual teacher alignment correlation coefficients were analyzed, the variability found between schools was supported by teacher differences. All teacher correlation coefficients were significant $p < .05$, and ranged from .37 to .87. Of the four

teachers with the lowest correlation coefficients, two were from School B and two were from School D. At School B, 75% or more of the students receiving A or B grades scored FCAT level 4 or 5, while only 35% or more of the students receiving D or F grades scored FCAT level 1 or 2. At School D, 72% or more of the students receiving A or B grades scored FCAT level 4 or 5, while only 55% or more of the students receiving D or F grades scored FCAT level 1 or 2. These findings suggest that teachers at School B and D need assistance to more closely align what they are teaching and counting towards grades with the tested FCAT skills. A factor might also be that these teachers don't know their students' individual needs as well as other teachers in this research. Teachers are important factors in determining students' grades (Smith, 1999). When looking at the large alignment variability in the lower range at School B, it is understandable that many parents would question why their child who scores an FCAT Reading achievement level of 1 or 2 has grades higher than a D or F in the final Reading grade. This lack of alignment and confusion about the meaning of grades supports Friedman and Frisbie's (2000) findings that the perceived meaning of report cards by parents may not necessarily be the teacher's intended meaning.

Another factor that may contribute to the low alignment between final Reading grades and FCAT Reading achievement levels is that the FCAT is not assessing what was taught by the teachers. Both the test and the classroom curriculum are based on Florida's Sunshine State Standards, which are the required benchmarks. If the teacher's curriculum is not aligned to these standards, it would be understandable that variability might exist between the test and the report card grade. Likewise, if the test is not aligned to the State benchmarks there would be a low alignment with the curriculum and ultimately with the

grades. Pedulla (2003) found that individual school district's curriculum was aligned to the state mandated test programs only 75% of the time. High stakes tests need to be evaluated to see if they are appropriately targeted to their goals. Pedulla's findings concerning test format and alignment were that only one-half of the teachers aligned their test with state tests.

The NELS 2002 study by Willingham et al. (2002) analyzed factors contributing to grade and test differences. The differences that were found by these researchers might also account for the low alignment of grades and Reading achievement levels. Factors analyzed in the NELS 2002 study included subjects covered by the teacher, grading variations, test reliability, student characteristics, and teacher ratings. They concluded that grades reflect what a student has been studying, but that state tests reflect progress on significant long-term educational objectives. Scholastic engagement can also contribute to the alignment variability because many teachers consider student behavior when assigning grades (Willingham et al., 2002).

In the current study, the *Third Grade Teacher Questionnaire* was used to investigate which factors were associated with the assessment relationship between final Reading report card grades and FCAT Reading achievement levels. Using nonparametric Spearman's rho correlations, items were identified which were associated with the teacher's alignment of grades and Reading achievement levels. A higher alignment of grades with FCAT levels was found with teachers responding that they did not have sufficient time to teach and help individual students. This might be interpreted that teachers with closer alignment have an improved understanding of the individual student's needs. A similar finding was that teachers who spent more time using

individualized methods for teaching Reading had a higher alignment between grades and FCAT levels. The higher alignment seems to support the idea that teachers who work more closely with the individual student and know the student's specific needs are those that will have less variability between report card grades and FCAT Reading achievement levels. While some researchers have investigated the possible sources of discrepancies between grades and test scores, research connecting the alignment measures between grades and test scores with variables such as time to teach and the understanding of individual student needs has not been published.

It was also found that there was a lower alignment between Reading grades and FCAT Reading achievement levels when teachers responded to the *Third Grade Teacher Questionnaire* that they took homework into account in the final Reading grade. This suggests that teachers accounting for homework in the report card grades leads to practices where the grading becomes further removed from what the student has learned. Homework assignments are often graded differently based on their importance, whereas test results may count as multiple grades. Consequently, grading using both homework and tests would not be as closely correlated as grading using only tests. Some research evidence indicates that the portion of homework completed each day has a stronger effect on grades earned, than does the time spent on the homework (Willingham, et. al., 1999). The divergent report card grading practices discussed in Chapter II are a result of the wide array of methods available to teachers, which contribute to the variability, unreliability, and subjectivity in grading (Marzano, 2000; Smith, 1999; Trumbull & Farr, 2000).

Further analyses of the *Third Grade Teacher Questionnaire* results and the alignment of report card grades and FCAT Reading achievement levels were conducted to ascertain if there were additional factors that could be associated with the alignment. There was no correlation of alignment with teacher beliefs about the FCAT, although 98% of the teachers felt there was too much emphasis being given to tests like the FCAT. None of the instructional activities in Reading and Language Arts was correlated with teacher alignment. Likewise, the Reading program used, the model of delivery of Reading instruction and the type of instructional planning done by teachers were not correlated to teacher alignment of grades with FCAT achievement levels. Of all the teachers surveyed, 97.9% implemented the District Comprehensive Reading Program and felt that the materials used fit their philosophy. Supplemental materials were used by all teachers when teaching Reading. However, none of these survey results were linked to higher teacher alignment between grades and Reading achievement levels.

Conclusions

The findings of this study support a strong relationship between final Reading report card grades and FCAT Reading achievement levels. This relationship showed 66% of students performing with the higher final Reading report card grades of A or B and scoring FCAT Reading achievement levels of 4 or 5. Similarly, the 70% of students performing in the lower range of D or F final Reading report card grades and scoring FCAT Reading achievement levels 1 or 2 was slightly larger than those in the higher range. The research question asking whether discrepancies existed between final Reading report card grades and FCAT Reading achievement levels can be answered affirmatively

based on these results. This is consistent with findings by Willingham et al. (2002) and Johnson (2001).

