

THE RELATIONSHIP OF 10TH-GRADE DISTRICT PROGRESS MONITORING
ASSESSMENT SCORES TO FLORIDA COMPREHENSIVE ASSESSMENT TEST
SCORES IN READING AND MATHEMATICS FOR 2008-2009

by

MARILYN K. UNDERWOOD
B.A. Rider University, 1977
M.Ed. University of Florida, 1988
Ed. S. University of Florida, 1994

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Education
in the Department of Educational Research, Technology and Leadership
in the College of Education
at the University of Central Florida
Orlando, Florida

Summer Term
2010

Major Professor: Rosemarye Taylor

© Marilyn Karin Underwood

ABSTRACT

The focus of this research was to investigate the use of a district created formative benchmark assessment in reading to predict student achievement for 10th-grade students on the Florida Comprehensive Assessment Test (FCAT) in one county in north central Florida. The purpose of the study was to provide information to high school principals and teachers to better understand how students were performing and learning and to maximize use of the formative district benchmark assessment in order to modify instruction and positively impact student achievement. This study expanded a prior limited study which correlated district benchmark assessment scores to FCAT scores for students in grades three through five in five elementary schools in the targeted county. The high correlations suggested further study. This research focused on secondary reading, specifically in 10th grade where both state and targeted county FCAT scores were low in years preceding this research. Investigated were (a) the district formative assessment in reading as a predictor of FCAT Reading scores, (b) differences in strength of correlation and prediction among student subgroups and between high schools, and (c) any relationships between reading formative assessment scores and Mathematics FCAT scores. An additional focus of this study was to determine best leadership practices in schools where there were the highest correlations between the formative assessment and FCAT Reading scores. Research on best practices was reviewed, and principals were interviewed to determine trends and themes in practice. Tenth grade students in the seven Florida targeted district high schools were included in the study. The findings of the study supported the effective use of formative assessments both in instruction and as

predictors of students' performance on the FCAT. The results of the study also showed a significant correlation between performance on the reading formative assessment and performance on FCAT Mathematics. The data indicated no significant differences in the strength of correlation between student subgroups or between the high schools included in the study. Additionally, the practices of effective principals in using formative assessment data to inform instruction, gathered through personal interviews, were documented and described.

This dissertation is dedicated to my husband whose pride in me serves as my greatest motivation, my mother whose encouragement of my efforts gives me confidence, and Spats & Sissy who “purr” me on.

ACKNOWLEDGMENTS

I would like to thank my advisor, dissertation chair, and mentor, Dr. Rosemarye Taylor, without whose consistent positive perspective, encouragement, early morning emails, and countless hours, I could not have been successful. To my committee, Drs. Walter Doherty, Janet McGee, and Conrad Katzenmeyer, I give my sincerest thanks and appreciation for the time and efforts you devoted to my dissertation. I count myself privileged to have taken your courses and worked with you in this study.

I offer special thanks to Dr. William Gaught who continued to support me after his initial review of elementary school data. I knew you were always an email away and I appreciate your advice, mentoring, and teaching.

To my Orlando family, CL, LR, JLG, Mammaw, and the Kitties, I give love and thanks for your pride and encouragement through this process. You never doubted that I would complete this and your continued support and love means so much more than you can ever know. Sincere appreciation to my wonderful Staff Development department, most especially Lisa and Tricia, for their patience, understanding, editing help, and support. To Dr. Mary Ann Lynn, I can only say that without your help I would still be trying to figure out Microsoft Office 07. Your patience and skill made it perfect!

Finally, to my husband and best friend who is always my greatest cheerleader and best problem solver, thank you for putting up with the late nights, hours away from home, time taken from you, and meals eaten alone.

TABLE OF CONTENTS

LIST OF FIGURES	x
LIST OF TABLES	xi
CHAPTER 1 THE PROBLEM AND ITS CLARIFYING COMPONENTS.....	1
Introduction.....	1
Purpose of the Study	2
Conceptual Framework.....	3
Accountability in Education	3
Florida Comprehensive Assessment Test	5
Formative Assessments as Progress Monitoring Tools	6
Reading Proficiency Related to Overall Academic Achievement	10
Leadership and the Effective Use of Assessment Data	13
The Targeted School District, Florida Research Study	17
Operational Definitions.....	18
Research Questions.....	18
Methodology and Procedures	19
Study Population.....	22
Data Collection and Instrumentation	22
Data Analysis	24
Delimitations of the Study	25
Limitations of the Study.....	25
Significance of the Study	26
Summary.....	26
CHAPTER 2 REVIEW OF LITERATURE AND RELATED RESEARCH	27
Introduction.....	27
Accountability in Education	28
Florida Comprehensive Assessment Test	33
Formative Assessments as Progress Monitoring Tools	36
Reading Proficiency Related to Overall Academic Achievement	48
Leadership and Effective Strategies for Progress Monitoring Programs.....	54
Summary	66
CHAPTER 3 METHODOLOGY AND PROCEDURES.....	68
Introduction.....	68
Research Questions.....	68
School District Demographics	69
Study Population.....	70
Instrumentation	74
District Benchmark Reading Assessment (DBMRA).....	74
Florida Comprehensive Assessment Test (FCAT)	75

High School Principal Interview Questions.....	77
Data Collection	77
Personal Interviews.....	78
Data Analysis	79
Summary.....	81
CHAPTER 4 ANALYSIS OF DATA	82
Introduction.....	82
Purpose of the Study	82
Research Questions.....	83
Description of Sample Population	84
Research Question 1	86
Research Question 2	89
Research Question 3	91
Research Question 4	94
Research Question 5	97
Summary.....	103
CHAPTER 5 SUMMARY, CONCLUSIONS, IMPLICATIONS, AND RECOMMENDATIONS.....	104
Introduction.....	104
Purpose of the Study	104
Summary and Discussion of the Findings	105
Research Question 1	105
Research Question 2	107
Research Question 3	109
Research Question 4	110
Research Question 5	112
Implications and Recommendations for Practice	118
Recommendations for Future Research	120
Summary.....	122
APPENDIX A SCHOOL DISTRICT APPROVAL.....	123
APPENDIX B DISTRICT BENCHMARK READING ASSESSMENT (DMBRA) SUMMARY REPORT.....	125
APPENDIX C ADMINISTRATOR INTERVIEW QUESTIONS.....	130
APPENDIX D INSTITUTIONAL REVIEW BOARD APPROVAL.....	132
APPENDIX E PRINCIPALS' INFORMED CONSENT.....	134
APPENDIX F SUMMARY OF RESEARCH.....	136

APPENDIX G CONSTANT COMPARISON RESPONSE DETAIL.....	138
LIST OF REFERENCES	142

LIST OF FIGURES

Figure 1. Scatter plot of FCAT Reading scale scores (vertical axis) and DBMRA scores (horizontal axis) for the study population.....	87
Figure 2. Scatter plot of FCAT Mathematics scale scores (vertical axis) and DBMRA scores (horizontal axis) for the study population.....	90
Figure 3. Correlations of DBMRA (x-axis) vs. Reading FCAT (y-axis) by School	92
Figure 4. Correlations of DBMRA (x-axis) vs. Reading FCAT (y-axis) by Subgroup....	95

LIST OF TABLES

Table 1	Research Questions and Sources of Data	21
Table 2	Study Population: Participating High Schools and 10th-Grade Students.....	70
Table 3	Demographics: Percentages of Students Enrolled in Participating Schools.....	72
Table 4	Demographics: Percentages of Students Enrolled in Study Population	73
Table 5	School Grades and Students Meeting High Standards in Reading (2008-2009)	73
Table 6	Study Population: Participating High Schools and 10th-Grade Students.....	84
Table 7	Principal Gender and Years of Experience.....	85
Table 8	Pearson Correlation Coefficients, (r), R^2 , d Values, and Cohen's Standard by School	93
Table 9	Pearson Correlation Coefficients, (r), R^2 , d Values, and Cohen's Standard by Subgroups	96
Table 10	Pearson Correlation Coefficients, (r), R^2 , d Values, and Cohen's Standard for Interview Schools.....	97
Table 11	Constant Comparison Analysis.....	99

CHAPTER 1 THE PROBLEM AND ITS CLARIFYING COMPONENTS

Introduction

In response to National accountability legislations and focus, educators have sought objective methods and data in order to assess and remediate student weaknesses. In addition to identifying student weaknesses, educators have sought to predict the score on the state test based on some valid, reliable instrument. Assessments for the purpose of monitoring student progress have been used in schools and districts to chart student progress and to predict performance on state standardized tests. These formative assessments have been the foundation of progress monitoring programs which track student proficiency on key skills in core content areas in short and frequent intervals for the purpose of remediating learning gaps and to increase performance on high stakes state tests. The stringent and inflexible accountability measures generated by the No Child Left Behind (NCLB) Act of 2001 (U.S. Department of Education, 2009) and the performance gaps delineated by Adequate Yearly Progress (AYP) reports, has prompted a strong response among educational leaders to implement progress monitoring and formative assessments prior to the state standardized test administration. Additionally, prediction of standardized test scores using cost effective, easily generated, and readily available formative assessments has been indicated. More importantly, there has emerged a continuous focus on the successful use of the data generated by such assessments and the strategies that effective leaders use to facilitate student achievement based on the assessments.

This research was conducted in the Marion County Public Schools in Florida. The District Benchmark Reading Assessment (DBMRA) was correlated to the Florida Comprehensive Assessment Test (FCAT) in Reading to determine the extent of correlation. Interviews were conducted with high school principals where there were strong correlations between the DBMRA and FCAT reading. Administrative best practices were documented based on the interviews.

Purpose of the Study

The purpose of the study was to (a) provide information to high school principals and teachers to better understand how students were performing and learning and (b) how to maximize use of the district benchmark assessment in order to modify instruction and positively impact student achievement. An additional focus of this study was to determine best leadership practices in schools where there were strong correlations between DBMRA and FCAT Reading. This study expanded on a 2008 study of third grade students (Gaught) using 2009 data to determine if the District Benchmark Reading Assessment (DBMRA) was a predictor of Florida Comprehensive Assessment Test (FCAT) performance. This study was focused on a larger sample of 10th-grade students. The researcher investigated the potential for predicting the FCAT Mathematics scale scores based on the DBMRA scores. Additionally, the study was conducted to determine whether there was a difference in the strength of prediction of DBMRA and FCAT Reading scale scores between high schools and subgroups.

Conceptual Framework

Accountability in Education

The educational accountability movement has grown into what Bracey (2003) termed a war against the nation's schools. This war has been fueled by public dissatisfaction with the skills of graduates in the workforce and the inability of the school system to reform itself. This dissatisfaction is not new; in fact, a review of legislation indicated the accountability movement in education was initiated in 1988 when, as a result of the Elementary and Secondary Education Act (ESEA), school districts were mandated to evaluate schools based on test scores (Guilfoyle, 2006; US Department of Education, 2009). The ESEA provided financial sanctions for schools which failed to meet state defined performance targets and consequences which required states to publicize school ratings and implement a structured system of reward and corrective actions. Among the required achievements outlined in the ESEA was adequate yearly progress (AYP). Each state was required to quantitatively define AYP and systematically raise the performance level year to year. Not only were schools held accountable under ESEA, but states were responsible for schools and districts under their governance; Districts and schools were sanctioned for performance failures (U.S. Department of Education, 2009).

In 2002, then President George W. Bush signed the No Child Left Behind (NCLB) Act into the law that governed educational accountability. Most educators have agreed that NCLB and accountability are synonymous. The clear and driving force behind

NCLB was standardized assessment as a means to measure achievement and hold educators accountable for the expected achievement of students in grades 3 through 10. This drive created an almost singular focus on teaching what was tested, particularly reading and mathematics, in the nation's schools (Guilfoyle, 2006). Upon enactment of NCLB, existing assessments were reviewed and, almost immediately, test scores were used to rate schools and place them in categories of required improvement. The accountability model grew in scope and focus. The result was the emergence of a tiered system of school improvement designations based on AYP and state assessment scores, and prescribing various levels of intervention (Barton, 2006). NCLB created a positive focus on the achievement of all students. Closing the achievement gap between sub groups emerged as a regular practice in schools across the nation (Weaver, 2006).

NCLB prompted investigation into standardized tests as instruments for accountability. Popham (2008) defined assessment alignment as the ability of the assessment to measure the curriculum goals of the state. Tests for accountability have been intended to determine whether the educational system has adequately prepared students to achieve the level set by the state on the state content standards. According to Popham (2008), in order to accurately measure student performance in this manner, direct alignment between the assessment and the standards was essential. As part of NCLB, states were required to demonstrate and provide evidence of this alignment.

Florida Comprehensive Assessment Test

State standardized tests have varied in scope and difficulty across the nation. Florida's state standardized testing program for accountability grew out of the Educational Accountability Act (Section 229.57, F.S.) of 1971 which mandated a state-wide testing program as the start of the Florida accountability movement. In 1996, the Florida Comprehensive Assessment Test (FCAT) was developed in response to legislation (Section 229.565, F.S.) which identified performance standards and academic benchmarks (Florida Department of Education, 2007).

The goal of the FCAT was to assist Florida in measuring the Sunshine State Standards (SSS) and achieving these higher standards of education for all students in the basic skills of reading, writing, and mathematics. The assessment was administered in the second school semester of the year, usually in March, to students in grades 3 through 10, with score reporting by the state in late May or June. The test produced three scores for each student: (a) the DSS, a criterion referenced score which was developmental, (b) a Scale Score which was criterion referenced, and (c) a norm referenced test (NRT) score. FCAT scores also determined remediation needs and, eventually, whether or not the student would graduate from a Florida public high school (Florida Department of Education, 2007).

FCAT questions have been comprised of both multiple choice and extended response, with the addition of gridded response in mathematics. Students earned points for each question answered correctly. More points were earned for written responses and gridded responses. In a written response question, students were presented with a short

answer question requiring analysis and synthesis skills applied to a reading passage. All students were expected to be tested based on the underlying philosophy of NCLB that all children can learn and make adequate progress. The A+ accountability program in Florida offered monetary rewards to schools that improved performance by a letter grade, improved the performance of the lowest quartile, and/or maintained an “A” school grade which was awarded according to a total number of points earned on both the reading and mathematics portion of the test (Florida Department of Education, 2007).

Formative Assessments as Progress Monitoring Tools

If, however, the goal of assessment is not only to hold educators accountable under NCLB, but also to improve student achievement, then an assessment must be given frequently and used as a gauge of student learning for the purpose of altering instruction to increase student proficiency on content standards (Leahy, Lyon, Thompson, & William, 2005). Standardized assessments do not achieve this goal. The United States Department of Education, in a technical assistance paper for state school governing bodies, advocated that the accountability model prescribed in NCLB should be directly tied to formative evaluations. Popham (2006) defined formative assessments as those that would regularly monitor student achievement in the classroom and provide instructional direction for teachers. Popham’s research indicated that short assessments where results were immediately available significantly impacted student learning, thus improving student performance on standardized state tests.

Formative assessments used to regularly monitor student progress have also been referred to as Curriculum Based Measurement or CBM (Foegen, Jiban, & Deno, 2007; Fuchs & Fuchs, 1999; Wayman, Wallace, Wiley, Ticha, & Espin, 2007). These assessments have been found to be viable in monitoring student progress, determining appropriate instructional interventions, and remediating for mastery (Underwood, 2008). Progress monitoring models emerged as effective educational programs from the research and work of Lezotte (1992) on quality schools. This work was based on the foundation that all students were capable of achieving proficiency and in a quality effective school, the instructional focus was regular measurement of learning and teacher accountability for student growth. Lezotte (1992) further held that student learning, tracked and documented, provided the evidence of effective instruction and the cornerstone of the quality school. In the accountability movement propelled by NCLB, it was determined that, while critical to student learning, a monitoring system or series of formative assessments was not widely implemented. Schmoker (2006) found that in the typical school, assessment for the purpose of monitoring student performance and teacher instructional practices was rare. Generally, assessment was sporadic at best, and the bulk of the learning for the school year was neither assessed nor matched to mastery of skills. Because assessment to analyze growth was rare, little evidence was provided aligning student learning to standards. Additionally, investigation of instruction and best practices was notably absent (Schmoker, 2006).

Researchers (Davenport & Anderson, 2002; Fuchs & Fuchs, 1999) have characterized formative progress monitoring as a systematic program of assessments that

are directly aligned with the curriculum standards, are embedded in the actual classroom structure, and are given frequently so that they provide continuous feedback. The intent of the assessment has been to measure the success of instructional strategies against the student learning gains and to provide a monitoring tool for charting continued growth. Foegen et al. (2007) defined progress monitoring to include CBM as a system of tools providing statistical data that, when utilized effectively, produced positive results in student learning, most particularly in literacy and reading. Foegen et al. found that when assessments were used effectively to monitor student growth, they were successful for increasing achievement in both general education classrooms and for students with disabilities. Smith (2005) supported this contention and indicated such progress monitoring programs provided useful data in mobile populations or in largely heterogeneous classrooms of varying levels.

In their research, Ainsworth and Viegut (2006) described separation of assessments into two categories: those assessments *of* learning and those *for* learning. These researchers indicated that, in most classrooms, assessments have been given for the purpose of assigning grades. These were assessments of learning. The researchers contended that the real purpose of assessment was to inform instruction. Assessments should, therefore, be formative and be assessments for learning. Popham (2003) contended that assessments should also define what teachers were teaching, what students already knew about the content, and what the student must learn in order to show mastery of the content. When the appropriate questions were asked prior to assessment development, large goals and content ideas could be divided into smaller, more

manageable chunks of instruction (Popham, 2003). Formative assessments properly developed and utilized provide what Ainsworth and Viegut called “predictive value as to how students are likely to do on the next level of assessment” (p. 19) or the high stakes tests present in most states (Ainsworth & Viegut, 2006).

Successful programs to monitor student progress are dependent on the quality of the assessments that make up the monitoring system, and thus must meet specific criteria. Lezotte (1992) found that locally generated systems were more readily implemented and more closely matched curriculum than those designed by vendors for a general market. Additionally, the assessment was required to be curriculum based and criterion referenced. Lezotte (1992) contended that formative assessments were expressly intended to measure the student proficiency of the standards and, therefore, were required to be aligned to the curriculum. When measured against Lezotte’s criteria, the assessment that was designed in each school or district had close correlation to the curriculum standards, the courses taught, and the philosophy of the teachers and leaders, leading to credibility and buy-in from teachers. Similarly, Ainsworth and Viegut (2006) stated that the assessments should be common between grade levels and content and that teacher teams of experts should be the assessment designers. Fuchs and Fuchs (1999) found that in order for formative assessments to be effective in measuring student progress, they were required to demonstrate incremental changes longitudinally and provide statistical data capable of informing instruction. Data generated by effective formative assessments enabled the educator to clearly chart growth over a finite period of time. Finally, the assessment had to specifically and adequately describe the learning of

each student and the growth achieved. Researchers at the University of Minnesota determined that assessments had to provide data that were easy to acquire and understand, specific and timely in order to effectively impact teacher decisions and instruction (Wayman et al., 2007).