Individual school score analyses document the fact that the alignment of grades with FCAT Reading achievement levels was not consistent across schools and by teachers. Variations in the correlation coefficients support these differences and can be attributed to individual teacher differences at the schools. Some schools had a much higher alignment correlation as did some of the teachers at the individual schools. Research by Howley et al. (1999) suggests that teacher grading practices differ by school and are shaped by the school culture. These differences and school culture might also affect the variations in the alignment correlations.

The investigation of specific factors contributing to the higher alignment relationship between third grade final Reading report card grades and FCAT Reading achievement levels was analyzed based on the *Third Grade Teacher Questionnaire* results. Teacher alignment was not associated with teacher beliefs concerning the FCAT, instructional activities in Reading and Language Arts, the Reading program used, the model of instructional delivery, or the type of instructional planning. It was, however, correlated to the teacher's belief that there was not sufficient time to teach and help individual students, and to the amount of time teachers spent on individualized methods of Reading instruction. Teachers, who spent more time on individualized instruction, had a higher alignment relationship between grades and FCAT Reading achievement levels. Since these results were based on questionnaires where teachers self-reported, it would be interesting to actually observe instruction to verify if the self-reporting matched what was actually happening in the classroom. There was also a higher alignment correlation

among teachers who did not take homework into account in the student's final Reading grade. This may be because some teachers use homework as a learning tool. Tests should be compared to tests. This result suggests that teachers should not count homework grades in the final Reading grade if they want a closer relationship between grades and FCAT levels.

Implications for Policy and Practice

A considerable amount of research exists concerning grading that substantiates the need for a nationwide emphasis on teacher professional development related to this topic. Too often the field of education neglects the continuing education of teachers currently in the field. There needs to be more extensive, meaningful professional development concerning testing, testing practices, reliability and validity of testing measures. This must include pre-service education to new teachers as well as post service education for experienced teachers. Professional development must also occur concerning grading so that grading practices may become more consistent. Finally, grades should be based on clear-cut measures of performance on assessments using clearly defined performance criteria (Burson, 2002; Guskey & Bailey, 2001).

Since the advent of high stakes tests, many states have mandated and passed laws stating that students may not progress to the next grade level if a specific test score was not attained. While laws such as these are meant to ensure all students are making educational progress, they are limiting because of the weight assigned to a single test score. Decisions concerning educational consequences should be made on a multitude of available data and not on one specific test score. Legislators should reconsider their assessment mandates and the impact they are having on the American education system.

While tests are merely a yardstick of progress (Reville, 2004) and can be considered a snapshot in time, report card grades are a cumulative documentation of a student's educational progress. Both of these indicators of achievement need to be continuously analyzed to ensure that they complement each other and provide a realistic and accurate assessment of educational progress. Our society's future generations need to be provided with academic excellence that is determined by good teaching, accurate assessment methods and a reliable monitoring system for educational progress. This will be accomplished by a closer alignment of state standardized assessment measures with a standards-based curriculum and accurate report card methods.

Recommendations for Future Research

In the educational system today, there is a multitude of methods and forms of report cards used as a means of documenting and communicating student progress. Additionally, standards-based state assessment and evaluation systems are being used to document student progress. These systems of testing are being mandated and legislated in all fifty U.S. states as a method of accountability. The fact that the assessment and report card systems do not always agree has been supported by this research and that of other scholars (Boser, 2000; Burson, 2001; Friedman & Frisbie, 2000; Seeley, 1994). The implication of this non-agreement, however, is poorly understood and could be aided by additional research. Further research needs to study those students whose final Reading grades are not in alignment with their FCAT Reading achievement level. The specific reasons for this lack of alignment must be further investigated.

Assessment research needs to continue to determine if there is an alignment of state standards with the standardized tests being used to verify progress. This is especially

important because the testing industry is largely unregulated and the assumptions being made by test consumers concerning test reliability and validity needs to be verified. Not only should the alignment of the test with the standards be investigated, but the alignment of the test with the curriculum must also be further explored. If the expectation is that report card grades are to align with standards-based state testing, then the curriculum used for instruction must support what is being tested. Further research should investigate the specific standards teachers are teaching and whether the curriculum and materials being used are truly supportive of those standards

In light of continued accountability requirements and ever-changing curriculum mandates, research needs to further explore report card grading practices in order to make responsible decisions about student progress. Specific data should be collected about methods teachers use when giving grades. For teachers utilizing electronic gradebooks, investigation should encompass the specific items being graded and what items receive weighted grades. With the advent of computer technology, it is possible that grading practices could be standardized much more effectively and efficiently than has been in the past.

Many school systems have begun to reform the way that educational progress is being reported. The topic of grading continues to be a matter requiring further research. There needs to be a more quantitative method applied to grades and what is being graded to ensure consistency and reduce factors contributing to grade variations. While the traditional grades of A, B, C, D, and F are still the most prevalent, many school systems have changed to a standards-based report card. Research needs to be done on these school systems to determine if this new type of reporting system is more effective. School

systems also need to reevaluate their report card grading system in light of the requirement to meet state standards and benchmarks. Research also needs to be done to find new, more meaningful ways to communicate educational progress to the parents and community. A next step would be to study parent perceptions of their child's report card and then follow-up with parent training. Additionally, parent and teacher involvement and education must occur if reporting systems change because oftentimes the parent perception does not match the teacher's intended meaning (Friedman & Frisbie, 2000).

Further studies should be done on more divergent populations in other settings as well. This investigation was limited to a single population in an urban setting in one county in the State of Florida. Many factors can be attributed to the discrepancy between report card grades and standard-based state test results. Researchers should study teachers across the nation that have a high alignment between final Reading report card grades and state standards-based standardized test results to find additional common factors that could contribute to the higher alignment. Studies should also include the alignment of mathematics grades and achievement test results.

REFERENCES

Abram, L. M., & Madaus, G. F. (2003). The lessons of high-stakes testing. *Educational Leadership, 61*(3), 31.

Anderson, R., Hiebert, E., Scott, J., & Wilkinson, I. (1985). *Becoming a nation of readers: the report of the commission on reading.* Washington, DC: National Academy of Education, Commission on Education and Public Policy.

Barksdale-Ladd, M. A., & Thomas, K. F. (2000). What's at stake in high-stakes testing: Teachers and parents speak out. *Journal of Teacher Education, 51*(5), 384-97.

Blank, R., Halbrook, A., & DuBois, N. (2003). *Surveys of enacted curriculum in English Language Arts and Reading.* Retrieved February 19, 2004 from http://www.ccsso.org/content/pdfs/SECelaSrv2004.pdf.

Boser, U. (2000). Teaching to the test? *Education Week, 19* (39), 1.

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

Burson, K. C. (2001). A correlation analysis of Pennsylvania system of school assessment, cognitive abilities tests and report card grades for students in grades 3, 5, 8, and 11. (Doctoral dissertation, Widener University, Chester, PA). *Dissertation Abstracts International, 62* (12), 4003. (UMI No. 3036998).

Bursuck, W., Polloway, E. A., Plante, L., & Epstein, M. H. (1996). Report card grading and adaptations: A national survey of classroom practices. *Exceptional Children, 62*(4), 301-318.

Cizek, G. J. (1996). Grades: The final frontier in assessment reform. *NASSP Bulletin, 80*(584), 103-110.

Colby, S. A. (1999). Grading in a standards-based system. *Educational Leadership, 56*(6), 52-55.

CRESST/UCLA. (1999). *CRESST assessment glossary.* Retrieved October 9, 2004, from http://cresst96.cse.ucla.edu/CRESST/pages/glossary.htm.

Davis, J. A. (1971). *Elementary survey analysis.* Englewood, NJ: Prentice-Hall.

Florida Department of Education. (2004). *Understanding FCAT reports 2004.* Retrieved October 9, 2004, from http://www.firn.edu/doe/sas/fcat/pdf/fc_ufr2004.pdf.

Foertsch, M. (n.d.). *A Study of Reading Practices, Instruction, and Achievement in District 31 Schools.* Retrieved January 19, 2004, from http://www.ncrel.org/sdrs/areas/liread/tchrquest.htm.

Friedman, S. J., & Frisbie, D. A. (2000). Making report cards measure up. *Education Digest, 65*(5), 45-50.

Friedman, S. J., & Frisbie, D. A. (1995). The influence of report cards on the validity of grades reported to parents. *Educational and Psychological Measurement, 55*(1), 5-26.

Guskey, T. R. (2001). Fixing grading policies that undermine standards. *The Education Digest, 66*(7), 16-21.

Guskey, T. R. (1996). *Communicating student learning: 1996 ASCD Yearbook.* Alexandria, VA: Association for Supervision and Curriculum Development.

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

Guskey, T. R., & Bailey, J. M. (2001). *Developing grading and reporting systems for student learning.* Thousand Oaks, CA: Corwin Press.

Heubert, J. P., Hauser, R. M., & Committee on Appropriate Test Use. (1999). *High stakes: testing for tracking, promotion, and graduation research.* Washington, DC: National Academy Press.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1998). *Applied statistics for the behavioral sciences* (4th ed.). Boston, MA: Houghton Mifflin Company.

Hoff, D. J. (2001). Teaching, standards, test found not aligned. *Education Week, 21*(9), 6.

Howley, A., Kusimo, P. S., & Parrott, L. (1999). *Grading and the ethos of effort* (143 Reports--Research No. TM030360). Ohio, U.S.: from the ERIC database.

Jacobson, H. (2003). Expert panel debates high-stakes testing: James V. Hoffman notes drawbacks to high-stakes approach. *Reading Today, 21*(3), 13.

Johnson, J. K. S. (2001). The grading of elementary student performance on a standards-based report card (Doctoral dissertation, University of Washington, Seattle). *Dissertation Abstracts International, 62* (05), 1726. (UMI No. 3013976)

Leinhardt, G., Zigmond, N., & Cooley, W. W. (1981). Reading instruction and its effects. *American Educational Research Journal, 18,* 343-361.

Lentz, D. L. N. (1997). The assessment, grading and reporting practices of selected elementary school teachers and principals. (Doctoral dissertation, University of Texas, Austin). *Dissertation Abstracts International, 59* (02), 421. (UMI No. 9825006)

Marzano, R. J. (2000). *Transforming classroom grading*. Alexandria, VA: Association for Supervision and Curriculum Development.

McMillan, S. M., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *Journal of Educational Research, 95*(4), 203-213.

McMillan, J. H., & Workman, D. (1999). *Teachers' classroom assessment and grading practices: Phase I and II.* Richmond, VA: Metropolitan Education Research Consortium. (Reports - Evaluative - Tests/Questionnaires No. TM031260).

Monsaas, J. A., & Engelhard, G. J. (1994). Teachers' attitudes toward testing practices. *The Journal of Psychology, 128*(4), 469-477.

Munk, D. D., & Bursuck, W. D. (2001). What report card grades should and do communicate: perceptions of parents of secondary students with and without disabilities. *Remedial and Special Education, 22*(5), 280-287.

O'Connor, K. (2002). *How to grade for learning: linking grades to standards* (2nd ed.). Arlington Heights, IL: Skylight Professional Development.

Pedulla, J. J. (2003). State-mandated testing: what do teachers think? *Educational Leadership, 61*(3), 42.