Ainsworth and Viegut (2006) outlined the characteristics for effective development and utilization of formative assessments. First, the curriculum must be completely aligned, both horizontally and vertically, to insure that the content taught at each level, and hence assessed, was appropriately placed. Similarly, this alignment assured that there were neither gaps nor overlaps in content instruction. To do this, curriculum must be prioritized. Reeves (2004) defined this prioritization as choosing the essential standards from the myriad standards in each content area based on their endurance, leverage, and necessity for the next level of instruction. Reeves (2004) stated that “By applying some criteria to each state standard, educators can add the important dimensions of prioritization and discernment to their state standards, . . . focus[ing]. . . attention on the standards that are most important for academic success” (p. 110). Ainsworth (2004) added that the prioritization must include standards necessary for the state test and the power standards in any content should be limited to the five to seven core concepts that are crucial to enduring understanding and application of information.

Reading Proficiency Related to Overall Academic Achievement

Program improvement has been most successful when the formative assessments and the program improvements focus on the area of reading. According to Schmoker

(1999, 2006), improved reading filtered into other content, causing improved achievement in mathematics, science, and social studies as well. Similarly, accountability for performance has been heavily weighted in the area of reading. Any academic performance improvement system must begin with the area of literacy and reading (Marzano, 2006).

To say that reading and literacy are pre-requisites of academic success seems an obvious statement. This realization preceded the present era of accountability and the focus on literacy in high stakes testing. As early as 1985, the Committee on Reading was formed to study literacy and produced *A Nation of Readers* (Anderson, Hiebert, Scott, & Wilkinson, 1985) which was intended to speak in lay terms to both educators and parents about the importance of reading. That report indicated that the formation of early literacy skills was crucial to future success in academics and productive citizenry. *A Nation of Readers* compared the performance of American students to those in Japan and Taiwan and showed that a disproportionate number of American children were in the lowest 33% in reading performance. The report issued a call to parents to focus on reading in the home to support children in the academic programs delivered in school. The charge for education was to recognize the power of literacy and make it a primary focus of the instructional program (Anderson et al., 1985).

Current literacy struggles have been centered in the secondary school. Investigation of the 2009 FCAT reading for 10th grade state-wide shows a meager 10 points gained in developmental scale score (DSS) as compared to hundreds of points gained in elementary schools at every level. It has been the 10th-grade FCAT that has

determined whether students will graduate from high school. At the time of the present study, many of the state's high school students were not meeting the necessary standards (Florida Department of Education, 2009). Secondary teachers have preferred to labor under the perception that reading is a skill developed and mastered in elementary school. They have isolated this skill from any relevant skill needed to master their subject matter, leaving adolescent learners struggling for meaning and understanding. Teachers at the secondary level have made the false assumption that because students can decode words, they can understand (Coutant & Perchemlides, 2005). The reaction from secondary educators, when it has been determined that students cannot read, has been to provide the information through lecture and notes, making comprehension even more elusive (Darwin & Fleischman, 2005).

Researchers have found that reading ability is the single greatest predictor of achievement in all other core areas. Marzano (2004), in his research, indicated that when students' reading skills and proficiency improved, they performed better on standardized tests in mathematics. Similarly, Reeves (2004) and Ainsworth (2006) found that when instruction targeted reading deficiencies in students, the growth in other content areas was significant, with effect sizes approaching one standard deviation. If proficiency in reading can lead to academic success, it follows that the lack of reading proficiency will result in negative effects for students (Ainsworth, 2006). Taylor and Collins (2003) found a strong relationship between discipline and reading. They found, in reviewing discipline records for middle school students, that students who were chronic discipline issues were reading at or below the 25th percentile.

Author and teacher, Beers (2003), found that not only did lack of skill inhibit performance, but low performance resulted in negative attitudes in students regarding reading. This compounded the spiral of low academic achievement. Without a positive perspective on reading, according to Beers, students were not able to effectively interact with the text to draw meaning. Thus, content skill achievement and assessment performance were hindered (Beers, 2003). Marzano (2004) found that when students were taught vocabulary to increase their reading skill and proficiency, the increase in achievement carried across content areas, building a case for the instruction of vocabulary, a prime reading component, in all academic areas.

Leadership and the Effective Use of Assessment Data

Formative progress monitoring program data when used as a tool to not only monitor student learning progress, but also to analyze teachers' instructional practices, has been shown to positively impact student achievement. Schmoker (2006) found that adjustments in instructional practices yielded results when informed by effective, frequent formative assessments that were directly aligned to the prescribed curriculum. Taylor and Collins (2003) suggested that it was important to align the curriculum, classroom assessments, and standardized tests. Without such alignment, the data gathered was not effective in suggesting instructional modifications. Similarly, Irvin, Meltzer, and Dukes (2007) held that school culture contributed in large part to the effective use of assessment data and the modification of the educational program as a result of the data. Without a collaborative culture of investigation of practices in order to

produce continual improvement, formative assessment data was treated as merely another grade and not a tool for diagnosis of learning.

Ainsworth and Viegut (2006) found that teachers who routinely utilize the data from formative assessments are able to effectively structure and modify instruction to meet the needs of students. Information that students already know is clearly identifiable and this identification eliminates re-teaching content. Teachers also immediately see trends in learning and learning gaps for the purpose of remediation and are able to make appropriate modifications to the instructional plan. Individual student differences are also highlighted, enabling differentiation of instruction to meet each student's need. Finally, regular use of common formative assessments serves to provide feedback to the student concerning academic progress (Ainsworth & Viegut, 2006).

Data analysis proficiency was found to be critical to the instructional impact of progress monitoring assessments for students and the remediation of learning gaps for both the student and subgroups of students (Irvin et al., 2006). Fuchs and Fuchs (1999) outlined several criteria that defined the data collection and analysis of effective progress monitoring systems. First, programs and assessments to monitor and improve student data were directly aligned to standards and required a measure of mastery which would assist in defining students' needs for teachers. Additionally, the assessment reporting clearly identified the relationship between the instructed content and the performance proficiency. Finally, successful systems allowed measurement and data collection for potentially large numbers of students. With general regard to progress monitoring programs, researchers contended that accurate and systematic data collection allowing for

routine measurement of student proficiency over time and effective analysis of that data enabled teachers to prescribe appropriate instructional interventions and remediation. Thus, the existence of the assessment, however effective, was not sufficient to impact achievement. Analysis and understanding of the data was necessary for impacting student growth and building instructional best practices (Fuchs & Fuchs, 1999).

Marzano, Waters, and McNulty (2005) stated that the single greatest indicator of an effective, high performing school was the quality of the school level administrator. Thus, the principal's mission is to use the assessments designed to monitor and improve achievement positively and proactively in regular practice in the school. According to Firestone (2009), the role of the principal as leader in the effective use of formative assessments to improve student performance on state standardized tests was critical. Firestone identified the leadership role as being a liaison between an accountability culture and a student learning culture. In an accountability culture, testing and central control would be evident. In a student learning culture, students would be considered a priority and teachers would be assumed to possess the professional skill and desire to help all students. Strong leadership from the top and mutual respect among the adults in the school building would be required. Firestone defined the student learning culture as one where there was shared responsibility for student growth and personal professional development. Senge (2006) referred to such a culture as the "fifth discipline" within the school and one that was crucial to the success of the institution and the individuals functioning within it. Senge characterized the successful organization as one "that discover[s] how to tap people's commitment and capacity to learn at all levels. . . ." (p.

4). Such organizations thrive on individuals who possess personal mastery, have great dedication to learning and a commitment to growth of every member of the organization. This growth was viewed as the norm for personal satisfaction and fulfillment, not simply for accountability. With such a learning culture modeled by the leader, Senge held that organizations could realize maximum fulfillment of the highest goals and achievements.

Strategies for successful leaders have been outlined by many current researchers. Schlechty (1997) described leadership in an environment of change as managing results and motivating others to manage the results in order to change. Fullan (2001) described five practices successful leaders share: leading with moral vision, building relationships, managing change, sharing knowledge, and building the bridge between new knowledge or practices and old paradigms. Marzano et al. (2005) cited 21 behaviors employed by successful principals. Among these, vision, culture, communication, and involvement were identified as keys to leading a successful school driven on student achievement. Monitoring and feedback were essential to improving student performance, and Marzano et al. (2005) advocated an approach where research and sharing data were as important as a hands-on approach to leading. Cotton (2003) found that the principal's vision and actions in conveying high expectations promoted a positive learning climate and a school culture that valued student achievement. Similarly, visible leadership of an approachable principal contributed significantly to school achievement. An underlying tenet of such leadership was the empowerment of teachers to make viable decisions about learning and school focus and shared leadership responsibilities in the school (Cotton, 2003).

The Targeted School District, Florida Research Study

This research was intended to build on Gaught's 2008 study of the relationship between student performance, formative assessment and the Florida Comprehensive Assessment Test (FCAT). Gaught conducted his research in six elementary schools in the public school district targeted in the present research. The two objectives of the study were to determine if a relationship existed in third, fourth, and fifth grades at the six schools between the county's assessments scores in mathematics and reading and the FCAT score in mathematics and reading. If a relationship existed, the goal was to create a linear regression model to predict FCAT performance based on the county's assessments scores (Gaught, 2008).

Reading and mathematics scores used for the study were the 2007 November county assessment percentage score in reading and mathematics and the 2008 scale score in FCAT reading and mathematics. Gaught also looked at one high school's data in grades nine and ten to determine if a similar correlation could be found at higher grade levels, however, did not complete a study for this school and these grades.

In the limited study, Gaught (2008) found that there was a relationship between student performance on the public school district's created formative assessment, (DBMRA) and the Florida Comprehensive Assessment Test (FCAT). This researcher conducted a consultation with Gaught to determine whether there was similar support for conducting a study using high school data. This resulted in a recommendation that further study would be reasonable (W.L. Gaught, personal communication, 2009). The

researcher in the present study has sought to extend Gaught's work to the secondary level and to a larger sample of schools.

Operational Definitions

Following are operational definitions developed by Underwood (2008).

Reading: This term relates only to the skills and competencies as identified by the State of Florida in nine (9) benchmarks that are tested in grades 3-10 on the FCAT.

Benchmark: A standard identified as the goal for student achievement in a specific skill area of a particular content area.

District Benchmark Reading Assessment (DBMRA): Criterion referenced regular assessment that measures the learning for each student in grades 3-10 in reading against the State identified reading benchmarks. Assessments mirror the format of the FCAT.

High School: Public schools, grades 9-12, with 900 or more students.
Florida Comprehensive Assessment Test (FCAT): Standardized test given yearly in March in grades 3-10 to determine proficiency in reading, mathematics, and science. The test is based on the state identified tested benchmarks in each content area.

Scale Score: A score which, for FCAT, can range from 100-500. The scale is used to determine the student's achievement level as measured against curriculum benchmarks (criterion) in each content area tested and is not based on growth (p. 5).

Research Questions

1. To what extent does the District Benchmark Reading Assessment (DBMRA) predict achievement level on the Florida Comprehensive Assessment Test (FCAT) in Reading for 10th grade for 2009?

2. To what extent does the District Benchmark Reading Assessment (DBMRA) predict achievement level on the Florida Comprehensive Assessment Test (FCAT) in Mathematics for 10th grade for 2009?
3. To what extent, if any, does the relationship between scores on the District Benchmark Reading Assessment (DBMRA) and FCAT Reading differ among high schools in the study?
4. To what extent, if any, does the relationship between scores on the District Benchmark Reading Assessment (DBMRA) and FCAT Reading differ by AYP subgroups?
5. What strategies do principals of high schools use where the correlation between the November, 2008 District Benchmark Reading Assessment (DBMRA) and the 2009 FCAT score in reading fall into or above Cohen's medium effect ($d \geq .50$)?

Methodology and Procedures

This study employed a non-experimental research design using secondary school data collected from the targeted public school district in Florida. The study extended the preliminary research initiated in a limited study conducted in 2008 by Gaught which used data from the 2007-2008 November DBMRA and FCAT in six selected elementary schools in the targeted county. This study was extended to the secondary level and focused on 10th-grade reading using the DBMRA 10th-grade student scores as predictors of FCAT Reading and Mathematics scores across the seven comprehensive public high schools in the targeted county. Because the correlation of reading to achievement in

other areas has been noted by researchers, the correlation of the DBMRA to FCAT Mathematics scores using the November 2008 DBMRA scores and the 2009 FCAT Mathematics scale scores was also investigated. The research questions, data sources and variables which guided the study are displayed in Table 1.

Table 1
Research Questions and Sources of Data

Research Question	Data Sources	Variables Tested
1. To what extent does the District Benchmark Reading Assessment (DBMRA) predict achievement level on the Florida Comprehensive Assessment Test (FCAT) in Reading for 10th grade for 2008?	District Benchmark Reading Assessment (DBMRA)	DBMRA raw score percentage (Independent)
	Florida Comprehensive Assessment Test Reading scale score (FCAT)	FCAT Reading scale score (Dependent)
2. To what extent does the District Benchmark Reading Assessment (DBMRA) predict achievement level on the Florida Comprehensive Assessment Test (FCAT) in Mathematics for 10th grade for 2008?	District Benchmark Reading Assessment (DBMRA)	DBMRA raw score percentage (Independent)
	Florida Comprehensive Assessment Test Mathematics scale score (FCAT)	FCAT Mathematics scale score (Dependent)
3. To what extent, if any, does the relationship between scores on DBMRA and FCAT Reading differ among high schools in the study?	District Benchmark Reading Assessment (DBMRA)	DBMRA raw score percentage (Independent)
	Florida Comprehensive Assessment Test Reading scale score (FCAT)	FCAT Reading scale score (Dependent)
	District High school list	
4. To what extent, if any, does the relationship between scores on the District Benchmark Reading Assessment (DBMRA) and FCAT Reading differ by AYP subgroups?	District Benchmark Reading Assessment (DBMRA)	DBMRA raw score percentage (Independent)
	Florida Comprehensive Assessment Test Reading scale score (FCAT)	FCAT Reading scale score (Dependent)
	AYP State Report	
5. What strategies do principals of high schools use where the correlation between the November, 2008 DBMRA and the 2009 FCAT score in reading falls into or above Cohen's medium effect ($d \geq .50$)?	Interviews of principals in schools showing a correlation at or above .50 between the November, 2008 DBMRA and the FCAT 2009 reading score.	

The data to be used in this study were collected through normal testing procedures outlined by district policy in school year 2008/2009. A data request was made to the student information department of the targeted district for the data to be provided in an Excel spreadsheet with student identifiers removed. Linear correlation models were developed across the sample distribution, and individual school associations were investigated. Additionally, personal interviews were conducted with selected principals and reported holistically, with the intent of identifying common trends or themes.

Study Population

The population in this research study were 10th graders from the seven high schools in the district. The sample was comprised of only those 10th-grade students in the entire 10th-grade population in the county who had both a 2008 November DBMRA score and 2009 FCAT Reading and Mathematics scale score.

Data Collection and Instrumentation

This study employed a non-experimental research design using secondary data collected from the targeted public school district in Florida during the 2008-2009 school year. A research approval letter (Appendix A) was secured from the district guidance and testing office. Data were requested from the district's student information system and were provided in an Excel spreadsheet. The data requested included the school assignment, gender, ethnicity, free and reduced lunch status, exceptional education status, English language learner status, DBMRA score, and FCAT scale score for 10th graders

in district high schools. The only stipulation in the data request was that all members of the sample had obtained scores for both the November, 2008 DBMRA and the 2009 FCAT for reading and mathematics.

District Benchmark Reading Assessment (DBMRA)

The District Benchmark Reading Assessment (DBMRA) was designed in 2005 by school district personnel. The assessment contained approximately 50 questions in a multiple choice format and was structured to measure the nine reading benchmarks which were tested on the FCAT (D. L. Greene, personal communication, 2008; Marion County Public Schools, 2007). Appendix B contains the District Benchmark Reading Assessment (DMBRA) Summary Report.

Florida Comprehensive Assessment Test (FCAT)

The Florida Comprehensive Assessment Test (FCAT) has been administered in mid to late March to students in grades 3-10 to assess reading and mathematics. The results of this standardized achievement test define the level of student proficiency targeted for mastery in both content areas up to and including the current grade (Florida Department of Education, 2007). For the purpose of this study, 2009 FCAT scale scores in reading and mathematics were utilized.

High School Principal Interviews

Interviews were conducted with principals of high schools where correlations between DBMRA and FCAT Reading were found to have a medium effect size on

Cohen's standard of effect size (Becker, n.d.; Cohen, 1988). The personal interviews were focused on the activities the principal used with the DBMRA data and were intended to determine trends used by effective principals. Data analysis of principal interview responses utilized Constant Comparison Analysis to identify themes or trends in best practices used by principals. The interview questions can be found in Appendix C.

Data Analysis

The data for this study were analyzed using multiple statistical procedures appropriate to the data and research questions to be answered. To determine the extent of the correlation and prediction of (a) the DBMRA and the FCAT Reading and (b) the DBMRA and the FCAT Mathematics, a Pearson correlation statistic was determined, and a regression test was used. Data analyses were accomplished using SPSS, Inc. Statistics, Version 17.0. To determine the extent to which the relationship between scores on DBMRA and FCAT differed among high schools and between subgroups, Pearson correlation coefficients were compared and evaluated on Cohen's (Becker, n.d.; Cohen, 1988) standard of effect size. Interview responses were analyzed using qualitative analysis, specifically a constant comparison analysis.

Delimitations of the Study

The study had the following delimitations:

1. The data were collected for 10th graders from all public high schools in the targeted county in Florida.
2. Data were collected for all 10th graders who had a DBMRA score and a Florida Comprehensive Assessment Test (FCAT) reading and mathematics scale score.
3. This study included all the high schools in the targeted district with the exception of two that served alternative populations.
4. Data in this study that were analyzed were delimited to that which was retrieved from the targeted county.

Limitations of the Study

The factors which limited the validity of this research included the following:

1. Data analyzed in this study were gathered using a single district assessment.
2. Scale score data used in the study were limited to that obtained from the 2009 FCAT administration in reading and mathematics.
3. The population for this study was comprised of 10th-grade students in public high schools in the targeted district. Students from charter schools in the county were excluded.
4. A sample of district high school principals were interviewed.

Significance of the Study

This study was intended to be of value to the targeted district by validating the district assessment as a viable formative assessment that could predict FCAT reading scores for 10th-grade students. Additionally, the study was intended to identify the administrative best practices that resulted in growth between the district assessment score, predicted FCAT score, and the actual FCAT score. Data from this study should be useful to educational systems in general that are attempting to build a program of formative assessments for the purpose of state test score prediction and student achievement gains as specified by NCLB. The administrative best practices that have been documented can be replicated and modeled by other administrators in high schools. Finally, the result of this study will both add to the current research and serve as a foundation for further research in the use of district assessments as viable formative assessments that could predict achievement on state tests.

Summary

This chapter has provided an introduction to the study along with an overview of the conceptual framework. The research questions that guided the study have been stated, and the methods and procedures that were followed in conducting the study have also been presented. Information on the sources of data (DBMRA, FCAT, and principal interviews) have been explained, and procedures used in data collection and analysis have been outlined. The significance of the study, its delimitations and potential limitations have been stated.