Plake, B. S. (2002). Evaluating the technical quality of educational tests used for high-stakes decisions. *Measurement and Evaluation in Counseling and Development, 35*(3), 144-152.

Reville, S. P. (2004). High standards + high stakes = high achievement in Massachusetts. *Phi Delta Kappan, 85*(8), 591-598.

Seeley, M. M. (1994). The mismatch between assessment and grading. *Educational Leadership, 52*(2), 4-6.

Smith, B. T. (1999). Grading practices of elementary teachers (Doctoral dissertation, University of Virginia, Charlottesville). *Dissertation Abstracts International, 60* (06), 1860 (UMI No. 9935036).

Solomon, P. G. (2002). *The assessment bridge: positive ways to link tests to learning, standards, and curriculum improvement.* Thousand Oaks, CA: Corwin Press.

Stiggins, R. J. (1999). Assessment, student confidence, and school success. *Phi Delta Kappan, 81*(3), 191-198.

Tankersley, K. (2003). *The threads of reading: strategies for literacy development.* Alexandria, VA: Association for Supervision and Curriculum Development.

Trepanier-Street, M. L., McNair, S., & Donegan, M. M. (2001). The view of teachers on assessment: a comparison of lower and upper elementary teachers. *Journal of Research in Childhood Education, 15*(2), 234-241.

Trumbull, E., & Farr, B. (2000). *Grading and reporting student progress in an age of standards*. Norwood, MA: Christopher-Gordon Publishers.

Urdang, L., & Flexner, S. B. (1968). *The random house dictionary of the English language* (college edition). New York: Random House, Inc.

Waltman, K. K., & Frisbie, D. A. (1994). Parents' understanding of their children's report card grades. *Applied Measurement in Education, 7*(3), 223-240.

Wiggins, G. (1994). Toward better report cards. *Educational Leadership, 52*(2), 28-37.

Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: accounting for observed differences. *Journal of Educational Measurement, 39*(1), 1-37.

Woodward, J. (2001). Using grades to assess student performance. *Journal of School Improvement, 2*(1), 44-45.

APPENDICES

APPENDIX A

GR-03 MB-CAM
DISTRICT-101
PHONE:
ADDRESS:

FL-ID

MIAMI-DADE COUNTY PUBLIC SCHOOLS
SCHOOL YEAR 2003/04
CYCLE-031

FINAL REPORT
AUG. 25, 2003 - JUNE 9, 2004

SCHOOL-
ELEMENTARY
PRINCIPAL-KRISTINE DITTMAR

06/28/04 17:18:

COURSE SEQ-SEC	SUBJECT DESCRIPTION	1st Grading Period GRD	E	C	ATT	CM	2nd Grading Period GRD	E	C	ATT	CM	3rd Grading Period GRD	E	C	ATT	CM	4th Grading Period GRD	E	C	ATT	CM	Final Grade	TEACHER
	Third Grade																						
	OFFICIAL ATTENDANCE																						
	DAYS EXCUSED ABSENT				01					01					02					03			
	DAYS UNEXCUSED ABSENT				00					00					01					00			
	TIMES TARDY				00					00					00					00			
	Language Arts	A	1	A			A	1	A			A	1	A			B	1	A			A	FINE
	Reading	A	1	A			B	1	A			B	1	A			A	1	A			B	FINE
	Mathematics	A	1	A			A	1	A			A	1	A			A	1	A		15	A	CAMPUZANO
	Science	A	1	A			A	1	A			B	1	A			A	1	A		15	A	CAMPUZANO
	Social Studies	B	2	A			A	1	A			B	1	A			B	1	A			B	CAMPUZANO
	Art	A	1	A			C	1	A			A	1	A			B	1	B			B	HALL
	Music	A	1	A			A	1	A			B	1	A			A	1	A			A	ROSSBACH
	Physical Education	A	1	A			A	1	A			A	1	A			A	1	A			A	DULEHAM
	Spanish as a Second Language	A	1	A			A	1	A			A	1	A			A	1	A			A	RODRIGUEZ
	Advanced Academics	A	1	A			A	2	A			A	1	A			A	1	A			A	FINE

------- TEACHER COMMENT (CM) -------

15-Is very self-directive; can work independently.
15-Puede dirigirse a si mismo muy bien.

15-Shows excellent class attitude.
15-Demuestra una excelente actitud en la clase.

EXPLANATION OF GRADING SYMBOLS

---(GRD)-ACADEMIC---
A: EXCELLENT
B: GOOD
C: SATISFACTORY
D: IMPROVEMENT NEEDED
F: UNSATISFACTORY

---(E)-EFFORT---
1: OUTSTANDING
2: SATISFACTORY
3: INSUFFICIENT

---(C)-CONDUCT---
A: EXCELLENT
B: GOOD
C: SATISFACTORY
D: IMPROVEMENT NEEDED
F: UNSATISFACTORY

-(ATT)- ATTENDANCE

FLORIDA LAW PROVIDES FOR PARENT INPUT ON TEACHER/ADMINISTRATOR PERFORMANCE, WHEN APPROPRIATE. FOR MORE INFORMATION, CONTACT THE SCHOOL PRINCIPAL OR THE ACCESS CENTER.

LA LEY DE LA FLORIDA PERMITE A LOS PADRES, SI DESEAN, OPINAR SOBRE EL TRABAJO DE MAESTROS/ ADMINISTRADORES. PARA MAS INFORME LLAME A LA ESCUELA O CENTRO DE ACCESO ESCOLAR.

READING LEXILE SCORES: 0880 AS OF 05/04

---ADDITIONAL COMMENTS---

NEXT GRADE LEVEL IS 04
ACADEMICALLY PROMOTED TO THE NEXT GRADE LEVEL
AS OF: 06/09/04

DCT A-0046 MIAMI-DADE COUNTY PUBLIC SCHOOLS ACADEMIC INFORMATION RECORD E

OIT-3547 Rev. (08-01)

APPENDIX B

NAME:
ID:
SCHOOL:
DISTRICT: 13-DADE

Spring 2004
Florida Comprehensive Assessment Test (FCAT)
SSS Reading Student and Parent Report
Grade 03

Your 2004 Reading Results

Your child's Reading score is below grade level and needs improvement. Talk to your child's teacher about ways to improve.

La calificación de Lectura de su hijo/a está por debajo del nivel y necesita mejorar. Hable con el/la maestro/a de su hijo/a sobre maneras de mejorarla.

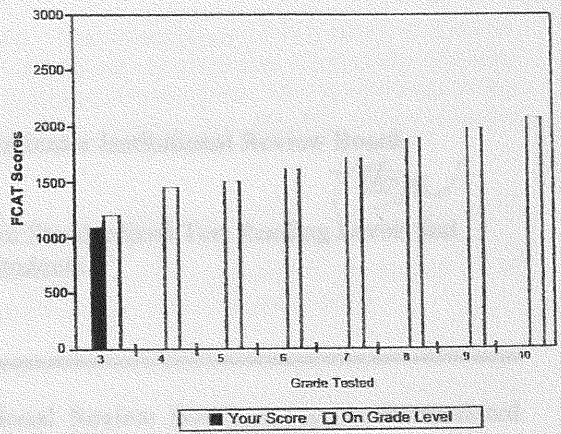
Nòt Lekti pitit ou pi ba pase nivo klas la epi li bezwen fè amelyorasyon. Pale ak pwofesè pitit ou sou kouman pou li fè amelyorasyon.

5	High
4	
3	
2	
1	Low

Your Reading FCAT Score History

Year	2001	2002	2003	2004
Grade Tested	NF	NF	NF	3
Achievement Level	NF	NF	NF	2
FCAT Score *	NF	NF	NF	1087

*These scores show your achievement on the day you were tested. If you were to take this same test again, it is likely that your 2004 FCAT Reading score would be between 1002 and 1172.



(Achievement levels 3 and above are considered on or above grade level.)

Your 2004 Reading Content Scores

Content Areas	Points Earned	Points Possible	Percent	Compared to other students L	M	H
Words/Phrases	7	10	70 %		X	
Main Idea/Purpose	8	15	53 %	X		
Comparisons	9	16	56 %	X		
Reference/Research	2	4	50 %	X		

L=Low, M=Middle, H=High

La table de la derecha muestra los niveles de rendimiento del FCAT y los rangos de calificación del FCAT para cada nivel.

El nivel de rendimiento 3 representa el trabajo "al nivel".

Antes de 2002, a los estudiantes se les otorgaba una calificación que seguía una escala de tres dígitos.

Tablo ki adwat la montre Nivo Rannman nan FCAT a ak Nechèl Nòt FCAT a pou chak Nivo.

Nivo rannman 3 reprezante travay "nan nivo klas" la.

Anvan 2002, yo te rapòte yon nechèl nòt ki gen twa chif ladan pou elèv yo.

Chart of FCAT Achievement Levels and FCAT Scores

Grade	Reading Level 1	Level 2	Level 3	Level 4	Level 5
3	86-1045	1046-1197	1198-1488	1489-1865	1866-2514
4	295-1314	1315-1455	1456-1689	1690-1964	1965-2638
5	474-1341	1342-1509	1510-1761	1762-2058	2059-2713
6	539-1449	1450-1621	1622-1859	1860-2125	2126-2758
7	671-1541	1542-1714	1715-1944	1945-2180	2181-2767
8	886-1695	1696-1881	1882-2072	2073-2281	2282-2790
9	772-1771	1772-1971	1972-2145	2146-2297	2298-2943
10	844-1851	1852-2067	2068-2218	2219-2310	2311-3008

On grade-level

Prior to 2002, a three-digit scale score was reported for students.

Reading Content –Content scores give more specific information about the skills on the FCAT. Grade level expectations for students include:

- **Words/Phrases** – uses skills to determine word meaning, including word parts and relationships between words.
- **Main Idea/Purpose** – determines a stated or implied essential message, details, author's purpose, or plot.
- **Comparisons** – knows similar and different, cause and effect, and contrast.
- **Reference/Research** – uses information from a variety of sources to reach conclusions.

Data Run Date: 04/09/2004

Calificaciones de contenido de lectura
Las calificaciones de contenido proporcionan información más específica sobre las habilidades en el FCAT. Las expectativas al nivel del grado para los estudiantes incluyen:

- **Palabras/Frases** – usa la habilidad para determinar el significado de una palabra, incluyendo las partes de palabras y las relaciones entre palabras.
- **Idea principal/Propósito** – determina un mensaje esencial expreso o implícito, detalles del propósito del autor o el argumento.
- **Comparaciones** – conoce similar y diferente, causa y efecto y contraste.
- **Referencia/Investigación** – usa la información de una variedad de fuentes para llegar a conclusiones.

Nòt Kontni pou Lekti
Nòt pou kontni yo bay enfòmasyon ki pi egzak sou kompetans elèv la nan FCAT a. Men sa yo alann de elèv la dapre nivo klas la:

- **Mo/Fraz** – itilize aplitid pou detèminen siyifikasyon mo yo, sa vle di chak pati nan mo yo epi relasyon ki genyen ant mo yo.
- **Ide Prensipal/Objektif** – detèminen yon mesaj esansyèl ki deklare oswa siyere, bay detay sou objektif oswa plan otè a.
- **Konparezon** – konnen menm ak diferan, tòz ak efè, epi kontras.
- **Referans/Rechèch** – itilize enfòmasyon ki soti nan plizyè sous pou rivwi konklizyon.