CHAPTER 2 REVIEW OF LITERATURE AND RELATED RESEARCH

Introduction

This chapter has been organized to provide a review of the literature and related research. Topics addressed include: (a) accountability in education, (b) the Florida Comprehensive Assessment Test, (c) formative assessments as progress monitoring tools, (d) reading proficiency related to overall academic achievement, and (e) leadership and the effective strategies for progress monitoring programs. A literature review was conducted investigating the available literature and research on these topics. The University of Central Florida research librarian was engaged on two separate occasions to conduct a thorough search of research surrounding the research questions in this study. Additionally, searches were conducted on the University of Central Florida Library database targeting key words and known author experts in the fields surrounding the research questions. Primary sources were reviewed using (a) reference lists in research obtained and (b) the Web of Science database in the University of Central Florida Library research program. This chapter contains a review of the literature that indicated effectiveness of formative assessments in reading as a tool for student achievement in reading and other content areas. Also discussed are studies investigating the relationship between reading skill and achievement in science and mathematics. Finally, the best practices of leaders in schools where there was student growth were detailed. The literature reviewed and cited served to create a basic foundation for understanding the ideas and concepts surrounding the research questions in this study.

Accountability in Education

The term “accountability” has become ubiquitous in education with the passage of legislation, the proliferation of assessments, and the award or removal of federal education funds, all tied to some definition of accountability for educators and schools. The definition most prevalent for accountability is responsibility, according to Linn (2003). The almost daily changes and legislation impacting education have defined accountability as a tangible body of standards, assessments, and Federal grants or programs designed to hold someone responsible for the state of education (Linn, 1994).

The educational accountability movement has grown into what Bracey (2003) termed a war against the nation’s schools. This war has been fueled by public dissatisfaction with the skills of graduates in the workforce and the inability of the school system to reform itself (Bracey, 2003). In reviewing the literature, evidence was found indicating that many of the reform movements initiated by educational institutions have been unsuccessful and have not yielded lasting change in curriculum or instructional processes (Priestley & Sime, 2005). Public dissatisfaction with education is not new. In fact, a review of legislation indicated the accountability movement in education was initiated in 1988 when, as a result of the Elementary and Secondary Education Act (ESEA), school districts were mandated to evaluate schools based on test scores (Guilfoyle, 2006; US Department of Education, 2009). The ESEA provided for (a) financial sanctions for schools that failed to meet state defined performance targets and (b) consequences which required states to publicize school ratings and implement a structured system of reward and corrective actions. Among the required achievements

outlined in the ESEA was adequate yearly progress (AYP). Each state was required to quantitatively define AYP and systematically raise the performance level of students year to year. Not only were schools held accountable under ESEA, but states were responsible for schools and districts under their governance; Districts and schools were sanctioned for performance failures (U.S. Department of Education, 2009).

Accountability movements changed dramatically after the mid 1980s, the ESEA, and the introduction of The Goals 2000: Educate America Act of 1994. With this legislation, the presence of the federal government in the realm of education was much stronger. The Goals 2000 program developed national panels and councils to develop content and performance standards on both the state and national levels. It also served to assist and judge state standards and programs and ensure the strict evaluation of Title I programs. While the program clearly indicated that the participation of states was voluntary, the national operations and development of Goals 2000 standards and programs made participation mandatory in order to receive funds, materials, and assistance. From the guidance offered by Goals 2000, states were charged with developing specific performance and content standards in each core content area for each grade level. At the same time, the assessments for (a) measuring the attainment of the national standards and (b) defining accountability for educators and students were created by the states (Linn, 1994).

In 2002, then President George W. Bush signed the No Child Left Behind (NCLB) Act into the law that governed educational accountability. Most educators would agree that NCLB and accountability are synonymous. The clear and driving force

behind NCLB was standardized assessment as a means to measure achievement and hold educators accountable for the expected achievement of students in grades 3 through 10. This drive created an almost singular focus on teaching what was tested, particularly reading and mathematics, in the nation's schools (Guilfoyle, 2006). NCLB increased the number and frequency of summative tests to one test per year in each school year from grade three through eight with a requirement that the student be tested at least once during high school. The goal of these summative assessments was to measure student performance against a predefined set of standards, or criteria, to determine level of proficiency. These assessments, because of their focus on accountability, were designed to be given as late in the year as possible, ensuring that students had been exposed to the most instruction on each standard (Perie, Marion & Gong, 2009). Upon enactment of NCLB, existing assessments were reviewed and, almost immediately, test scores were used to rate schools and place them in categories of required improvement. The accountability model grew in scope and focus. The result has been the emergence of a tiered system of school improvement designations based on AYP and state assessment scores, and prescription of various levels of intervention (Barton, 2006).

The original goal of NCLB was to improve student learning. The system embedded in the legislation provided for accountability and high stakes testing to improve achievement. The goal and the means seem to have been at odds. The volume of summative assessments, given late in the year and used for ranking or grading purposes, precluded use of data for improving instruction or programs. There were no diagnostic assessments or reports built into NCLB that would have furthered the original

goal. Instead, scores reflected total performance on large groups of standards without information on specific skills needed for instruction to improve achievement for each student. At the same time, NCLB held, as its basic tenets, that all students could learn and that education was charged with the responsibility to remediate each student so that proficiency was achieved and learning gaps among sub groups could be closed. The system did not provide the formative data through interim assessments necessary for educators to address student learning and provide appropriate instruction to close achievement gaps (Perie et al., 2009).

While there were numerous problems inherent within the NCLB system, NCLB did create a positive focus on the achievement of all students. The concept of learning for all was brought to the forefront of education, and schools were forced to consider those students that previously might have been regarded as incapable of grade level work. Closing the achievement gap between sub groups emerged as a regular practice in schools across the nation (Weaver, 2006). The targets that were set by NCLB prompted educators to find ways to improve achievement for every student and brought an outgrowth of smaller, classroom assessments used to define students' current state of proficiency, the intended standard of proficiency, and the distance between the two (Nichols, Meyers & Burling, 2009).

NCLB was a national reform movement designed to place American students in line with the advanced students of other countries. As early as 1998, Bracey's report concluded that the American obsession with mathematics and science and the comparison of American student scores with countries like Japan would lead American legislators

and educators to make incorrect assumptions (Bracey, 1998). After the first year of NCLB, President Bush cited the growth of math and reading scores on the 2003 National Assessment of Educational Progress (NAEP) as an indicator that NCLB was working. Researchers contended that this growth may have been a result of previous reforms implemented by schools in response to Goals 2000 (Loveless as cited in Fuller, Gesicki, Kang, & Wright, 2006).

The question of whether NCLB was successful still remains. Based on NAEP scores released in 2005, indications were that the progress slowed among fourth, eighth, and twelfth graders tested. Scores were, at best, flat with little growth indicated (Fuller et al., 2006, Fuller, Gesicki, Kang, & Wright, 2007). The United States Department of Education (2009) presented similar data in a longitudinal study of NAEP scores from 1971-2008. In this study, indications were that scores for fourth, eighth, and twelfth graders improved little from 1971 to 2008. Across the years, scores remained flat, showing little or no increase with the advent of NCLB, and in reading, showing a decline for twelfth graders (US Department of Education, 2009). Because NCLB was so heavily focused on standardized test scores, this then brought into question the adequacy of assessments as a tool to measure student success in small increments over time (Fuller et al., 2006, 2007).

NCLB prompted investigation into standardized tests as instruments for accountability. The apparent lackluster data surrounding the longitudinal NAEP data suggested that when narrowly defined parameters were attached to the accountability program, instruction became very narrowly focused and the assessment became the

driving force of education (Linn, 2003). Fuller et al. (2007) questioned reliance on the summative standardized test to assess student learning. Instead, they held that the focus must be placed on assessment over time to determine content and performance achievement in alignment with state or national standards (Fuller et al., 2007). Popham (2008) defined assessment alignment as the ability of the assessment to measure the curriculum goals of the state. Tests for accountability have been intended to determine whether the educational system has adequately prepared students to achieve the level set by the state on the state content standards. According to Popham (2008), in order to accurately measure student performance in this manner, it is essential that there is a direct alignment between the assessment and the standards. As part of NCLB, states have been required to demonstrate and provide evidence of this alignment.

Florida Comprehensive Assessment Test

State standardized tests have varied in scope and difficulty across the nation. Florida's state standardized testing program for accountability grew out of the Educational Accountability Act (Section 229.57, F.S.) of 1971 which mandated a state-wide testing program as the start of the Florida accountability movement. In 1974 and 1976, the Educational Accountability Act was modified and expanded to include reading, writing and mathematics for students in grades 3, 5, 8, and 11. A requirement for passing the State test for the graduating class of 1979 was also added as a prerequisite for a high school diploma. In 1991, as a result of Blueprint 2000, schools in Florida operated under stricter accountability and testing provisions designed to identify and reward high

performing schools while prescribing interventions for lower performing schools. During the period between 1974 and 1995, Florida's State assessment underwent substantive revision in scope and depth. Finally, in 1996, the Florida Comprehensive Assessment Test (FCAT) was developed in response to legislation (Section 229.565, F.S.) which identified performance standards and academic benchmarks (Florida Department of Education, 2007).

The goal of the FCAT was to assist Florida in measuring the Sunshine State Standards (SSS) and achieving these higher standards of education for all students in the basic skills of reading, writing, and mathematics. The assessment was administered in the second school semester of the year, usually in March, to students in grades 3 through 10, with score reporting by the state in late May or June. The test produced three scores for each student: (a) the DSS, a criterion referenced score which was developmental, (b) a Scale Score which was criterion referenced, and (c) a norm referenced test (NRT) score. The developmental scores (DSS) ranged from 0-3000 and tracked student growth over time, allowing comparison between years and measuring adequate yearly progress and learning gains required by the State. The scale scores ranged from 100-500 and were divided into five levels (1-5, with level 3 being required for proficiency). FCAT scores also determined remediation needs and, eventually, whether or not the student would graduate from a Florida public high school. The cost for developing the FCAT was reported to be over \$40 million in 2007 (Florida Department of Education, 2007).

The FCAT was designed to measure the SSS which were adopted by the State Board of Education and mandated to be part of every school's curriculum. Teaching to

the standards in every classroom was determined to be essential to student success on the standards based test. The premise underlying the FCAT was that quality teaching of standards in every classroom would result in high achievement on the FCAT. Teaching to the test was not necessary as the test embodied the standards. Thus, teaching the standards would produce proficiency. In fact, the state discouraged educators from spending instructional time on activities intended to prepare or practice for FCAT (Florida Department of Education, 2007).

FCAT questions have been comprised of (a) multiple choice, (b) extended response, and (c) and gridded responses in mathematics. Questions were written by a professional team of writers who used comprehensive item specifications that closely related to the standards tested. Students earned points for each question answered correctly. More points were earned for written responses and gridded responses. In a written response question, students were presented with a short answer question requiring analysis and synthesis skills applied to a reading passage. The gridded response was associated with a mathematics question where students were asked to record their response on a bubble grid with the correct placement of all numbers and decimals.

All students were expected to be tested based on the underlying philosophy of NCLB that all children can learn and make adequate progress. The A+ accountability program in Florida offered monetary rewards to schools that improved performance by a letter grade, improved the performance of the lowest quartile, and/or maintained an “A” school grade which was awarded according to a total number of points earned on both the reading and mathematics portions of the test (Florida Department of Education, 2007).

Formative Assessments as Progress Monitoring Tools

Leahy, Lyon, Thompson and William (2005) discussed the multiple goals of assessment. They thought that if the goal of assessment was not only to hold educators accountable under NCLB, but also to improve student achievement, assessments needed to be given frequently. These assessments could be used as a gauge of student learning for the purpose of altering instruction to increase student proficiency on content standards. Standardized assessments do not achieve this goal. The United States Department of Education, in a technical assistance paper for state school governing bodies, advocated that the accountability model prescribed in NCLB should be directly tied to formative evaluations (n.d., 2009). Popham (2006) defined formative assessments as those that would regularly monitor student achievement in the classroom and provide instructional direction for teachers. Assessment research indicated that short assessments where results were immediately available significantly impacted student learning, thus improving student performance on standardized state tests (Popham, 2006).

Formative assessment has also been defined as an active process where the teacher and student intentionally pair to systematically gather evidence in order to improve learning (Moss & Brookhart, 2009). The primary purpose of an assessment is to improve teaching and, thus, learning. Only assessments at intervals, or formative assessments that consider growth over time in multiple perspectives, were found to significantly impact learning (International Reading Association & the National Council of Teachers of English, 2010). Formative assessments have emerged as tools to check understanding in a developmental mode rather than use of a summative or state

achievement test (Fisher & Frey, 2007). Perie et al. (2009) distinguished formative assessments from other kinds of assessments indicating that formative assessments were not only diagnostic to determine where students stood in their current learning, but they were also at the classroom level and embedded into the current learning so that the assessment was not far removed from the actual instruction. Black and Wiliam (1998a, 1998b) held that an assessment became formative when its sole purpose was to modify instruction in order to meet the identified needs of the students. Critical to the success of formative assessments was the feedback to the student and teacher. The formative assessment was not a stand-alone tool, but rather part of the larger instructional and learning program to assist with proficiency in meeting standards or prescribed learning benchmarks (Black & Wiliam, 1998a, 1998b). Wiliam (2006) stated that the key factor in making an assessment formative was the data and interpretation of that data for use in instruction. Formative assessments used to regularly monitor student progress have also been referred to as Curriculum Based Measurement (CBM) by some researchers (Foegen, Jiban, & Deno, 2007; Fuchs & Fuchs, 1999; Wayman, Wallace, Wiley, Ticha, & Espin, 2007). These assessments have been found to be viable in monitoring student progress, determining appropriate instructional interventions, and remediating for mastery (Underwood, 2008).

Formative assessments or CBM are valuable tools in a progress monitoring model where the assessment informs instruction by reporting aggregate scores in specific skills. This model then prescribes remediation or enhancement and is again monitored for progress using the formative assessment (Perie et al., 2009). Researchers (Davenport &

Anderson, 2002; Fuchs & Fuchs, 1999) characterized formative progress monitoring as a systematic program of assessments that are directly aligned with the curriculum standards, are embedded in the actual classroom structure, and are given frequently so that they provide continuous feedback. The intent of the assessment has been to measure the success of instructional strategies against the student learning gains and to provide a monitoring tool for charting continued growth.

Progress monitoring models emerged as viable educational programs from the research and work of Lezotte (1992) on quality schools. This work was based on the foundation that all students were capable of achieving proficiency, and in a quality effective school, the instructional focus was regular measurement of learning and teacher accountability for student growth. Lezotte (1992) further held that student learning, tracked and documented, provided the evidence of effective instruction and the cornerstone of the quality school. In the accountability movement propelled by NCLB, it was determined that, while critical to student learning, a monitoring system or series of formative assessments was not widely implemented. Black and Wiliam (1998a, 1998b), in their early synthesis of existing research surrounding formative assessment, found compelling evidence that formative assessment, appropriately implemented in a progress monitoring model, significantly impacted student achievement.

Schmoker (2006) found that in the typical school, assessment for the purpose of monitoring student performance and teacher instructional practices was rare. Generally, assessment was sporadic at best, and the bulk of the learning for the school year was neither assessed nor matched to mastery of skills. Because assessment to analyze growth

was rare, little evidence was provided aligning student learning to standards.

Additionally, investigation of instruction and best practices was notably absent (Schmoker, 2006).

Elwood found studies in England that indicated a lack of formative assessment progress monitoring programs implemented either regularly or as part of an entire district, school, or grade level assessment initiative (Elwood, 2006). Black and Wiliam (1998a, 1998b) cited several reasons that formative assessments were not widely used in classrooms. First, the assessment for formative purposes was not understood by teachers, thus not implemented or used for purposes other than grades. Second, the national requirements for assessments tied to accountability exerted a strong influence on the use of assessments in the classroom. Finally, implementation of assessment for learning, or formative assessments, required a dramatic paradigm shift in the roles of both students and teachers in the interpretation of scores and feedback for the learning program (Black & Wiliam, 1998a, 1998b).

The goal of progress monitoring has been to assess instruction and student learning and to monitor the continuous growth of the student. In order to accomplish this goal and produce useful and meaningful results, the assessment and monitoring program needed to correlate with standards (Davenport & Anderson, 2002; Fuchs & Fuchs, 1999). Key to effective use of formative assessments for the purpose of increasing student achievement was feedback. This feedback occurred between the test results and the educator to identify gaps between what the student knew and what was required. The interpretation of the test data provided a prescription for instruction. The feedback was

also between the educator and the student so that students became active in the learning process and fully aware of the steps to be taken. For the assessment to truly be a formative one, the feedback had to be used to close the gap of knowledge for the student (Heritage, Kim, Vendlinski & Herman, 2009).

Foegen et al. (2007) defined progress monitoring to include CBM as a system of tools providing statistical data that, when utilized effectively, produced positive results in student learning, most particularly in literacy and reading. Foegen et al. found that when assessments were used effectively to monitor student growth, they were successful for increasing achievement in both general education classrooms and for students with disabilities. Smith (2005) supported this contention and indicated such progress monitoring programs provided useful data in mobile populations or in largely heterogeneous classrooms of varying levels.

Fuchs, Fuchs, and Zumeta (2008) found that progress monitoring using formative assessments in the content areas of mathematics and reading was effective in terms of raising student achievement and guiding instructional modification. Scriven (as cited in Newton, 2007) held that formative evaluation systems administered assessments more than once in relation to a program of study or unit of learning for the express intent of improvement. In a study conducted by Hintze, Christ, and Methe (2006), formative assessments and progress monitoring were described as assessments that defined proficiency in a particular skill in the curriculum. The progress monitoring system that the researchers studied delineated the wholesale curricula and standards into short-term

instructional objectives for the purpose of program improvement and student achievement.

In their research, Ainsworth and Viegut (2006) described separation of assessments into two categories: those assessments *of* learning and those *for* learning. Ainsworth and Viegut indicated that, in most classrooms, assessments had been given for the purpose of assigning grades. These were assessments of learning. The researchers contended that the real purpose of assessment was to inform instruction. Assessments should, therefore, be formative and be assessments for learning (Ainsworth & Viegut, 2006). Stiggins (2002) described assessments of learning as the test given once a year which provided information on student achievement in a broad category. These assessments of learning, while intended to produce data on student performance, did not provide enough data early enough to be diagnostic. Rather, they reflected whole group increases or decreases and were more useful in describing the state of student performance in a school, district, or state. Assessments for learning, on the other hand, provided immediate feedback and monitoring for growth. Such formative assessments were useful in that they enabled the student to see growth and encouraged more learning, particularly for sub groups where there was assessment history of low achievement (Stiggins, 2002). Popham (2003) contended that assessments should also define what teachers were teaching, what students already knew about the content, and what the student needed to know in order to show mastery of the content. When the appropriate questions were asked prior to assessment development, it was possible to break large goals and content ideas into smaller, more manageable chunks of instruction (Popham,

2003). Formative assessments properly developed and utilized provided what Ainsworth and Viegut called “predictive value as to how students are likely to do on the next level of assessment” (p. 19) or the high stakes tests present in most states.

Successful progress monitoring programs are dependent on the quality of the assessments that make up the monitoring system, and thus must meet specific criteria. Lezotte (1992) found that locally generated systems were more readily implemented and more closely matched curriculum than those designed by vendors for a general market. Additionally, the assessment was required to be curriculum based and criterion referenced. Lezotte (1992) contended that formative assessments were expressly intended to measure the student proficiency of the standards and, therefore, were required to be aligned to the curriculum. When measured against Lezotte’s criteria, the assessment that was designed in each school or district had close correlation to the curriculum standards, the courses taught, and the philosophy of the teachers and leaders, leading to credibility and buy-in from teachers.