NT=Not Tested, NR=Not Reported, NF=Not Found

FIU
FLORIDA INTERNATIONAL UNIVERSITY
Miami's public research university

The Division of Sponsored Research & Training
Office of Research Compliance, MARC 430

MEMORANDUM

To:	Kristine Dittmar
CC:	Peter Cistone
From:	Niurca E. Marquez-Castro, Coordinator Institutional Review Board
Date:	March 24, 2004
Proposal Title:	Predictors for Alignment of State Standardized Test Reading ~~Levels~~ and Report Cards for Third Grade Students
	Approval # 031504-02

Your study was deemed Exempt by the Institutional Review Board at Florida International University.

As a requirement of IRB approval you are required to:

1) ***Submit a completion report*** (Form B) upon completion of your project in order for the file to be closed.
2) Submit a proposal and receive approval for any additions or changes in the procedures involving human subjects
3) Provide immediate written notification to the IRB of every serious or unusual or unanticipated adverse event as well as problems with the rights or welfare of the human subjects. You must confirm the receipt of serious AE reports with the IRB office.
4) And, if applicable, utilize copies of the date stamped consent document(s) for the recruitment of subjects and receive annual renewal of consent documents.

Special Conditions: Informed consent approved for use.

Please note your approval number is indicated above. For further information, you may contact the IRB Coordinator by email at irbiacuc@fiu.edu or visit the DSRT – Human Subject web site at www.dsrt.fiu.edu.

APPENDIX C

IRB Approved
Date: 3/24/04
No.: 031504-02

CONSENT TO PARTICIPATE IN A RESEARCH STUDY

TITLE: PREDICTORS FOR ALIGNMENT OF STATE STANDARDIZED TEST READING LEVELS AND REPORT CARD GRADES FOR THIRD GRADE STUDENTS

You are being asked to participate in a research study. The investigator of this study is Kristine Dittmar and she is a student at FIU. The study will include about thirty-five third grade teachers. Your participation will require approximately twenty minutes of your time. We are looking at the predictors that affect alignment of Reading report card grades and Florida Comprehensive Achievement Test (FCAT) Reading achievement levels.

If you decide to be a part of this study you will be asked to complete a survey. You will be asked questions about your education, teaching experience, instructional focus, attitudes, instructional methods, materials used and grading practices. The researcher would appreciate if you answer all of the questions. If you get upset or uncomfortable during the survey, you may take a break.

We do not expect any harm to you by being in this study. The information you provide will be kept confidential. There is no cost or payment to you as a subject. You will not get any direct benefit from being in the study. However, your help will give us information about third grade teachers' attitudes, instructional focus, instructional methods, materials used and grading practices in Reading.

Your survey will be identified only by your school number and the last five digits of your social security number, not your name. Your name will not appear anywhere. Surveys will be assigned a number. Only the researcher will know the code that will be kept in a locked file cabinet in the researcher's private residence. All of your answers are private and will not be shared with anyone unless required by law. Your data will be compared to the data from others subjects. We will present the research results as a group. You may ask questions about the study at any time. If you choose not to participate, no one will be upset with you. You may also choose to stop your participation before you finish the survey.

If you would like more information about this research after you are done, you can contact Dr. Peter J. Cistone or me at 305 235-5738. If you would like to talk with someone about being a subject in this study, you may contact Dr. Bernard Gerstman, the Chairperson of the FIU Institutional Review Board at 305 348-3115 or 348-2494.

Department of Educational Leadership & Policy Studies
College of Education
11200 SW 8 Street, ZEB 313 • Miami, FL 33199 • Tel: (305) 348-3418 • Fax: (305) 348-1515 • www.fiu.edu

APPENDIX C

FLORIDA INTERNATIONAL UNIVERSITY
Miami's public research university

Your signature below indicates that all questions have been answered to your liking. You are aware of your rights and you would like to be in the study.

Signature of Participant	Printed Name	Date

I have explained the research procedure, subject rights and answered questions asked by the participant. I have offered him/her a copy of this informed consent form.

Signature of Researcher	Date

APPROVED

Department of Educational Leadership & Policy Studies
College of Education
11200 SW 8 Street, ZEB 313 • Miami, FL 33199 • Tel: (305) 348-3418 • Fax: (305) 348-1515 • www.fiu.edu

APPENDIX D

Miami-Dade County Public Schools

giving our students the world

Office of Evaluation and Research
Executive Director
Dr. Robert A. Collins

April 23, 2004

Kristine Dittmar
12520 SW 108 Avenue
Miami. FL 33176

Dear Ms. Dittmar:

I am pleased to inform you that the Research Review Committee of the Miami-Dade County Public Schools (MDCPS) has approved your request to conduct the study, "Predictors for Alignment of State Standardized Test Reading Levels and Report Card Grades for Third Grade Students." The approval is granted with the following conditions:

1. Participation of a school in the study is at the discretion of the principal. A copy of this approval letter must be presented to the principal.
2. The participation of all subjects is voluntary.
3. The anonymity and confidentiality of all subjects must be assured.
4. The computer-generated data which are provided by the MDCPS will be either aggregated or coded to ensure the subjects' anonymity.
5. The study is based on anonymous student records, so parent permission forms are not required.
6. The study will involve approximately 1106 MDCPS students in grade 3.
7. Teacher participation is voluntary.
8. Disruption of the school's routine by the data collection activities of the study must be kept at a minimum.
9. The MDCPS internal school mail system cannot be used in conducting the study.

1500 Biscayne Boulevard, Suite 225 • Miami, Florida 33132
305-995-7501 • FAX 305-995-2691 • bcollins@sbab.dade.k12.fl.us

It should be emphasized that the approval of the Research Review Committee does not constitute an endorsement of the study. It is simply a permission to request the voluntary cooperation in the study of individuals associated with the MDCPS. It is your responsibility to ensure that appropriate procedures are followed in requesting an individual's cooperation, and that all aspects of the study are conducted in a professional manner. With regard to the latter, make certain that all documents and instruments distributed within the MDCPS as a part of the study are carefully edited.

The computer-generated data for the study will be provided by Ms. Gisela Feild of the Division of Data Quality Management of the MDCPS. Contact her at (305) 995-7511 to arrange a meeting to review your request and determine the cost.

The approval number for your study is 1078. This number should be used in all communications to clearly identify the study as approved by the Research Review Committee. The approval expires on June 30, 2006. During the approval period, the study must adhere to the design, procedures and instruments which were submitted to the Research Review Committee. If there are any changes in the study as it relates to the MDCPS, it may be necessary to resubmit your request to the committee. Failure to notify me of such a change may result in the cancellation of the approval.

If you have any questions, please call me at (305) 995-7501. Finally, remember to forward an abstract of the study when it is complete. On behalf of the Research Review Committee, I want to wish you every success with your study.

Sincerely,

Joseph J. Gomez, Ph.D.
Chairperson
Research Review Committee

JJG:fp

cc: Ms. Gisela Feild

APPROVAL NUMBER: 1078	APPROVAL EXPIRES: 6-30-06

APPENDIX E

Third Grade Teacher Questionnaire[1]

School Number _____ Last five digits of Social Security Number _____

By using a check (√) or filling in the blank please respond to the following questions as candidly and completely as possible. Your responses will be kept confidential. Thank You!

1. Gender: O Male O Female

2. Check √ Degrees and list major:
   O Bachelor's _____ O Masters _____
   O Specialist _____ O Doctorate _____

3. Years taught? _____

4. Years taught language arts/reading _____

5. How many ***hours each week*** do you spend on: (You may use fractions or decimals for part of an hour)						
	In school	**Out of school**			**In school**	**Out of school**
Grading classwork				Curriculum planning		
Grading homework				Classroom instruction		
Lesson planning				Administration		
Student records				Student mentoring/ counseling		

	Please check (√) one answer for each question below.	**Yes, Very Much**	**Yes**	**No**
6	Are you more enthused about teaching now than when you started?			
7	Would you recommend your school to a friend looking for a job?			
8	Are you proud of your school?			
9	Do you enjoy your work?			
10	Is teacher excellence recognized/rewarded at your school?			
11	Is your workload too high?			
12	Do you have enough time to teach & help individual students?			
13	Do you have enough time to prepare quality lessons?			
14	Do you have enough time to deliver quality lessons?			
15	Does the leadership encourage professional risk taking and experimentation?			
16	Does the leadership participate with the staff in discussions of educational issues?			
17	Do you feel you are informed about what high performance means?			
18	Do you feel there is too much emphasis placed on tests such as Florida Comprehensive Assessment Test (FCAT)?			
19	Do you feel that the FCAT is a good test?			
20	Do you feel that the FCAT should be abolished?			
21	Do you feel that the FCAT accurately measures what your students have learned?			
22	Do you feel the FCAT accurately measures what you have taught your students?			

1 (Adapted from "A Study of Reading Practices, Instruction, and Achievement in District 31 Schools Survey" Report by Mary Foertsch. Copyright © North Central Regional Educational Laboratory, Oak Brook, Illinois.)

	Please check (√) one answer for each question below.	Yes, Very Much	Yes	No
23	Do you feel that you are "teaching to" the FCAT?			
24	Do you feel that the FCAT adequately measures the Sunshine State Standards in reading?			
25	Does the leadership put pressure on you to get your students to do well on the FCAT?			
26	Do you feel the parents of your students put pressure on you to have your students do well on the FCAT?			
27	Do you feel the parents of your students put pressure on their child to earn high grades?			
28	Do you feel the current report card grading system is adequate?			
29	Do you feel the district's grading system is fair?			
30	Do you feel your child's reading report card grades reflect how well they will do on the FCAT?			
31	Do you feel your students' language arts report card grades reflect how well they will do on the FCAT?			
32	Do you grade students according to their achievement on specific skills you teach?			
33	Do you grade students according to the reading Sunshine State Standards?			
34	Do homework grades count in students' final grades?			
35	Do test grades count more than assignments (daily or homework) when entered into your grade book?			

36. Is Reading/language arts taught mainly as a separate subject (i.e., not integrated with other subjects) to your class? (Check √ one and fill in the blank.)

O **Yes** How many minutes per week in your class? _____ minutes/week

O **No** Average minutes per week spent on reading related instruction? _____ minutes/week

APPENDIX E

INSTRUCTIONAL ACTIVITIES IN READING AND LANGUAGE ARTS[2]

To get a sense of the content you cover in your class, please estimate the time a student will spend in that activity over the course of a school year. The activities are not necessarily mutually exclusive across activities. Consider each technique alone. Put a check (√) inside the box indicating the amount of time you estimate you will have spent on reading strategies this year. Your answers will undoubtedly exceed 100%.