The explicit statement of fundamental knowledge in the content was found to be critical to the success of formative assessments as progress monitoring tools. The assessment had to be criterion referenced and tied directly to one or more of a set of identified standards required for proficiency in the content. The performance of the student did not depend on performance of other students, rather on his or her own proficiency as it related to the end goal. If the goal of formative assessments and progress monitoring was to teach to mastery, that possibility was enabled only with a

clear statement of required learning in increments specific and small enough to be diagnostically assessed (Lalley & Gentile, 2009).

In a high school study, researchers determined that effective classroom assessments were directly tied to a specific outcome. In order to implement formative assessment progress monitoring programs, teachers were coached to develop clear curriculum learning goals for students. Within those goals, criteria for student work and proficiency were explicitly defined and provided to students (Allen, Ort, & Schmidt, 2009). Such an approach was supported by Wiggins (2005) in backward design. He held that standards informed everything teachers did in the classroom. According to Wiggins, teachers were guided by national, state, or local standards to determine what students must know, and they used this knowledge to develop specific goals that drove the lesson and the learning goals for students (Wiggins, 2005). Once subskills were identified for students, the assessments needed to be small and specific enough to measure the proficiency within the subset. The assessment items also were designed to measure the subset and elicit the performance that would be indicative of proficiency (Hintze et al., 2006).

A challenge researchers found in development of formative assessments for progress monitoring was the inclusion of tasks that not only measured the current content but also included a requirement to synthesize various skills necessary for competency in the content as a whole. Two approaches were found to be successful in insuring that students were able to integrate skills. The first was creating a task that was better correlated than other tasks with the required end-of-course competencies. Such tasks

were shown to successfully elicit performance that integrated essential skills. The second approach was to include the essential tasks in individual assessments and carefully map the instructional calendar and intensity of instruction so that each skill received equal attention and time (Fuchs et al., 2008). Fuchs and Fuchs (1999) found that in order for formative assessments to be effective in measuring student progress, they were required to demonstrate incremental changes longitudinally and provide statistical data capable of informing instruction. Data generated by effective formative assessments enabled the educator to clearly chart growth over a finite period of time. Finally, the assessment had to specifically and adequately describe the learning of each student and the growth achieved.

Ainsworth and Viegut (2006) outlined the characteristics for effective development and utilization of formative assessments. First, the curriculum must be completely aligned, both horizontally and vertically, to ensure that the content taught at each level, and hence assessed, was appropriately placed. Similarly, this alignment ensured that there would be neither gaps nor overlaps in content instruction. To do this, curriculum must be prioritized. Reeves (2004) defined this prioritization as choosing the essential standards from the myriad standards in each content area based on their endurance, leverage, and necessity for the next level of instruction. Reeves stated that “By applying some criteria to each state standard, educators can add the important dimensions of prioritization and discernment to their state standards, . . . focus[ing]. . . attention on the standards that are most important for academic success” (p. 110). Ainsworth (2004) added that the prioritization included standards necessary for the state

test and that the power standards in any content area should be limited to the five to seven core concepts that were crucial to enduring understanding and application of information.

A second criterion for successful formative assessment progress monitoring was full understanding by teachers and buy-in from teachers (Lezotte, 1992). Black and William (1998a, 1998b) supported this contention in their finding that lack of teacher knowledge, or buy-in, prevented the effective implementation of formative assessments in the classroom. Similarly, Ainsworth and Viegut (2006) stated that the assessments should be common between grade levels and content, and that teacher teams of experts should be the assessment designers. Researchers at the University of Minnesota determined that assessments had to provide data that were easy to acquire and understand, specific and timely in order to effectively impact teacher decisions and instruction (Wayman et al., 2007).

In successful implementations, expectations and classroom culture changed to a model where the teacher and student worked together. Additionally, the teacher's role changed from one of information giver to learning coach. Such global changes were risky and threatening. It was essential for the teacher to be empowered in the implementation (Black, Harrison, Lee, Marshall, & William, 2004).

Researchers have found that a key to effective use of formative assessments was interpreting the data and applying the appropriate judgment. Summative judgments and purposes were distinctly different from formative purposes. There could be no formative judgments if the assessment was designed to identify areas of proficiency and non-proficiency. To make judgments based on the data was simply slipping back to the

traditional model of assessment when the test equated to a grade, or a judgment. Only summative assessments yielded judgments, as they were designed to paint the total picture of student proficiency. Confusion between the two led to confusion in instruction and abandonment of the initiative (Newton, 2007). Because the success of implementation relied so heavily on teacher action, it became imperative that teachers not only understood and agreed, but had an on-going professional learning model to assist in building skill (Fullan, 2006).

The effectiveness and benefits of formative assessment progress monitoring in the classroom have been outlined extensively by Black and Wiliam (1998a) in their review of research on assessment. A key component of Black and Wiliams' research was the appropriate use and the correct type of feedback suggested by the assessment data. The feedback was immediate, based on current conditions, diagnostic, and specific. Feedback was based on teacher assumptions about learning as well as the tasks that would lead to the desired outcome. For the success of the progress monitoring, it was essential that the teacher make the correct assumptions and choices of tasks and skills in instruction. Similarly, assumptions were made regarding learner motivation and character. It was equally essential that subjective assumptions were avoided in favor of focusing on the objective needs as documented by the data (Black & Wiliam, 1998a).

Black & Wiliam (1998a, 1998b) found that there were two crucial actions necessary to successful formative assessments. First, it was important that the learner perceived a gap between current knowledge and desired outcome. The teacher was required to interpret the data to identify the gap and then provide the appropriate

feedback to the student for complete understanding. The other action critical to successful formative assessments was the learner as active participant in closing the learning gap, with teacher coaching and instruction. Black and Wiliam determined that the first action was easily achieved. However, the second would require positive actions in relation to study skills, collaboration, and honest self-assessment. Learners needed to understand not only the specific learning objective but the nature and complexity of the task to be learned. Students needed specific feedback about the types and description of errors. Similarly, teachers needed to modify instructional time to offer some students more time while others were moving ahead. Essential in Black and Wiliams' findings was the need for students to have time in small, collaborative peer groups to discuss the learning and assist one another in finding solutions for the learning that was still difficult. Without meaningful interaction with and among learners, no successful use of formative assessments was possible (Black & Wiliam, 1998a, 1998b).

The validity of assessment and the conclusions drawn from it have often been questioned. To determine whether the formative assessment held any validity, Nichols et al. (2009) found that both empirical data and reasoned judgment had to be paired in the interpretation and the feedback to students. The formative assessment was valid to the extent that (a) it was accurately portrayed the current state of proficiency against the target of achievement and (b) this was accurately conveyed to and received by the learner. The formative assessment labored under a potentially false set of endowments as educators and students assigned an almost magical power to the assessment's ability to pinpoint skill deficiency and prescribe intervention. Without consideration of design

methods, theoretical perspectives, and empirical evidence, this was not possible (Nichols et al., 2009). For formative assessments to be effective in measuring the identified standard or criteria, the assessments must have alternate forms, varying in both questions and complexity so that growth can be measured, rate of learning can be assessed, and the danger of test memorization can be avoided (Fuchs et al., 2008). Shepard (2009) found that the use of the formative assessment, more than the assessment itself, was the determinant in validity. The assessment needed to yield data that could impact instructional decisions, and the instructional decisions made based on the data also needed to be effective in improving student achievement (Shepard, 2009). Nichols et al. held that in order for a formative assessment to be valid, it had to lead to instructional modifications and learning that positively impacted student proficiency (Nichols et al., 2009).

Reading Proficiency Related to Overall Academic Achievement

Program improvement has been most successful when the formative assessments and the program improvements focus on the area of reading. According to Schmoker (1999, 2006), improved reading filtered into other content and caused improved achievement in mathematics, science, and social studies. Similarly, accountability for performance has been heavily weighted in the area of reading. Any academic performance improvement system must begin with the area of literacy and reading (Marzano, 2004, 2006; Marzano et al. 2005).

To say that reading and literacy are pre-requisites of academic success seems an obvious statement. This realization preceded the present era of accountability and the focus on literacy in high stakes testing. As early as 1985, the Committee on Reading was formed to study literacy and produced *A Nation of Readers* (Anderson, Hiebert, Scott, & Wilkinson, 1985) which was intended to speak in lay terms to both educators and parents about the importance of reading. That report indicated that the formation of early literacy skills was crucial to future success in academics and productive citizenry. *A Nation of Readers* compared the performance of American students to those in Japan and Taiwan and showed that a disproportionate number of American children were in the lowest 33% in reading performance. The report issued a call to parents to focus on reading in the home to support children in the academic programs delivered in school. The charge for education was to recognize the power of literacy and make it a primary focus of the instructional program (Anderson et al., 1985).

Literacy struggles have been centered in the secondary school. Investigation of the 2009 FCAT reading for 10th grade state-wide showed a meager 10-point gain in developmental scale score (DSS) as compared to hundreds of points gained in elementary schools at every level. It has been the 10th-grade FCAT that has determined whether students will graduate from high school. At the time of the present study, many of the state's high school students were not meeting the necessary standards (Florida Department of Education, 2009). Many secondary teachers have preferred to labor under the perception that reading was a skill developed and mastered in elementary school. They isolated this skill from any relevant skill needed to master their subject matter,

leaving adolescent learners struggling for meaning and understanding. Teachers at the secondary level have made the false assumption that because students can decode words, they can understand (Coutant & Perchemlides, 2005). The reaction from secondary educators, when it has been determined that students cannot read, has been to provide the information through lecture and notes, making comprehension even more elusive (Darwin & Fleischman, 2005).

Researchers have found that reading ability is the single greatest predictor of achievement in all other core areas. Marzano (2004), in his research, indicated that when students' reading skills and proficiency improved, they performed better on standardized tests measuring other contents. Similarly, Reeves (2004) and Ainsworth (2006) found that when instruction targeted reading deficiencies in students, the growth in other content areas was significant, with effect sizes approaching one standard deviation. Moxley and Taylor (2006) held that by embedding literacy strategies in content area classes, there was a consistency that led to student assimilation of the strategies, relationship between and among contents, and a uniform tool for raising student achievement across the school. Marzano (2004) found that when students were taught vocabulary to increase their reading skill and proficiency, the increase in achievement carried across content areas, building a case for the instruction of vocabulary, a prime reading component, in all academic areas. Similarly, Thompson (2008) found that in exemplary schools, vocabulary was the foundation of instruction in all content classes.

If proficiency in reading can lead to academic success, it follows that the lack of reading proficiency will result in negative effects for students (Ainsworth & Viegut,

2006). Taylor and Collins (2003) found a strong relationship between discipline and reading. They found, in reviewing discipline records for middle school students that students who presented chronic discipline issues were reading at or below the 25th percentile.

Author and teacher, Beers (2003), found that not only did lack of skill inhibit performance, but low performance resulted in negative attitudes in students regarding reading. This compounded the spiral of low academic achievement. Without a positive perspective on reading, according to Beers, students were not able to effectively interact with the text to draw meaning. Thus, content skill achievement and assessment performance were hindered (Beers).

Limited research was found concerning the relationship of reading proficiency to achievement in core content areas in the secondary school. Visone (2009) conducted a study to determine if a relationship existed between reading and achievement on the Connecticut standardized science assessment, the Connecticut Academic Performance Test (CAPT). In the study, reading scores of a demographically diverse random sample of 767 tenth-grade students were compared with the science score on CAPT. The reading proficiency level used was the raw score students earned on the Reading For Information (RfI) test on the Spring administration of the CAPT. The science scores reported were the raw score, scale score, and a corresponding performance level from 1 to 4, with 4 being advanced. The findings of the study documented a moderate to strong correlation ($r=.54-.79, p \leq .05$) between the RfI proficiency raw score and the science CAPT raw score across all subgroups reported (gender, ethnicity, free and reduced lunch, and

English language learners). The data reported confirmed the hypothesis that there was a relationship between reading and science achievement. While this study did not employ statistical tests to establish predictability, given the moderate to strong correlation, it is possible to surmise that a prediction could have been determined. The researcher's recommendation for replication of the study in other states with varying contents would help in determining causality and predictability (Visone, 2009).

Goddard (2003) conducted a study concerning relationships and norms of fourth grade students. While not designed to study content relationships and reading correlation to content areas specifically, research conducted to test the study hypotheses included the use of reading and mathematics assessment scores for fourth graders. Goddard found that reading and mathematics achievement correlated positively on a moderate level. Additionally, he found that reading achievement in prior years was also positively correlated on a moderate level to later mathematics achievement. Findings in the study implied that reading proficiency was related to achievement in other areas for fourth grade students (Goddard, 2003).

Kelly and Gaustad (2006) conducted a study measuring the relationship between deaf college students' reading and language proficiency and their mathematical skills. Reading proficiency was assessed using the Michigan Test of English Language Proficiency (MTELP), The California Achievement Test for Comprehension (CATC), and a morphological knowledge test developed by Gaustad. Mathematics skill was assessed using the American College Testing (ACT) Mathematics Subtest. The study results indicated that mathematics proficiency and performance was significantly related

to morphological knowledge and reading proficiency. The study found that scores of students who took the ACT Mathematics subtest ($n=37$) correlated from moderate to strong with the reading assessments given (MTELP: $r = .44$, $p \leq .05$; CATC: $r = .67$, $p \leq .01$; Gaustad split decision: $r = .68$, $p \leq .01$; Gaustad meaningful parts: $r = .50$, $p \leq .01$). Findings of this study seemed to suggest that the reading and morphological knowledge and proficiency of deaf students was directly related to mathematics performance and course grades. If the conclusions can be generalized, then these results could suggest that reading knowledge and mathematics performance was also correlated for non-deaf students.

Lutz-Doemling (2007) conducted a study measuring the relationship between the Pennsylvania 4Sight Reading Benchmark Assessment and the Pennsylvania System of State Assessment (PSSA) mathematics score in sixth graders who participated in the 4Sight Reading Benchmark Assessment in fifth grade. The 4Sight Reading Benchmark Assessment was developed commercially for Pennsylvania with a Cronbach Alpha of .91 for fifth grade reading. The study utilized an experimental group ($n=438$) and a control group ($n=413$) that did not participate in the 4Sight Reading Benchmark Assessment. A convenience sample, which included diverse populations matching the percentages in the county, was used for both groups. The purpose of the study was to determine whether a statistically significant relationship existed between the 4Sight Reading Benchmark Assessment in fifth grade and the PSSA score in mathematics in sixth grade. Lutz-Doemling found that the correlation between the fifth grade 4Sight Reading Benchmark Assessment and the PSSA in mathematics in sixth grade was both positive and

statistically significant in the experimental group ($n=438$, $r = .77$, $p \leq .0005$). Based on the data, the researcher concluded that there was a strong relationship between the 4Sight Reading Benchmark Assessment and the PSSA mathematics score. While the data revealed a strong relationship, the researcher indicated that effective use of the data required a full understanding of data analysis and would require action on the understanding to impact school improvement. The researcher recommended that further study be conducted in other grades (Lutz-Doemling, 2007).

Leadership and Effective Strategies for Progress Monitoring Programs

Formative progress monitoring program data used as a tool to not only monitor student learning progress, but also to analyze teachers' instructional practices, has been shown to positively impact student achievement. Schmoker (2006) found that adjustments in instructional practices were productive when informed by effective, frequent formative assessments that were directly aligned to the prescribed curriculum. Irvin et al. (2007) expressed their belief that there must be a culture of collaborative investigation of data for the purpose of continuous improvement. Taylor and Collins (2003) suggested that it was important to align the curriculum, classroom assessments, and standardized tests. Without such alignment, the data gathered was not effective in suggesting instructional modifications.

Ainsworth and Viegut (2006) found that teachers who routinely utilized the data from formative assessments were able to effectively structure and modify instruction to meet the needs of students. Information that students already knew was clearly

identifiable and this identification eliminated reteaching content. Teachers also immediately perceived trends in learning and learning gaps for the purpose of remediation and were able to make appropriate modifications to the instructional plan. Individual student differences were also highlighted, enabling differentiation of instruction to meet each student's need. Finally, regular use of common formative assessments served to provide feedback to students concerning academic progress (Ainsworth & Viegut, 2006).

Analysis and understanding of student data facilitated the modification of instruction to remediate the gaps in performance of individuals or subgroups (Irvin et al., 2006). Fuchs and Fuchs (1999) outlined several criteria that defined the data collection and analysis of effective progress monitoring systems. First, programs and assessments to monitor and improve student data were directly aligned to standards and required a measure of mastery which would assist in defining students' needs for teachers. Additionally, the assessment reporting clearly identified the relationship between the instructed content and the performance proficiency. Finally, successful systems allowed measurement and data collection for potentially large numbers of students. With general regard to progress monitoring programs, researchers contended that accurate and systematic data collection allowing for routine measurement of student proficiency over time and effective analysis of that data enabled teachers to prescribe appropriate instructional interventions and remediation. Thus, the existence of the assessment, however effective, was not sufficient to impact achievement. Analysis and understanding

of the data was necessary for impacting student growth and building instructional best practices (Fuchs & Fuchs, 1999).

Marzano et al. (2005) stated that the single greatest indicator of an effective, high performing school was the quality of the school level administrator. Thus, the principal's mission is to use the assessments designed to monitor and improve achievement positively and proactively in regular practice in the school, providing opportunities and resources for teachers to implement and utilize the formative assessment data effectively (Marzano et al., 2005). In many schools, teacher leadership required to initiate and sustain collaborative efforts in the effective use of data, study groups to investigate research, planning for future lessons, and modeling of effective strategies was weak or non-existent. Marzano et al. stated that the principal as school leader was charged with the role of working with teachers to nurture and provide the necessary support. This support included providing resources, finances, time, and materials to maintain a commitment to the process of utilizing the data from the formative assessments in a manner that modified instruction to positively impact student achievement. A primary component to building a collaborative culture where meaningful discussions concerning data and student achievement occurred, was the role of principal as cheerleader for professional development to teach and enhance instructional skills among teachers (Stigler & Hiebert, 1999).

According to Firestone (2009), the role of the principal as leader in the effective use of formative assessments to improve student performance on state standardized tests was critical. Firestone identified the leadership role as being a liaison between an

accountability culture and a student learning culture. In an accountability culture, testing and central control would be evident. In a student learning culture, students would be considered a priority, and teachers would be assumed to possess the professional skill and desire to help all students. Strong leadership from the top and mutual respect among the adults in the school building would be required. Firestone defined the student learning culture as one where there was shared responsibility for student growth and personal professional development. Senge (2006) referred to such a culture as the fifth discipline within the school and one that was crucial to the success of the institution and the individuals functioning within it. Senge characterized the successful organization as one “that discover[s] how to tap people’s commitment and capacity to learn at all levels. . . .” (p. 4). Such organizations thrive on individuals who possess personal mastery, have great dedication to learning and a commitment to growth of every member of the organization. This growth was viewed as the norm for personal satisfaction and fulfillment, not simply for accountability. With such a learning culture modeled by the leader, Senge held that organizations could realize maximum fulfillment of the highest goals and achievements.

Strategies for successful leaders have been outlined by many current researchers. Schlechty (1997) described leadership in an environment of change as managing results and motivating others to manage the results in order to change. Cotton (2003) found that the principal’s vision and actions in conveying high expectations promoted a positive learning climate and a school culture that valued student achievement. Similarly, visible leadership of an approachable principal contributed significantly to school achievement. An underlying tenet of such leadership was the empowerment of teachers to make viable

decisions about learning, school focus, and shared leadership responsibilities in the school (Cotton, 2003).

Fullan (as cited in Crow, 2009) described the essential role of principal leadership within the school as motion leadership. Effective principals identified and focused on the key components of the change desired and created positive movement. Leadership responsibility was to work on the fundamental goals and then advertise and advocate these to the staff for positive results. Inherent in Fullan's motion leadership, were the key components of effective leadership necessary to bring any change (Crow, 2009).

Fullan (2001) described five practices successful leaders share: leading with moral vision, building relationships, managing change, sharing knowledge, and building the bridge between new knowledge or practices and old paradigms. Marzano et al. (2005) cited 21 behaviors employed by successful principals which echo those identified by Fullan (2001). Among these, vision, culture, communication, and involvement were identified as keys to leading a successful school focused on student achievement. Monitoring and feedback were essential to improving student performance, and Marzano et al. (2005) advocated an approach where research and sharing data were as important as a hands-on approach to leading. The effective leadership practices cited in research most pertinent to the present study were building a culture of student achievement and a genuine belief that all students can learn, collaboration and sharing, and effective professional development. These areas can be specifically defined to clarify the scope of the effective leadership practice.

The creation of school culture has been the topic of volumes of research and books. Among the recognized culture experts, Peterson and Deal (2002) indicated that the school improved through a shared system of values, beliefs, and norms, or culture. This shared system gave the school its meaning and vision, a sense of purpose, and a definition of expectations and actions. These authors defined it as the heart and passion of the organization. Culture, whether positive or negative was immediately recognizable in schools studied by Peterson and Deal (2002). The culture developed by the school leader defined behavior, infused commitment, and built motivation. Culture facilitated school effectiveness and was shaped by the leader through actions and interactions with staff, students, and community (Peterson and Deal, 2002). Schein (1986) considered culture a deeply ingrained set of beliefs whose tangible evidence was behavior. This behavior, then, became the norm in the organization and was tested against daily tasks, issues, and events to measure the success of the behavior. Successful behavior was adopted as the organizational way of doing things and evolved into the ingrained culture. Essential to culture, according to Schein, was a common language and a common set of expectations and rules for interacting with the organization, tasks, and each other. The assumptions that the individuals in the organization learned became the values and the culture (Schein, 1986).

Gordon (as cited in St. John, 2009), in a survey of 143 principals, documented that great principals established positive learning and nurturing cultures. The studied principals accomplished this through actions that demonstrated their caring for others, finding out what was needed, providing the resources, and recognizing other's

achievements. Within the culture, principals established and communicated strong vision and expectation, but also supported these through facilitative teams and support systems. Principals who were able to establish strong cultures did so by demonstrating caring and support for the individual and individual decisions while demonstrating academic and performance expectations (St. John, 2009).

Gordon (2006) found that the application of a modified Maslow's hierarchy could assist principals in supporting the organization so that optimum performance and culture could be attained. The hierarchy was a four-level system where basic needs were level one at the base and growth was at level four, the apex. In level one, principals needed to insure that the most basic provisions were available to teachers. Items such as materials, equipment, and time were essential as well as knowing specifically what was expected of the teacher in basic, task oriented language (Gordon, 2006). Whitaker (2003) supported this basic level by stating that great principals recognized the difficulty of changing people's values. When the principals developed culture, they focused first on clear, definable behavior. Whittaker held that repeated behaviors eventually translated to internal beliefs and values. The second level of Gordon's (2006) hierarchy was management support where there was encouragement of development, caring supervisors, recognition, and a confirmed feeling on the part of employees that they were doing something they did well. Level three was teamwork. In teamwork, there was a clear sense of the mission and vision and the role that the employee played in making the accomplishment of those a reality. There was also a sense that employees had friendship with peers and that their opinions mattered to the management and functioning of the

organization. Finally, the pinnacle of the hierarchy was growth. In growth, there were clear opportunities provided for learning and development and a measureable progress in a six-month period (Gordon, 2006).

Barth (2006) defined excellence in collaboration in schools through characterization of the relationships among the adults in the school. Barth held that the relationships apparent between adults defined all the relationships in the school. Without positive collaborative relationships, limited progress could be made in school improvement, improvement of teacher efficacy, and student achievement. Inherent in Barth's relationships were what he termed "nondiscussables" (p. 8) or issues that were difficult to openly discuss. Usually, these were discussed in gossip sessions, but Barth held that these issues had to be brought to the forefront and handled before any productive relationships could be established and maintained. Barth defined four basic relationships in the school. Parallel play existed when adults did not interact in professionally related ways. They worked side by side yet did not collaborate or combine efforts to learn, grow, or solve problems. They operated in silos and alone. Adversarial relationships were non-productive and negative interactions that were often blatantly antagonistic or subversive. There was no sharing, and the feeling that individuals had to experience the hard knocks for themselves was prevalent. Characteristic of adversarial relationships was the fact that in such an environment, the lack of sharing caused history and best practices to vanish when the individual vanished. Congenial relationships existed when adults were cordial and pleasant on a social, caring level. Such relationships were professionally non-productive as they did not help individuals grow,

learn, or solve problems in the profession. Finally, collegial relationships provided the culture for collaboration. In such relationships, individuals shared professional knowledge and had substantive discussions concerning their skill and craft. When such relationships existed, it was possible to have a meaningful collaborative culture and professional learning community (Barth, 2006).

Fullan (as cited in Crow, 2009) described the principal's role in building collaboration and opportunities for sharing as one where the leader mobilized people to learn from each other and do the important work in the school. This group was defined as a leadership team, literacy council, school improvement team, or any group that came together to learn. Important to the function of the group, however, was that they were empowered to make real and important decisions, not merely insignificant decisions or recommendation for a decision to be made by the leader in a top down approach. Such groups made leadership decisions that were able to be shared. Fullan further described an extension of this mobilization where the group was an extension of groups in the school, and the school was an extension of groups of schools in the district (Crow, 2009). Successful leaders built teams that were mutually supportive. While it was important that teachers retained autonomy and independence in some areas of their work, it was recognized that interdependence was critical to the emotional support and development that was essential to the craft of teaching (St. John, 2009).

Without effective professional development, building a culture of achievement and collaboration was not possible. Easton (2008) found that in many school districts, a training department existed to provide what some still referred to as training. This was a

misnomer, as professional development evolved into actual, active learning rather than a “sit-and-get” session from an expert. Educators have found that to address the specific needs of schools and students, they must change. Easton stated that this change required real learning. Both time and format needed to be altered to provide effective professional development. Time for professional development had to be real, current, and in-contract hours time. The location of professional learning had to be school based, and the format was most appropriately embedded in teachers’ classrooms with their students. Support for teachers embarking on this learning and change was essential. Support occurred in the form of coaching, mentoring, observation, and collaboration with peers. Similarly, evaluation of professional development had to change from a feeling assessment to an implementation assessment and student work improvement measure (Easton, 2008).

Schmoker (1999) indicated that much of professional development looked good on paper but provided no substance or real learning. Such professional development was useless, as it provided no structure for measurement of results or improvement. Professional development structures were successful when developed by the district and when district goals along with research based best practices were used to select the professional development to be offered. Success had to be measured in actual student performance and growth. Teacher action research that focused on research and student work was found to be a highly effective professional development method and one where collaboration was essential (Schmoker, 1999). Easton (2008) stated that this collaboration facilitated effective professional learning efforts. These included regular teacher visits to other classrooms and results-focused learning groups or professional

learning communities where, when the need for change was detected based on student work, the change was initiated without waiting for direction from a top-down leadership system. Cotton (2003) cited that principals of high achieving schools structured on-going professional development opportunities and utilized resources creatively to find both time and materials to offer the best professional learning. Similarly, they expected attendance and implementation of the learning to be supported through collaborative groups.

Thompson (2008) stated that professional development was essential to both the implementation of exemplary practices and the sustaining and monitoring of those practices. Effective leaders chose two or three professional development initiatives based on research of exemplary practices and supported these throughout the year. Key to Thompson's vision was that the leader was part of all professional development activities alongside the teachers. Evaluation and observation was structured around the expectation that teachers implement the learning in the classroom. In each classroom walkthrough, principals had identified what they would be looking for based on the initiatives and what specific questions they would ask teachers directly related to the implementation's effect on student learning and achievement. The initiatives from the professional development were consistent in the school and pervasive based on the expectations of the principal, and teachers were held accountable for implementing the learning and sharing with peers around the focus of actual student work (Thompson, 2008).

In a study conducted by Quick, Holtzman, and Chaney (2009), interviews with teachers indicated that effective professional development was not only directly related to school goals but was also provided in a safe and trusting environment. Collaboration and

opportunities for modeling and practice were important to the implementation and use of the professional learning. Similarly, effective professional learning was not an isolated incident, but was sustained and connected to both previous and subsequent learning. Perhaps most important was that professional learning was based on the actual identified needs of the teachers among which was the actual implementation with real students rather than theoretical discussion using imaginary students (Quick et al., 2009).

The National Staff Development Council (NSDC) (2001) developed standards to govern effective professional development. These standards were divided into context, process, and content and clearly outlined what effective professional learning encompassed. Collaborative learning communities utilizing student data to determine priorities were critical to the effective learning model set by NSDC (2001). Also important was the school leader's role in guiding and directing continuous learning for continuous improvement and the resources needed to implement the learning. The State of Florida Department of Education (FLDOE) (2006) developed the Florida Professional Development Evaluation Protocol to direct and guide districts in planning, selecting, delivering and evaluating professional learning. The Protocol defined three levels of professional development: faculty, school, and district. The standards at each level addressed four components defined as (a) planning, (b) delivery, (c) follow-up, and (d) evaluation. In each component, requirements were well defined and clearly supported the research surrounding effective professional development. The planning domain required that professional learning be not only based on research but on student data, student needs, and teacher needs. In the delivery component, indicators were focused on

collaborative groups, professional learning communities, and action research teams. Follow-up focused on a coaching model or other systems where modeling, practice, and observation were prevalent to support the teacher in implementing new learning. Finally, evaluation focused on student growth as a result of the professional learning and documentation that the growth could be directly tied to the learning and effective classroom implementation (Florida Department of Education, 2006).

Summary

The American public school systems have embarked on an era of accountability initiated with the Elementary and Secondary Education Act and continued by The No Child Left Behind Act. A result of this accountability movement was the movement toward objective measures to monitor student achievement and impact student growth for performance on high stakes state tests.

The review of literature indicated that the use of formative assessments was an educationally sound practice for monitoring and measuring student growth. Formative assessments, when used effectively, focused on student improvement. The process for utilizing formative assessments in a progress monitoring program relied heavily on appropriate interpretation of data and feedback. Essential in the success of formative assessment was not only identifying the gap between current skill level and desired skill level, but also student recognition and motivation for learning and growth.

Progress monitoring formative assessments in reading were shown to be effective in monitoring improvement in other content areas. Limited studies showed a relationship

between reading skill and achievement in science and mathematics. In order to realize the most positive results in student achievement from a progress monitoring program using formative assessments, the role of the principal as instructional leader has been critical and was well defined in research. Leaders who built positive learning cultures, structured opportunities and expectations for collaborative teams, and planned and delivered effective professional development were successful in providing the teacher learning necessary to implement change in the classrooms for student growth.

CHAPTER 3 METHODOLOGY AND PROCEDURES

Introduction

This chapter presents the methodology and procedures used to conduct the study. Addressed are the five primary research questions that were used to guide the study, the demographics of the targeted county and the specific study population. Also described are the three instruments which were utilized and the data collection process. Finally, the methods employed in the analysis of data are presented as they relate to each research question.

Research Questions

The following research questions were used to guide the research:

1. To what extent does the 2008 District Benchmark Reading Assessment (DBMRA) predict achievement level on the Florida Comprehensive Assessment Test (FCAT) in Reading for 10th grade for school year 2008-2009?
2. To what extent does the 2008 District Benchmark Reading Assessment (DBMRA) predict achievement level on the Florida Comprehensive Assessment Test (FCAT) in Mathematics for 10th grade for school year 2008-2009?
3. To what extent, if any, does the relationship between scores on DBMRA and FCAT Reading differ among high schools in the study?

4. To what extent, if any, does the relationship between scores on the DBMRA and FCAT Reading differ by AYP subgroups?
5. What strategies do principals of high schools use where the correlation between the November, 2008 District Benchmark Reading Assessment (DBMRA) and the 2009 FCAT score in reading fall into or above Cohen's medium effect ($d \geq .50$)?

School District Demographics

The targeted county is located in north central Florida. The district, the fifth largest public school district in Florida, has 51 urban and rural schools and a student population of slightly over 40,000. A total of 56% of all students are on free/reduced lunch. Ethnicity of the student population is as follows: White, 59%; Black 19%, Hispanic, 15%, and Other 7%. The annual budget of the district is approximately \$675 million with a per pupil expenditure of \$8,296. The land area encompasses 1657 square miles. The District FCAT results and State assigned school grades have reflected a steady increase in A, B, and C school grades for elementary and middle schools since 2004, the year of implementation of the Reading Progress Monitoring Program. Furthermore, the District percentage of students scoring at level 3 proficiency in FCAT reading has increased for elementary and middle school students since the program's inception, starting with 51% in 2004/2005 and increasing to 56% in 2006/2007.

Study Population

The specific population for this study were the 2,263 tenth-grade students enrolled in the seven public high schools of the targeted district during the 2008-2009 school year. Excluded from the study were two high schools which were considered alternative schools due to the limited grade levels and alternatively assessed population of students with disabilities and/or discipline assignments. Students deemed eligible and included in the study population were only those that participated in both the November, 2008 DBMRA for reading and the FCAT Reading and Mathematics for the 2008-2009 school year. The study population was distributed almost evenly among schools, with 100 students separating the smallest (School B = 283) and largest (School C = 383) schools. Table 2 lists each school and the number of 10th-grade students included in the study.

Table 2
Study Population: Participating High Schools and 10th-Grade Students

High Schools	Students
School A	290
School B	283
School C	383
School D	326
School E	291
School F	344
School G	346
Total	2263

Demographics and AYP sub groups for each school in this study are shown in Table 3. The school population was comprised of a fairly even distribution of males and females with a range of 48% to 52% in each gender group. The only exception occurred

in School F where there was a 5% difference in males (45%) and females (55%). All schools had a majority of White students. The Black population was approximately 20% except for a 9% Black population in school A. The Hispanic population was less evenly distributed and ranged from a low of 8% in Schools E and F to 23% in School B.

The number of students indicated as eligible for free and reduced lunch, a measurement of socioeconomic status, was determined by self-filing paperwork to qualify. In most cases, schools had additional students that may have been eligible but chose not to complete and return the necessary forms. Therefore, data on the free and reduced population were not considered as accurate or complete. The reported free and reduced lunch percentages ranged from 36% at School C, the largest school, to 65% at School D. No high school received Title 1 funds. Percentages indicated that six of the seven high schools included in the study would place above the district cut score of 40% for Title 1 funds if the district included secondary schools in the Title allocation.

The population of students with disabilities was between 10% and 16% at all schools. The percentage of English Language Learners (ELL) was between 2% and 4% at six of the seven schools. School B, the smallest school, had almost double the ELL population with 6% of its students being categorized as ELL.

Table 3
Demographics: Percentages of Students Enrolled in Participating Schools

School	Male	Female	Black	White	Hispanic	Other	Free/ Reduced Lunch	SWD	ELL
School A	50%	50%	9%	73%	14%	4%	41%	12%	3%
School B	42%	48%	19%	53%	20%	6%	60%	18%	5%
School C	53%	47%	18%	65%	11%	5%	40%	12%	2%
School D	50%	50%	18%	66%	11%	5%	66%	13%	2%
School E	47%	43%	28%	62%	8%	3%	61%	13%	2%
School F	46%	54%	24%	60%	7%	8%	49%	10%	2%
School G	50%	50%	20%	50%	22%	6%	54%	11%	6%

Note. SWD = Students with Disabilities; ELL = English Language Learners.

Schools differed in demographic group percentages of students included in the study. However, the study population reflected percentages equal to the school populations in most categories. Noted exceptions were the reduction of the Hispanic percentage in the study population at School B ($n = 395$) of three percentage points and increase of the Hispanic percentage in the study population at School G ($n = 437$) of two percentage points. Table 4 shows the percentages of students in the study for each school in each demographic group.

Table 4
Demographics: Percentages of Students Enrolled in Study Population

School	Male	Female	Black	White	Hispanic	Other	Free/ Reduced Lunch	SWD	ELL
School A	49%	51%	9%	73%	14%	4%	43%	13%	4%
School B	52%	48%	19%	53%	20%	8%	59%	18%	5%
School C	53%	47%	18%	65%	11%	6%	37%	12%	2%
School D	50%	50%	18%	66%	11%	5%	66%	13%	2%
School E	47%	53%	28%	62%	8%	2%	61%	13%	2%
School F	46%	54%	24%	60%	7%	8%	49%	10%	2%
School G	50%	50%	20%	50%	22%	8%	54%	11%	5%

Note. SWD = Students with Disabilities; ELL = English Language Learners.

Schools also differed in school grade assigned by Florida Department of Education based on previous FCAT data. Table 5 shows the school grades for the 2008-2009 school year and the percentage of students in the school meeting high standards, or scoring at Levels 3, 4, or 5, in Reading on the 2009 FCAT.

Table 5
School Grades and Students Meeting High Standards in Reading (2008-2009)

High Schools	School Grade 2007-2008	School Grade 2008-2009	Students Meeting High Standards in Reading
School A	A	C	47%
School B	C	C	41%
School C	A	C	52%
School D	B	C	39%
School E	B	D	35%
School F	A	C	56%
School G	C	B	46%

Five of the seven high schools in the study (71%) decreased by at least one letter grade from 2008 to 2009. Only one school, School G, increased one letter grade from 2008 to 2009. Four schools (57%) decreased by two letter grades, and two schools (28%) decreased by one letter grade. The grade for School B remained the same from 2008-2009. Of the seven high schools, only Schools B and G showed 50% or more of the lowest quartile making learning gains in reading for the 2008-09 school year.

Instrumentation

District Benchmark Reading Assessment (DBMRA)

The District Benchmark Reading Assessment (DBMRA) was designed in 2005 by school district personnel. The assessment contained approximately 50 questions in a multiple choice format that were structured to measure the nine reading benchmarks which were tested on the FCAT. Items were written by district curriculum experts and teachers following guidelines for multiple choice question development prescribed by the district. DBMRA item content was aligned with FCAT item specifications for each grade level. The goal of the DBMRA was to assess student proficiency in tested benchmark skills, identify learning gaps for remediate, and define instructional interventions to address student weaknesses. The goal was to utilize the DBMRA and the results to measure, remediate, and retest with the intent of closing learning gaps to produce increased performance on FCAT (D. L. Greene, personal communication, 2008).