	Percentage of Time Taught	0	<10%	11-25%	26-50%	51-75%	>75%
	Words and Phrases in Context						
37	Predict content and purpose						
38	Identify words and construct meanings						
39	Determine meaning and increase vocabulary						
40	Clarify understanding (reread, summarize, etc.)						
41	Recognize effects of language (rhymes, vocabulary, story structure and patterns)						
	Main Idea, Plot and Purpose						
42	Determine main idea						
43	Identify author's purpose						
44	Recognize persuasive text						
45	Identify personal preferences of fiction or non-fiction						
46	Recognize fact and opinion						
47	Identify characteristics of non-fiction, fiction, poetry and drama						
48	Understand plot development and conflict resolution						
49	Identify major theme in story or non-fiction text						
50	Form ideas from text and use information to support ideas						

	Percentage of Time Taught	0	<10%	11-25%	26-50%	51-75%	>75%
	Comparisons and Cause/Effect						
51	Recognize comparison and contrast						
52	Recognize similarities and differences in characters, settings and events						
53	Identify attitudes and values of time period from works written during that time period						
54	Identify and use literature terminology						
55	Recognize cause and effect relationships						
56	Explain motives of characters or causes of events and their comparison with those in his or her own life						
	Reference and Research						
57	Use reference materials (maps, charts, etc.)						
58	Organize information (reports, interviews, etc.)						

[2] (Adapted from English Language Arts and Reading Survey: Copyright © 2003 by the Council of Chief State School Officers, Washington, DC; the Wisconsin Center for Education Research, Madison, WI; and Learning Point Associates/NCREL, Naperville, IL. *All rights reserved with the exception of reproduction for educational purposes.*)

APPENDIX E

Instructional Reading Materials

59	Which one of the following District Reading programs to you use? (Check √ one) O Comprehensive Reading Program O Success for All O Direct Instruction O Other (specify)_____			
		Circle correct answer below:		
60	Are you using a basal reading program as a basis for your reading instruction? If yes, check √ which one? O Scott Foresman O Houghton-Mifflin O Harcourt-Brace O Macmillan/McGraw-Hill O Other_____	Yes	No	Don't Know
61	Do you use supplemental reading materials when teaching? Which ones?_____	Yes	No	Don't Know
62	How do you plan reading instruction?	By myself	With teachers at my grade level	I Don't Plan for Reading Instruction
63	How do you plan for reading?	Short term	Long term	Both
64	Do the instructional materials (approach) you use fit with your philosophy of reading?	Yes	No	Don't Know
65	Does each teacher at your grade level teach reading?	Yes	No	Don't Know
66	Do children switch to other teachers for reading, perhaps by reading ability level?	Yes	No	Don't Know

Reading Methods (Please indicate with a check √ how often you use these techniques.)

	Never	Some	Often	Daily		Never	Some	Often	Daily
Teacher Directed Whole Group					Teacher Directed Small Group Guided at group's instructional level				
Individualized (one to one)					Vocabulary Development				
Independent Reading (at instructional level)					Reciprocal Teaching (small groups practice critical reading strategies)				
Question and text discussion					Read and Retell				
CRISS (CReating Independence through Student-owned Strategies)					Read Alouds (teacher reads challenging material to whole group)				
Accelerated Reader software program					Buddy Reading (students read to each other)				
Marginal note-taking (students write notes in margins)					Graphic Organizers				
Multi-sensory activities (auditory listening devices, tactile, kinesthetic, realia)					Dramatization (role play, music/dance, poetry, puppets)				
Brainstorming									

VITA

KRISTINE L. DITTMAR

June 27, 1951	Born, Iola, Wisconsin, USA
1973	B.S., University of Wisconsin Madison, Wisconsin Major: Child Development and Preschool/Kindergarten Education
1977	M.S., Florida International University Miami, Florida Major: Early Childhood Education
1980	M.S., Florida International University Miami, Florida Major: Diagnostic Teaching – Specific Learning Disabilities

Professional Experience

1973-1984	Teacher Miami-Dade County Public Schools Miami, Florida
1984-1987	Exceptional Student Education Placement Specialist Miami-Dade County Public Schools Miami, Florida
1987-1995	Assistant Principal Miami-Dade County Public Schools Miami, Florida
1995-Present	Principal Miami-Dade County Public Schools Miami, Florida
2004	ACCESS Center V Principal of the Year Miami-Dade County Public Schools Miami, Florida
2004	State of Florida Region 5 Principal of the Year Finalist Marco Island, Florida

PRESENTATIONS

Dittmar, K. and Jaureguizar, M. (2003, September). *Kendale Elementary School: Our Road Map to Excellence.* Nova Southeastern University Education Express Bus Tour, Miami, Florida.

Dittmar, K. (2003, October). *Know When To Let Go—Diversified Strategic Leadership Is The Key.* Kendale Elementary School Sterling Best Practices Conference, Miami, Florida.

Dittmar, K., Rivas, E., Viera, C. and Caballero, K. (2004, March). *Meeting School Site Challenges Through Systematic Problem-Solving Approaches and Innovative Solutions.* National Association for Supervision and Curriculum Development Conference, New Orleans, Louisiana.

Dittmar, K. (2004, June). *Shared Leadership: Developing and Empowering Leaders Throughout Your Organization.* Florida Sterling Conference for Performance Excellence, Orlando, Florida.

Dittmar, K. (2004, August). *Shared Leadership: Developing and Empowering Leaders Throughout Your Organization.* Miami Dade County Community Action Agency Conference, Miami, Florida.

Dittmar, K. (2004, October). *Continuous Improvement Through Shared Leadership.* Florida Association for Supervision and Curriculum Development Conference, Tampa, Florida.

Dittmar, K. and Webster, E. (2004, December). *Take the "I, I, I" Out of Teamwork.* National Blue Ribbon Schools Blueprint for Excellence Conference, Myrtle Beach, South Carolina.

Dittmar, K. and Webster, E. (2005, April). *Continuous Improvement Through Shared Leadership.* National Association for Supervision and Curriculum Development Conference, Orlando, Florida.