The 10th-grade reading assessment consisted of 50 multiple choice questions, with 3-11 items in each of the nine reading benchmarks tested on FCAT. The DBMRA was written to mirror the FCAT format, without extended response items, using only multiple choice questions. A test summary report is included in Appendix B. The completion time for the DBMRA is approximately 50 minutes in a typical classroom period, administered by the instructor within district secure testing guidelines (Marion County Public Schools, 2007). After assessment completion, school administrators scan answer documents and scores are transmitted electronically to the district and entered into the Student Information System. The DBMRA is curriculum/benchmark criterion referenced (D. L. Greene, personal communication, 2008).

The scores reported for the DBMRA are raw scores on a scale of 0-100 with the score indicating the percentage of correct answers. Scores are grouped into levels ranging from 1-5 where level 5 = superior (90-100%), level 4 = proficient (80-89.9%), level 3 = marginal (60-79.9%), level 2 = non-proficient (40-59.9%), and level 1 = non-mastery = 0-39.9%. Level 3 is considered proficiency for mastery of the tested benchmarks on the DBMRA. Reliability has been established through the calculation of Cronbach's Alpha (.88) for the 10th grade for the 2008 DBMRA (Pearson Education, 2008). Results from the November 2008 test administration were used for this study.

Florida Comprehensive Assessment Test (FCAT)

The Florida Comprehensive Assessment Test (FCAT) has been administered yearly in mid to late March to students in grades 3-10 to assess reading and mathematics.

The results of this standardized achievement test define the levels of student proficiency targeted for mastery in both content areas up to and including the current grade. Student scores reflect three distinctly different scores: (a) a scale score which is a raw score of percentage of items correct on a scale of 100-500 for one particular test in one particular year, (b) a developmental scale score from 0-3000 that provides the ability to track student growth and progress over time and allows comparison from year to year to identify growth, and (c) a proficiency level from 1-5 which corresponds to the scale and developmental score. Level 3 and above has been considered to be proficiency. (Florida Department of Education, 2007). For the purpose of this study, 2009 FCAT scale scores in reading and mathematics were utilized.

Reliability for the FCAT has been measured for internal consistency using Cronbach's Alpha coefficients. The reliability coefficients have been expressed in a number from 0.0 to 1.0, where zero is a lack of reliability and one is high reliability. The 2003 Cronbach's Alpha reliability for FCAT in 10th-grade reading and mathematics were .85 and .88 respectively.

Validity for the FCAT, according to the Florida Department of Education (2007), refers to the interpretation of the test score as valid. Content validity indicates that the test measures what it is intended to measure. The Florida Department of Education has used several specific steps to insure the validity of the FCAT. Those are:

1. Educators and citizens judged the standards and skills acceptable.
2. Item specifications were written.
3. Test items were written according to the guidelines provided by the item specifications.
4. The items were pilot tested using randomly selected groups of students at appropriate grade levels.

5. All items were reviewed for cultural, ethnic, language, and gender bias and for issues of general concern to Florida citizens.
6. Instructional specialists and practicing teachers reviewed the items.
7. The items were field tested to determine their psychometric properties.
8. The tests were carefully constructed with items that met specific psychometric standards.
9. The constructed tests were equated to the base test to match both content coverage and test statistics. (Florida Department of Education, 2007, p. 40)

Because the FCAT was developed using these criteria, it has been considered to have content validity.

High School Principal Interview Questions

Interview questions were written as a qualitative instrument for personal interviews with principals (Appendix C). A pilot study was conducted by sending the four questions developed to three principals not participating in the study. These principals were asked for input on clarity and focus of the questions. Based on input, no modifications were made.

Data Collection

This study used a non-experimental research design using secondary data collected through district testing procedures in school year 2008-2009. The data collected were school assignment, DBMRA reading score and FCAT Reading and Mathematics score for any 10th-grade student who participated in and had a score for the DBMRA given in November, 2008.

Additionally, demographic data including gender, ethnicity, free and reduced lunch and disability for any 10th grader included were collected. A research approval

letter (Appendix A) was secured from the Public School's Guidance and Testing Office. An Institutional Review Board application and study overview was submitted to the University of Central Florida's Institutional Review Board, and an exempt status was secured (Appendix D).

Data elements were requested from the county's Public Schools Student Information Office as a data file sent to the researcher electronically. Data were compiled by a third party who did not have any knowledge of the research questions or focus of this research study. Data were sent without any identifying indicators for individual student identity. The researcher complied with all District and FERPA student privacy requirements.

Personal Interviews

Personal interviews were conducted with all seven of the high school principals to answer Research Question 5. Participation in the interviews was voluntary, and the interviews were conducted during May, 2010 in a structured format using open-ended questions. There were four interview questions that sought to identify (a) the importance of data, particularly the District Reading Benchmark Assessment (DBMRA), to the principal and (b) the successful best practices used with the data by principals to impact instruction. The interview questions (Appendix C) are listed below.

1. To what extent is the District Benchmark Reading Assessment implemented in 10th grade classrooms with fidelity? Can you provide examples? How do you know?

2. To what extent do you consider the District Benchmark Reading Assessment an important tool in improved student achievement on FCAT Reading at your school? Why?
3. When the data comes in, what do you do with it? Is it ever used to modify instruction and how? Can you provide examples?
4. How do you monitor what you told me takes place?

Prior to the interviews, principals were contacted by electronic mail to request an interview. The electronic mail contained an informed consent document (Appendix E) and a Summary Explanation of the research (Appendix F). In agreeing to the interviews, principals responded by electronic mail or phone. The researcher contacted each principal by phone to schedule an appointment for a personal interview.

Data Analysis

1. To what extent does the 2008 District Benchmark Reading Assessment (DBMRA) predict achievement level on the Florida Comprehensive Assessment Test (FCAT) in Reading for 10th grade for school year 2008-2009?
2. To what extent does the 2008 District Benchmark Reading Assessment (DBMRA) predict achievement level on the Florida Comprehensive Assessment Test (FCAT) in Mathematics for 10th grade for school year 2008-2009?

For Research Questions 1 and 2, regression analysis was used to determine the mathematical model for prediction. A scatter plot was prepared for each high school in the study as well as a composite of all seven high schools before computing the Pearson correlation statistic. A Pearson correlation coefficient was computed for DBMRA and

FCAT Reading to determine the relationship between the DBMRA and the FCAT Reading score. The Pearson correlation coefficient was computed by school and for the schools as a group. Cohen's (Becker, n.d.) standard for effect size was used to characterize scores computed.

The d statistic created by Cohen was a scale for operational definitions of effect size for a more descriptive and qualitative way of discussing statistical effect size in social sciences research. Cohen's d can be calculated from Pearson's Product Moment r statistic with the following formula $d = 2r/\sqrt{1-r^2}$. Cohen then assigned the resulting d value to the corresponding descriptor of effect size based on a range of d values (Cohen, 1988, 1992).

Cohen assigned descriptors to his measure of effect size as small ($d = .2$), medium ($d = .5$), and large ($d = .80$). While Cohen recognized the risk in assigning general label to effect size for use in analysis of power, the categories were considered appropriate for use as descriptors of results (Becker, n.d.; Cohen, 1988).

3. To what extent, if any, does the relationship between scores on DBMRA and FCAT Reading differ among high schools in the study?
4. To what extent, if any, does the relationship between scores on the DBMRA and FCAT Reading differ by AYP subgroups?

For Research Questions 3 and 4, Pearson's Correlation Coefficient was used to determine the relationship. Cohen's (Becker, n.d.) standard for effect size was used to determine strength and compare correlations of the seven high schools and differences between subgroups.

5. What strategies do principals of high schools use where the correlation between the November, 2008 DBMRA and the 2009 FCAT score in reading falls into or above Cohen's medium effect ($d \geq .50$)??

A qualitative interview instrument was used to record responses. Constant comparison analysis was used to analyze the answers to the open-ended questions. In this analysis, the researcher listed all answers for each question for interview one. When subsequent interviews were reviewed, the answers were compared against those already listed. Any new answer or trend was added and duplicate responses were tallied. The same process was applied to each interview for each question to determine the complete list of responses with frequency of response. These responses were chunked into phrases and then grouped into similar categories to identify themes in practices of principals. This analysis provided supplementary information to the statistical data provided in Research Questions 1, 2, 3, and 4 and described the importance of the findings in practice.

Summary

This chapter provided an overview of the targeted public school district's demographics and the specifics of the seven high schools included in the study. A description of the instruments used in the study was provided, and the validity and reliability were documented. The data collection and analysis procedures were also described. Information was provided regarding approval by the district and the university to ensure the integrity of the research.

CHAPTER 4 ANALYSIS OF DATA

Introduction

This study sought to determine the effectiveness of a locally created formative assessment in reading as a predictor of performance on the state standardized test in 10th-grade reading and mathematics. Additionally, the best practices principals used regarding data with teachers and within the instructional program were investigated. This chapter provides the results of the data analysis used to answer each of the research questions and establishes the foundation for conclusions and recommendations for further study presented in Chapter 5.

Purpose of the Study

The purpose of the study was to (a) provide information to high school principals and teachers to better understand how students were performing and learning and (b) how to maximize use of the district benchmark assessment in order to modify instruction and positively impact student achievement. An additional focus of this study was to determine best leadership practices in schools where students exceeded the predicted FCAT scores. This study expanded on a 2008 study of third grade students (Gaught) using data from the 2007-2008 school year to determine if the District Benchmark Reading Assessment (DBMRA) was a predictor of Florida Comprehensive Assessment Test (FCAT) performance. This study was focused on a larger sample of 10th-grade students. The researcher investigated the potential for predicting the FCAT Mathematics

scale scores based on the DBMRA scores. Additionally, the study was conducted to determine whether there was a difference in the strength of prediction of DBMRA and FCAT Reading scale scores between high schools and subgroups. Finally, principals were interviewed to identify best practices related to formative assessment data usage in high schools.

Research Questions

This study was guided by the following research questions:

1. To what extent does the District Benchmark Reading Assessment (DBMRA) predict achievement level on the Florida Comprehensive Assessment Test (FCAT) in Reading for 10th grade for 2009?
2. To what extent does the District Benchmark Reading Assessment (DBMRA) predict achievement level on the Florida Comprehensive Assessment Test (FCAT) in Mathematics for 10th grade for 2009?
3. To what extent, if any, does the relationship between scores on the District Benchmark Reading Assessment (DBMRA) and FCAT Reading differ among high schools in the study?
4. To what extent, if any, does the relationship between scores on the District Benchmark Reading Assessment (DBMRA) and FCAT Reading differ by AYP subgroups?
5. What strategies do principals of high schools use where the correlation between the November, 2008 District Benchmark Reading Assessment

(DBMRA) and the 2009 FCAT score in reading meets or exceeds Cohen's medium effect ($d \geq .50$)?

Description of Sample Population

The specific population for this study were the 2,263 tenth-grade students enrolled in the seven public high schools of the targeted district during the 2008-2009 school year. Excluded from the study were two high schools which were considered alternative schools due to the limited grade levels and alternatively assessed population of students with disabilities and/or discipline assignments. Students deemed eligible and included in the study population were only those that participated in both the November, 2008 DBMRA for reading and the FCAT Reading and Mathematics for the 2008-2009 school year. The study population was distributed almost evenly among schools. Only 100 students separated the smallest (School B = 283) and largest (School C = 383) schools. Table 6 lists each school and the number of 10th-grade students included in the study.

Table 6
Study Population: Participating High Schools and 10th-Grade Students

High Schools	Students
School A	290
School B	283
School C	383
School D	326
School E	291
School F	344
School G	346
Total	2263

Schools differed in demographic group percentages of students included in the study (Table 3). However, the study population reflected percentages equal to the school populations in most categories. Noted exceptions were (a) the reduction of the Hispanic percentage in the study population at School B ($n = 395$) of three percentage points and (b) increase of the Hispanic percentage in the study population at School G ($n = 437$) of two percentage points. Data for the seven high school principals that were interviewed was gathered during the interview and are presented in Table 7.

Table 7
Principal Gender and Years of Experience

School	Principal Gender	Years Experience as Principal
School A	Male	10
School B	Female	4
School C	Male	3
School D	Female	6
School E	Female	5
School F	Male	5
School G	Female	18

The gender distribution of those interviewed favored females in that more than half of the principals were female (male = 42.8%; female = 57.2 %). The majority of the principals (71.4 %) interviewed had between three and six years experience as a principal. Only two principals interviewed had greater than nine years of experience as a principal. One male had 10 years and one female had 18 years experience as a principal.

Research Question 1

To what extent does the District Benchmark Reading Assessment (DBMRA) predict achievement level on the Florida Comprehensive Assessment Test (FCAT) in Reading for 10th grade for 2009?

A simple linear regression between DBMRA raw score and the FCAT Reading scale score was conducted to answer this question. Data collected to answer Research Question 1 consisted of 10th-grade student raw scores in the November, 2008 DBMRA and the 2009 FCAT Reading scale scores. Students included in the analysis were those at the seven district high schools who had both scores. A Pearson correlation coefficient, r , was computed and analyzed for the sample. A scatter plot was created prior to calculation of the Pearson correlation coefficient to determine whether there was a strong positive correlation, thus ensuring the validity of pursuing the Pearson correlation coefficient. Figure 1 contains the scatter plot of FCAT Reading scale scores (vertical axis) and DBMRA scores (horizontal axis) for the study population.

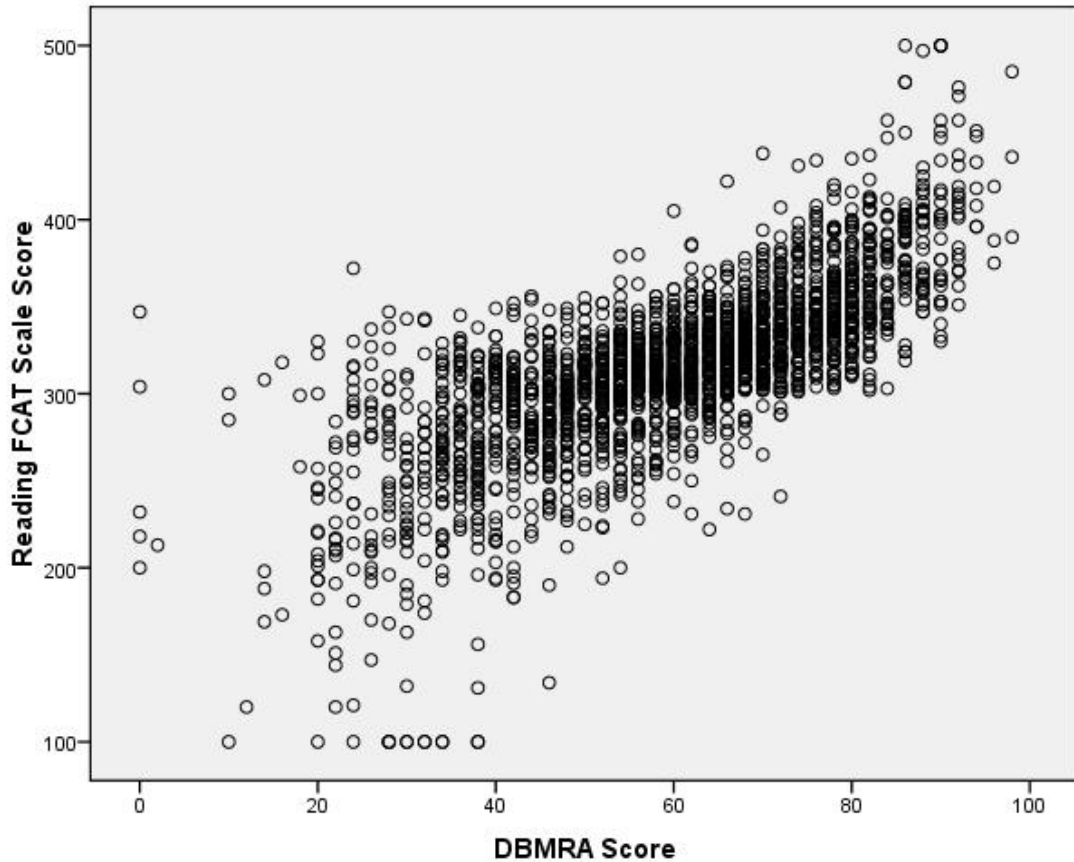


Figure 1. Scatter plot of FCAT Reading scale scores (vertical axis) and DBMRA scores (horizontal axis) for the study population

A strong positive correlation was observed in the scatter plot for the study population ($r = .74$, $n = 2,263$). The Pearson correlation coefficient, ($r = .74$), fit Cohen's (Becker, n.d.) requirements for large ($d \geq .80$) effect size.

The d statistic created by Cohen was a scale for operational definitions of effect size for a more descriptive and qualitative way of discussing statistical effect size in social sciences research. Cohen's d can be calculated from Pearson's Product Moment r

statistic with the following formula $d = 2r/\sqrt{1-r^2}$. Cohen then assigned the resulting d value to the corresponding descriptor of effect size based on a range of d values (Cohen, 1992).

Cohen assigned descriptors to his measure of effect size as small ($d = .2$), medium ($d = .5$), and large ($d = .80$). While Cohen recognized the risk in assigning general label to effect size for use in analysis of power, the categories were considered appropriate for use as descriptors of results (Becker, n.d.). The correlation and resulting effect size of the linear regression in this study indicated that a strong positive relationship existed between the scores students received on DBMRA and FCAT Reading scale scores.

The relationship between the DBMRA and FCAT Reading scale scores was sufficiently strong to utilize linear regression to calculate a mathematical equation for prediction of FCAT Reading scale score based on DBMRA score. The prediction model was considered significant ($F(1, 2261) = 2,690.41, p < .01$) which means the DBMRA score was a good predictor of FCAT Reading scale score. The R^2 value was .54 which indicated that 54% of the variability in FCAT Reading scale score was explained by the model. The model found that the two scores were related with the formula:

$$\text{FCAT Reading scale score} = 187.99 + 2.13(\text{DBMRA})$$

Research Question 2

To what extent does the District Benchmark Reading Assessment (DBMRA) predict achievement level on the Florida Comprehensive Assessment Test (FCAT) in Mathematics for 10th grade for 2009?

Data collected to answer Research Question 2 were 10th-grade student scale scores in the November, 2008 DBMRA and the 2009 FCAT Mathematics scale scores. Students included in the analysis were those at the seven district high schools who had both scores. Research Question 2 was investigated using a simple linear regression between DBMRA and the FCAT Mathematics scale score. The Pearson correlation coefficient was computed and analyzed for the sample. A scatter plot was created prior to calculation of the Pearson correlation coefficient to determine whether there was a strong positive correlation and it was valid to pursue the Pearson correlation coefficient. Figure 2 contains the scatter plot of FCAT Mathematics scale scores (vertical axis) and DBMRA scores (horizontal axis) for the study population.

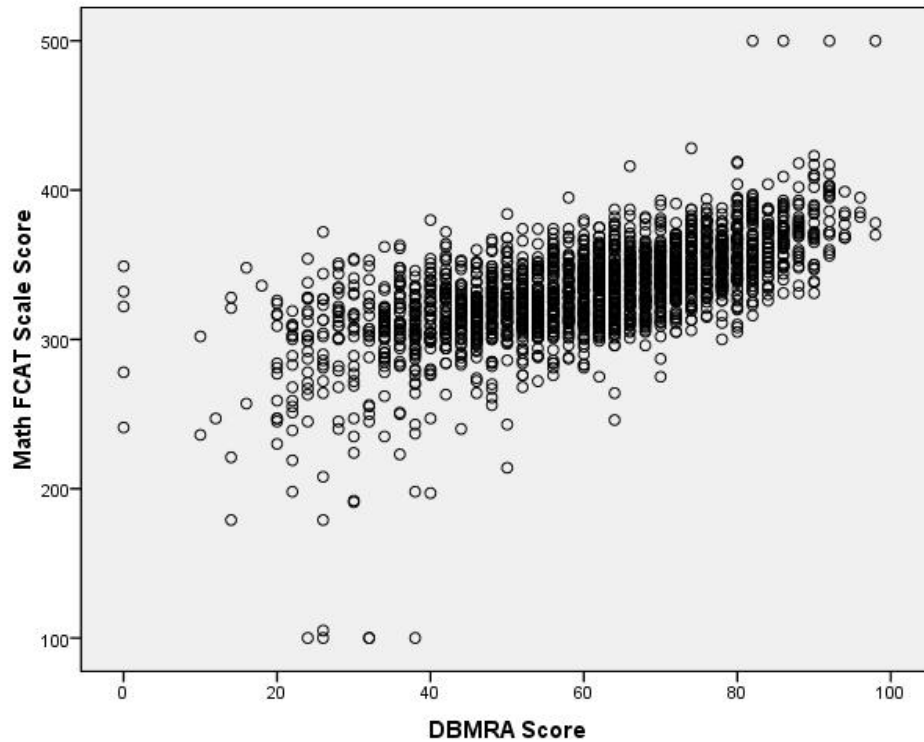


Figure 2. Scatter plot of FCAT Mathematics scale scores (vertical axis) and DBMRA scores (horizontal axis) for the study population

A strong positive correlation can be observed in the scatter plot ($r = .64$, $n=2166$) for the study population. The Pearson correlation coefficient, ($r = .64$), fit Cohen's (Becker, n.d.) requirements for large ($d \geq .80$) effect size. The correlation and resulting effect size indicated that a strong positive relationship existed between the scores students received on DBMRA and FCAT Mathematics scale scores.

The prediction model was considered significant ($F(1,2164) = 1468.90$, $p < .01$) which means that DBMRA score was a good predictor of FCAT Mathematics scale score. The R^2 value of .40 indicated that 40% of the variability in FCAT Mathematics

scale score was explained by the model. The model found that the two scores were related with the formula:

$$\text{FCAT Mathematics scale score} = 259.41 + 1.23(\text{DBMRA})$$

Research Question 3

To what extent, if any, does the relationship between scores on the District Benchmark Reading Assessment (DBMRA) and FCAT Reading differ among high schools in the study?

Data collected to answer Research Question 3 were 10th-grade student scale scores in the November, 2008 DBMRA and the 2009 FCAT Reading scale scores for each school in the study. Students included in the study for each school were those who had both scores. Research Question 3 was investigated using a Pearson correlation coefficient, r , computed and analyzed for the sample at each school. A scatter plot was created prior to calculation of the Pearson correlation coefficient to determine whether there was a sufficiently strong positive correlation to suggest it was valid to pursue the Pearson correlation coefficient. Figure 3 contains the scatter plots of each school for FCAT Reading scale scores (vertical axis) and DBMRA scores (horizontal axis) for the study population.

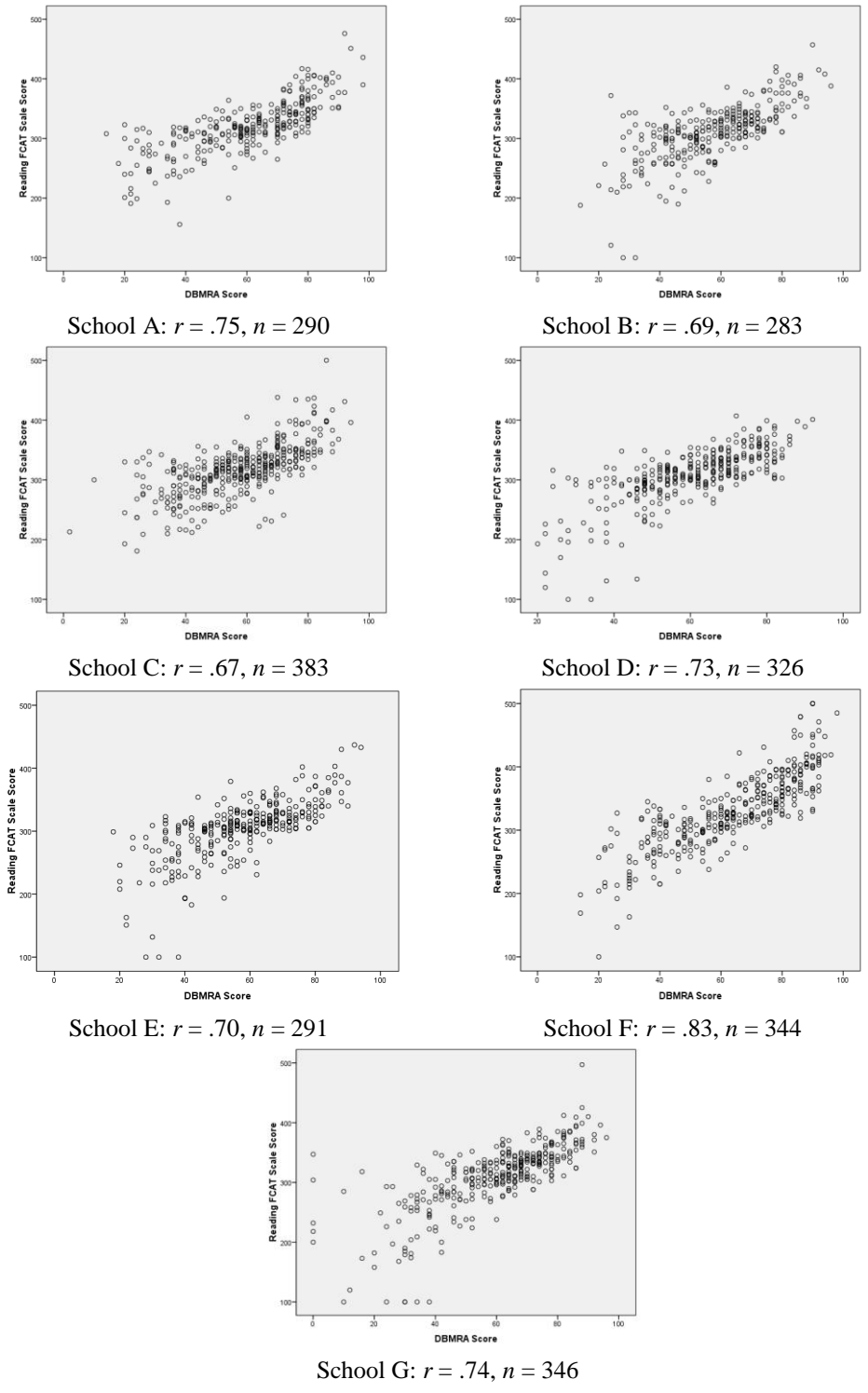


Figure 3. Correlations of DBMRA (x-axis) vs. Reading FCAT (y-axis) by School

The correlation (r), R^2 , and d values, as well as the corresponding Cohen standard for effect size (Becker, n.d.), for the relationship between the November 2008 DBMRA scale score and 2009 FCAT Reading scale score by school are illustrated in Table 8.

Table 8
Pearson Correlation Coefficients, (r), R^2 , d Values, and Cohen's Standard by School

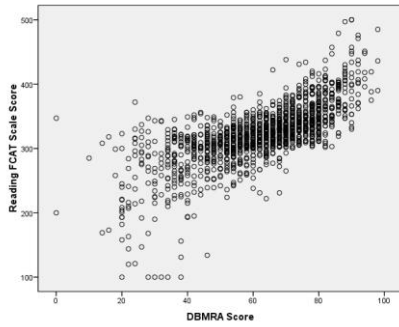
School	Pearson r	R^2	d	Cohen's Standard for Effect Size
School A	.75	.56	>2.0	Large
School B	.69	.48	1.9	Large
School C	.67	.45	1.8	Large
School D	.73	.53	>2.0	Large
School E	.70	.49	2.0	Large
School F	.83	.69	>2.0	Large
School G	.74	.55	>2.0	Large

The correlations ranged between .67 and .83, which indicated fairly strong relationships between the two variables. All correlations were also highly significant ($p < .01$). It should also be noted that correlations signified a relationship only and did not prove that increases in DBMRA scores increase FCAT Reading scores. The R^2 values between .45 and .69 indicated that there was a strong possibility that future outcomes would be similar. Cohen's (Becker, n.d.) standard for effect size indicated a large treatment effect.

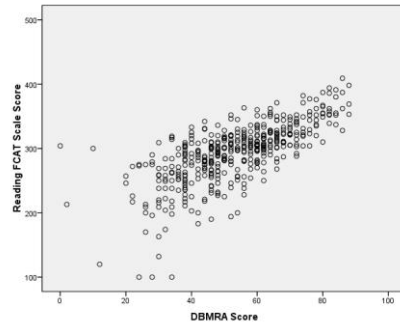
Research Question 4

To what extent, if any, does the relationship between scores on the District Benchmark Reading Assessment (DBMRA) and FCAT Reading differ by AYP subgroups?

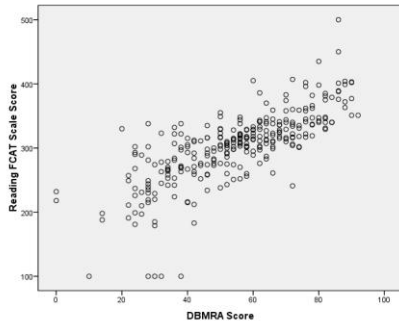
Data used to answer Research Question 4 included 10th-grade student scale scores by sub groups for the November, 2008 DBMRA and the 2009 FCAT Reading scale scores for the sample population in the study. Students included in the analysis were those who had both scores. The use of AYP in the research question was only as a descriptor of the groups used. The analysis categorized data only by the subgroups identified in the AYP process and not the AYP score. Research Question 4 was investigated using a Pearson correlation coefficient, r , computed and analyzed for each subgroup in the sample population. A scatter plot was created prior to calculation of the Pearson correlation coefficient to determine whether there was a strong positive correlation and that it was valid to pursue the Pearson correlation coefficient. Figure 4 contains the scatter plots of each subgroup for FCAT Reading scale scores (vertical axis) and DBMRA scores (horizontal axis) for the subgroup population.



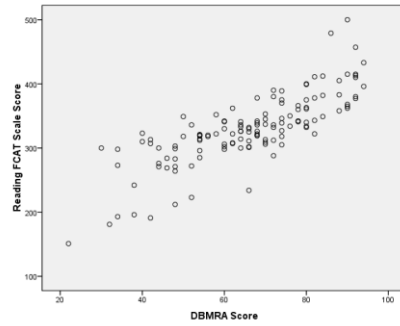
White: $r = .71$, $n = 1,390$



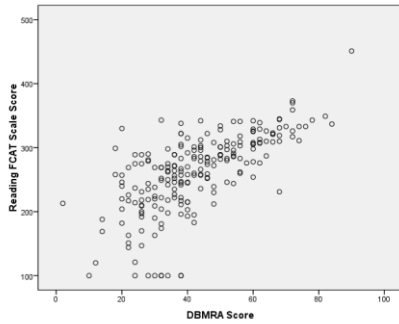
Black: $r = .70$, $n = 450$



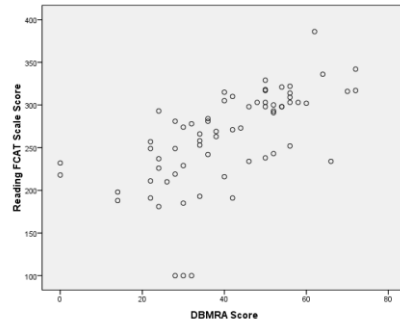
Hispanic: $r = .75$, $n = 292$



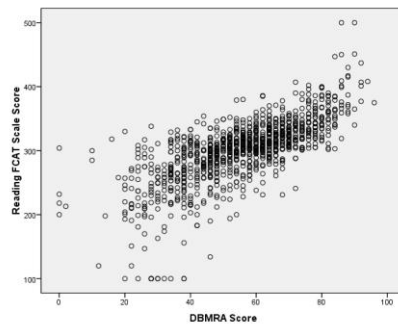
Other: $r = .80$, $n = 131$



SWD Students: $r = .67$, $n = 216$



ELL Students: $r = .63$, $n = 68$



FRL Students: $r = .71$, $n = 1,068$

Figure 4. Correlations of DBMRA (x-axis) vs. Reading FCAT (y-axis) by Subgroup

The correlation (r), R^2 , and d values, as well as the corresponding Cohen standard for effect size (Becker, n.d.), for the relationship between the November 2008 DBMRA scale score and 2009 FCAT Reading scale score by subgroup are illustrated in Table 9.

Table 9
Pearson Correlation Coefficients, (r), R^2 , d Values, and Cohen's Standard by Subgroups

Subgroup	Pearson r	R^2	d	Cohen's Standard for Effect Size
White	.71	.50	>2.0	Large
Black	.70	.49	2.0	Large
Hispanic	.75	.56	>2.0	Large
Other	.80	.64	>2.0	Large
SWD	.67	.45	1.8	Large
ELL	.63	.40	1.6	Large
FRL	.71	.50	>2.0	Large

Note. SWD=Students with Disabilities, ELL=English Language Learners, FRL=Free and Reduced Lunch

The correlations ranged between .63 and .80, which indicated fairly strong correlations between the two variables in every subgroup. All correlations were also highly significant ($p < .01$). It should also be noted that correlations simply signified a relationship existed and could not prove that increases in DBMRA scores increased FCAT Reading scores. The R^2 values between .40 and .64 indicated that there was a strong possibility that future outcomes would be similar. Cohen's (Becker, n.d.) standard for effect size indicated a large treatment effect.

Research Question 5

What strategies do principals of high schools use where the correlation between the November, 2008 District Benchmark Reading Assessment (DBMRA) and the 2009 FCAT score in reading meets or exceeds Cohen's medium effect ($d \geq .50$)?

Data collected for this question was from personal interviews of principals at schools in the study that met the Research Question criteria of Cohen's standard (Becker, n.d.) of $d \geq .50$). Table 10 lists the schools in the study with correlation values, d values, and corresponding Cohen's standard (Becker, n.d.) that were included in the interviews.

Table 10
Pearson Correlation Coefficients, (r), R^2 , d Values, and Cohen's Standard for Interview Schools

School	Pearson r	R^2	d	Cohen's Standard for Effect Size
School A	.75	.56	>2.0	Large
School B	.69	.48	1.9	Large
School C	.67	.45	1.8	Large
School D	.73	.53	>2.0	Large
School E	.70	.49	2.0	Large
School F	.83	.69	>2.0	Large
School G	.74	.55	>2.0	Large

A constant comparison qualitative data analysis described in Chapter 3 was used to analyze the data collected in personal interviews by the researcher to identify themes and trends. Detail on the interview protocol was presented in Chapter 3 and interview questions are contained in Appendix C. Interviewee responses were chunked, coded by trend or theme, and compared to determine the best usable codes. Where the researcher considered the response chunks to be essentially synonymous, they were combined for

coding. Chunked responses that contained elements making them distinctly different were treated separately for coding. Responses were categorized by code and enumerated in both number of responses per code and percentage of respondents who answered in the listed codes. Table 11 presents the constant comparison analysis by interview question with codes, and frequency. The trends indicated in Table 11 in the code column represent themes and trends in open-ended questions without prompt or confirmation from the researcher. A listing of chunks derived from the answers to each question and from which the codes were developed can be found in Appendix G.

Table 11
Constant Comparison Analysis

Interview Question	Code	Number of Responses	%
1A. To what extent is the District Benchmark Reading Assessment implemented in 10 th -grade classrooms with fidelity?	A.Consistently.	5	71
1B. Can you provide examples?	B.Given in 9/10 English.	5	71
	School-wide administration 9-12.	2	29
1C. How do you know?	C.Teach the standards and measure.	6	86
	Indicator of where we are.	2	29
	Remediation at specific time.	2	29
	Curriculum Calendar.	6	86
	Data is turned in.	5	71
	Data meetings.	6	86
	Classroom walk through.	4	57
	Administrators attend and/or run data meetings.	2	29
2A. To what extent do you consider the District Benchmark Reading Assessment an important tool in improved student achievement on FCAT Reading at your school?	A.Use it and rely on it.	5	71
	Results may not be accurate.	2	29
	Strong teacher buy-in.	1	14
2B. Why?	B.Students over-tested.	7	100
	Students not motivated to do best.	7	100
	Test used as grade to encourage best.	2	29
	Make adjustments in instruction.	1	14
	Indicator of where we are.	3	43
	Teachers lack content skills.	2	29
	Teachers give it but don't use it.	1	14
3A. When the data comes in, what do you do with it?	A.Collect teacher data.	7	100
	Collect district and school-wide data.	7	100
	Give data to teachers.	4	57
	Require teacher to pull data.	3	43
3B. Is it ever used to modify instruction and how?	B.Is used to modify.	4	57
	Should be used to modify.	3	43
3C. Can you provide examples?	C.Collaborative groups.	7	100
	Data walls.	2	29
	Department meetings.	4	57
	One-on-one meetings with administrator.	3	43

3C. Can you provide examples? (continued)	Student meetings.	3	43
	Administrator analyzes data.	7	100
	Teachers analyze data.	3	43
	Compare years.	4	57
	Share best practices.	3	43
	School-wide training and skill focus.	5	71
	Analyze weakness and reteach.	2	29
	Celebrate.	1	14
	Focus on next skill.	1	14
4. How do you monitor what you told me takes place?	Classroom walk through monitoring.	6	86
	Administrator assigned to mentor grade/content.	4	57
	Collaborative meetings.	7	100
	Administrator led meetings.	1	14
	Hold trainings.	4	57
	Administrator delivers trainings.	4	57
	Teachers keep data notebooks.	2	29
	Expectation of compliance.	7	100
	Model.	2	29
	Performance evaluations.	1	14
Conversations with teachers.	5	71	

For interview question one as to the use of the DBMRA, 71% of the principals responded that the DBMRA was used consistently as designed in the school. Among the strongest practices emerging from principals to describe how this determination was made were data meetings (86%). Most meetings were conducted and led by teachers with only two administrators (29%) indicating that they both attended and led the meetings. Further determination that the DBMRA was in use with fidelity was the practice of classroom visitations or short walk-through visits (57%) to gather information, trends, and monitor instruction.

Also primary among practices contributing to the effective and consistent use of formative data was the use of curriculum calendars or focus calendars (86%) that outlined the year assigning time to each skill based on the need indicated by prior year data. A majority of principals (86%) responded that this calendar or continuum enabled them to

clearly determine that standards were being taught and that what had been taught was measured.

Interview question two was used to investigate the perception of principals as to the importance of the DBMRA and to what they attributed this perception. The majority of principals (71%) indicated that the DBMRA was a viable tool on which they and their teachers could rely. At the same time, 100% of the principals indicated that the results of the DBMRA might not have been entirely accurate as they said that students were over-tested and, therefore, did not provide their best effort on every test. To combat this, principals employed a number of practices to encourage students to exert their best efforts. These included practices ranging from providing positive reinforcement by facilitating adjustments in instruction (14%) to using the test for a class grade (29%).

Interview question three sought to document the principal practices in the immediate use of the data. Seven principals, (100%), indicated that they collected teacher, school, and district data for the purpose of comparing it and gauging their place among the schools in the district. Principals were almost evenly split in distributing the data to teachers (57%) and requiring the teachers to access their own data (43%). All principals stated that the data either was used to modify instruction (57%) or should be used to modify instruction (43%). Consistently in interview responses, principals (100%) indicated that it was their responsibility to fully understand data and its implications. Additionally, it was important for the principal to compare data from year to year (57%) to chart growth, identify recurring areas of concern, and find points of celebration.

Among the strategies to garner focus and support of data analysis in instruction, the use of data walls or visible representations of the data (29%) were cited. Other ways that principals focused on the use of data to inform and modify instruction was one-on-one meetings with teachers (43%) to specifically analyze teacher data. One of the most cited practices, and one considered to be most valuable, was determining a school-wide skill focus based on data and training all teachers in strategies (71%). Sharing data and best practices (43%) among teachers, re-teaching (29%), and holding departmental content specific meetings (57%) for this purpose were considered to be essential in effective use of formative assessment data.

Interview question four sought to identify how the principal effectively monitored the practices and activities that were taking place at the school regarding data. All principals (100%) indicated that they witnessed and attended collaborative meetings throughout the school at various times and in various contents. Equally strong (100%) was the contention by principals that there was an expectation of compliance from teachers. Principals indicated that modeling (29%), participation in meetings (57%), and training (57%) were successful leadership behaviors in school-wide data usage.

Classroom visits and walk-throughs (86%) were cited as key behaviors in monitoring practices and expectations. Attending collaborative meetings (57%) or leading collaborative meetings (14%) were once again mentioned as typical practices. Among the principals, more than half (57%) indicated that they divided the administrative team and assigned each administrator a grade level or content area to mentor and lead. Several principals (57%) stated that they routinely made opportunities

for training to be delivered. Four principals (57%) indicated that they actually delivered the training to staff based on training they were provided or research they had acquired.

Summary

This chapter has provided a restatement of the purpose of the study and a demographic description of the population. The presentation of the analysis of the data has been organized around the five research questions. A summary and discussion of the findings along with implications and recommendations for practice and future research will be provided in Chapter 5.

CHAPTER 5 SUMMARY, CONCLUSIONS, IMPLICATIONS, AND RECOMMENDATIONS

Introduction

This chapter contains a summary and discussion of the results of the data analysis presented in Chapter 4. The purpose and significance of the study are restated. The summary and discussion of findings are organized around the five research questions and their respective null hypotheses. Also provided are implications and recommendations for future research.

Purpose of the Study

The purpose of the study was to (a) provide information to high school principals and teachers to better understand how students were performing and learning and (b) how to maximize use of the district benchmark assessment in order to modify instruction and positively impact student achievement. An additional focus of this study was to determine best leadership practices in schools where there was a high correlation of DBMRA to FCAT scores. This study expanded on a 2008 study of third grade students (Gaught) to determine if the District Benchmark Reading Assessment (DBMRA) was a predictor of Florida Comprehensive Assessment Test (FCAT) performance. This study was focused on a larger sample of 10th-grade students. The researcher investigated the potential for predicting the FCAT Mathematics scale scores based on the DBMRA scores. Additionally, the study was conducted to determine whether there was a difference in the strength of prediction of DBMRA and FCAT Reading scale scores

between high schools and subgroups. Finally, principals were interviewed to identify best practices related to formative assessment data usage in high schools.

Summary and Discussion of the Findings

Secondary data collected through school district normal testing procedures and demographic data requested for 10th-grade students in the seven high schools included in the study were used to answer research questions one through four. Quantitative data analysis methods were used to analyze the data collected for Research Questions 1-4. Data collected through personal interviews were used to answer Research Question 5. Qualitative analysis was used to analyze the data gathered in the interviews. A discussion of the findings for each research question follows.

Research Question 1

To what extent does the District Benchmark Reading Assessment (DBMRA) predict achievement level on the Florida Comprehensive Assessment Test (FCAT) in Reading for 10th grade for 2009?

Research Question 1 sought to determine the relationship between the DBMRA and the FCAT Reading scale score for 10th-grade students. In order to identify the relationship, data were collected from all 10th graders in the seven public high schools in the target district who had a November, 2008 DBMRA score and a 2009 FCAT Reading scale score.

The relationship between DBMRA and FCAT Reading scores was investigated by calculating a Pearson correlation coefficient and simple linear regression. The analysis

indicated a strong positive correlation ($r = .74$, $n = 2,263$). Linear regression considered the prediction model significant ($F(1, 2261) = 2,690.41$, $p < .01$) indicating that DBMRA was a possible predictor of FCAT Reading scale score. The R^2 value of .54 indicated that 54% of the variability in FCAT Reading scale score was explained by the model.

The null hypothesis for Research Question 1 (H_{01}) was that there was no relationship between the DBMRA and FCAT Reading scale score. The analysis indicated that there was a strong positive relationship between the DBMRA and FCAT Reading, thus enabling the researcher to reject the null hypothesis. These results were consistent with the review of literature in which it was found that the use of formative assessments like the DBMRA, in content areas such as reading, had an impact on student achievement (Black & Wiliam, 1998a, 1998b; Fuchs, Fuchs, & Zumeta, 2008). Furthermore, if FCAT tests the Florida standards and if the DBMRA is a good predictor of FCAT performance, then it must follow that the DBMRA correlates with the standards. This was a primary characteristic of an effective formative assessment found in the literature review (Black & Wiliam, 1998a; Davenport & Anderson, 2002; Fuchs & Fuchs, 1999).

The linear regression model provided a mathematical prediction formula (FCAT Reading scale score = $187.99 + 2.13(\text{DBMRA})$). This finding would suggest further study to determine potential uses of DBMRA and formative assessments like it in classroom instruction to impact student performance.

In summary, data analysis for Research Question 1 indicated that across the survey population of 10th-grade students ($n = 2,263$) in the public high schools in the

target district, the DBMRA as a formative assessment was a possible predictor of student achievement on FCAT Reading. The null hypothesis ($H_01 =$ There is no relationship between DBMRA and FCAT Reading) was rejected as a result of the high correlation ($r = .74$) between the two instruments.

Research Question 2

To what extent does the District Benchmark Reading Assessment (DBMRA) predict achievement level on the Florida Comprehensive Assessment Test (FCAT) in Mathematics for 10th grade for 2009?

Research Question 2 sought to determine the relationship between the DBMRA and the FCAT Mathematics scale score for 10th-grade students. In order to identify the relationship, data were collected from all 10th graders in the seven public high schools in the target district who had a November, 2008 DBMRA score and a 2009 FCAT Mathematics scale score ($n = 2,166$).

The relationship between DBMRA and FCAT Mathematics score was investigated by calculating a Pearson correlation coefficient and simple linear regression. The analysis indicated a strong positive correlation ($r = .64, n = 2,166$). Linear regression considered the prediction model significant ($F(1, 2164) = 1468.90, p < .01$) indicating that DBMRA was a possible predictor of FCAT Mathematics scale score. The R^2 value of .40 indicated that 40% of the variability in FCAT Mathematics scale score was explained by the model.

The null hypothesis for Research Question 2 (H_02) was that there was no relationship between the DBMRA and FCAT Mathematics scale score. The analysis

indicated that there was a strong positive relationship between the DBMRA and FCAT Mathematics, thus enabling the researcher to reject the null hypothesis. These results support findings in the literature review which indicated that formative assessments in reading were related to skill development in other content areas (Marzano, 2004; Schmoker, 1999, 2006). In schools where there was positive achievement as measured by standardized tests, the foundation of instruction in all content areas was vocabulary (Thompson, 2008) and a focus on literacy strategies embedded in the content (Taylor & Moxley, 2006).

Similarly, the findings for Research Question 2 supported the research correlating reading to other content areas such as 10th-grade science (Visone, 2009), fourth grade mathematics (Goddard, 2003), college mathematics (Kelly & Gaustad, 2006), and sixth grade mathematics (Lutz-Doemling, 2007). Results in this study positively correlating FCAT Mathematics to the DBMRA would suggest that it might be possible to generalize across grade levels and content, stating that reading skill is directly related to skill attainment in other content areas.

In summary, data analysis for Research Question 2 indicated that across the survey population of 10th-grade students ($n = 2,166$) in the public high schools in the target district, the DBMRA as a formative assessment was a possible predictor of student achievement on FCAT Mathematics. The null hypothesis (H_0) stating that there was no relationship between DBMRA and FCAT Mathematics, was rejected as a result of the high correlation ($r = .64$) between the two instruments.

Research Question 3

To what extent, if any, does the relationship between scores on the District Benchmark Reading Assessment (DBMRA) and FCAT Reading differ among high schools in the study?

Research Question 3 sought to determine whether there was a difference in the relationship between the DBMRA and the FCAT Reading scale score for 10th-grade students between high schools. In order to identify any differences in the relationship, data were collected from all 10th graders in the seven public high schools in the target district who had a November, 2008 DBMRA score and a 2009 FCAT Reading scale score and calculated by school.

The relationship between DBMRA and FCAT Reading score was investigated by calculating a Pearson correlation coefficient for each school in the study. The analysis indicated strong positive correlations ($r = .67 - .83$). All correlations were considered significant ($p < .01$). Data analysis for Research Question 3 compared strong positive correlations to determine if any differences existed in the strength of the correlation. The null hypothesis (H_03) was that there were no differences between schools in strength of correlation. Using Cohen's (Becker, n.d.) standard, all correlations fell into the large effect size, with only minor d value differences. The d values range from 1.8 to 2.0 indicating very little if any difference between them. Based on this analysis, the researcher concluded that there were no significant differences in the strength of correlation among schools and failed to reject the null hypothesis.

Research Question 4

To what extent, if any, does the relationship between scores on the District Benchmark Reading Assessment (DBMRA) and FCAT Reading differ by AYP subgroups?

Research Question 4 sought to determine whether there was a difference in the relationship between the DBMRA and the FCAT Reading scale score for 10th-grade students between subgroups. In order to identify any differences in the relationship, data were collected from all 10th graders in the seven public high schools in the target district who had a November, 2008 DBMRA score and a 2009 FCAT Reading scale score and calculated by subgroups defined as White, Black, Hispanic, Other, Free/Reduced Lunch (FRL), Students with Disabilities (SWD), and English Language Learners (ELL).

The relationship between DBMRA and FCAT Reading score was investigated by calculating a Pearson correlation coefficient for each subgroup in the study. The analysis indicated fairly strong positive correlations ($r = .63 - .80$). All correlations were considered significant ($p < .01$).

Data analysis for Research Question 4 compared fairly strong positive correlations to determine if any differences existed in the strength of the correlation. All correlations were found to be considered large on Cohen's (Becker, n.d.) standard, the majority (57%) hovering between .70 and .75. The weakest correlations, while still considered to be large on Cohen's (Becker, n.d.) standard of effect size, existed in the ELL subgroup ($r = .63$) and in the SWD subgroup ($r = .67$). This finding is consistent with trends indicated in state FCAT Reading results and AYP reports (Florida Department of Education, 2009). The null hypothesis (H_04) was that there were no

differences between subgroups in strength of correlation. Using Cohen's (Becker, n.d.) standard, all correlations fell into the large effect size, with only minor d value differences. The d values range from 1.6 to 2.0 indicating very little if any difference between them. Based on this analysis, the researcher concluded that there were no significant differences in the strength of correlation among subgroups and failed to reject the null hypothesis.

These results support findings in the literature review that indicated that formative assessments used to monitor student growth were successful for both general education populations and for SWD populations (Foegen et al., 2007). Smith (2005) concurred saying that formative assessments provided data for highly mobile populations and in classrooms where there was a wide variety of students and abilities. Key to the effective use of the formative assessment in largely diverse populations was the feedback provided to the student for full understanding of the gaps in learning and the actions and practice needed to gain proficiency (Heritage, Kim, Vendlinski & Herman, 2009). In ethnically diverse populations and populations where language proficiency could be a barrier, it may be more problematic to convey this understanding and provide the needed interventions.

Research Question 5

What strategies do principals of high schools use where the correlation between the November, 2008 District Benchmark Reading Assessment (DBMRA) and the 2009 FCAT score in reading fall into or above Cohen's medium effect ($d \geq .50$)?

This research question sought to identify and document principal best practices relating to data usage and instructional program modification at schools that had strong correlations between DBMRA and FCAT Reading. The best practices were identified in personal interviews with principals of the targeted schools and analyzed using constant comparison analysis. Table 11, presented in Chapter 4, displays the codes determined in each question and the frequency of response for each code. Interview questions can be found in Appendix C.

Data meetings or meetings for collaboration emerged as a key practice in almost every interview question. These meetings might have been named differently from principal to principal, but they were all essentially collaborative groups of teachers or professional learning communities who shared data and analyzed student progress based on the data. Principals, for the most part, structured their practice so that teachers were leading the meetings and the data discussions. Principals indicated that it was important for both teacher buy-in and learning that the groups be peer directed. This finding was consistent with information documented in the literature review concerning the need for collaborative cultures in schools (Barth, 2006; Fullan, 2001; Irvin et al., 2007; Marzano et al., 2005; Peterson & Deal, 2002). While it was incumbent on the principal to provide the leadership and vision for such collaboration, Ainsworth and Viegut (2006) found that

when teachers investigated and manipulated the data gathered from formative assessments, they were better able and more likely to share and modify instruction based on their findings. Specific modifications made by teachers in classroom instruction to address formative assessment data and to differentiate instruction are unclear in this study as responses from principals did not indicate observation of specific instructional modifications decided on in collaborative meetings.

Classroom visits in short and frequent occurrences were also a recurring theme in the interview responses. Walk-through visits were not part of a structured program designed by vendors to be sold as evaluations or monitoring. The visits were termed *learning walks* that followed the principal's own set of steps and offered feedback based on each principal's focus. While the walks differed in both focus and structure, they had in common the fact that they were frequent, consistent, and specific in focus.

Curriculum calendars that provided structure for instruction and the timeline for the year assisted principals in accomplishing monitoring and school-wide skill focus. Weak skills were instructed first and longest with opportunity to revisit the skill prior to formative assessment administration. Use of such calendars, according to administrators, left very little opportunity for failure to follow prescribed progress monitoring programs and formative assessments as scheduled. It was obvious to administrators that teachers were teaching the standards that the formative assessment (DBMRA) measured. This was consistent with best practices indicated by researchers regarding use and development of formative assessments. Wiggins (2005) held that standards drove all instruction and, therefore, had to be at the heart of formative assessment development. Assessments

should define not only what teachers were teaching, but where students were in the standards and what more they were expected to learn (Popham, 2003). Researchers have stated that a systematic alignment of standards to assessments was critical to the effectiveness of the assessment (Ainsworth & Viegut, 2006; Popham, 2003; Stiggins, 2002; Taylor & Collins, 2003; Wiggins, 2005).

Generally, principals expressed the belief that the DBMRA was valuable in providing a picture of the status of the learner and a prescription of intervention. That students were not performing at optimal levels due to over-testing or lack of motivation was a concern for principals, and they implemented several programs to encourage and motivate students. Veteran principals responded that the best practice was student data conversations and viewed application of the reason for the test to some student reality as a way to motivate students to do their best.

Access to data for teachers emerged as a crucial component to leading a school based on data-driven decision making. The most veteran principal stated that unless teachers were trained to be proficient in the mechanics of accessing the data and expected to routinely look at the data, there would be a general lack of understanding and buy-in. The principal indicated that for there to be a collaborative team approach, or what Senge (2006) referred to as a learning organization, each teacher needed to be part of the process. Principals indicated that they needed to fully understand the data and the implications of the data in order to lead a data-driven school or facilitate use of data in instructional decisions. Comparison of data yielded information about growth and provided opportunities to celebrate success, however small. Celebration was important to

one principal of a school making efforts to reform instruction across content areas for the purpose of sustaining student achievement in the lowest quartile.

Practices intended to build trusting relationships were important and viewed as essential to developing a collaborative culture. One-on-one meetings with teachers to analyze weaknesses and develop a specific action plan to modify instruction and re-teach content were successful in building a team relationship between administrator and teacher. Professional development was considered to be a necessary component to changing instruction, building strong research-based instructional strategies, and enhancing the skill in specific content areas. This practice is aligned to findings documented in the review of literature that indicated the importance of professional development in building a culture of collaboration (Easton, 2008).

According to Senge (2006), best administrative practices made it essential for teachers to be part of the learning organization. In the targeted schools, this would mean that teachers would be active stakeholders in the progress monitoring and improvement of student achievement. While principals did not expressly state that professional learning communities (PLCs) were ongoing at the school, the description of the function of the groups and meetings were indicative of this concept. Such practices achieved what Barth (2006) characterized as collegial groups that functioned on a professional level to discuss their craft and enhance their practice. Opportunities for collaborative practice provided by administrators were a critical role of the principal. The administrator served as a cheerleader and liaison between a typical high school culture where each teacher had an isolated focus and an accountability culture where all were responsible for the

achievement of students (Firestone, 2009). Such opportunities enabled those in the organization to move from operating in isolation or in simply congenial relationships to true professional learning communities (Barth, 2006; Senge, 2006).

The principal responses indicated that collaboration opportunities were important, but that an expectation of compliance was also present. At first analysis, this appeared to be paradoxical as one practice, collaboration, is a facilitative leadership approach while the other, expectation of compliance, appears to be a top-down authoritative management approach. Upon further analysis, however, it became apparent that through modeling, participation in meetings, and training, the principal formed and outlined the culture of participation where the expectation of being part of the *team* was the norm in the culture. Assignment of administrators to content teams or grade levels enabled administrators to be active participants in the team collaboration. Part of the role of this mentorship was to attend meetings, facilitate data discussion, assist in identifying and planning corrective actions or needed training, and provide resources and support to teacher groups to enable them to be successful.

Training was cited numerous times as a crucial factor in leading the school in data driven instructional decision making. Principals stated that they consistently held training on the designated district training days and expected all teachers to attend. Principals attended the training as a participant with teachers. The strongest principal practice relating to training was the principal as training presenter. These principals indicated that they gained credibility by being the trainer and were able to not only model

the expectation but convey that they knew what they would be looking for in terms of instruction.

In summary, several best practices emerged from the interviews with principals. First, in effective schools, data were readily available and easily accessible to teachers. Principals also insured that teachers received training to access the data. Principals who successfully facilitated data usage were very comfortable with data themselves and analyzed the data prior to teachers receiving it so that they could lead discussions with teachers. While principals agreed that students were over-tested in the current era of accountability, common to successful principals was the belief that the formative assessment was a valid and viable instrument that, when used correctly, could positively impact student achievement.

Effective principals who used data to drive instruction provided opportunities, support, and clear expectations for collaborative groups or small learning communities of professionals within the school. In all cases, principals attended the data meetings or led them, being an active participant in either practice. Expectations for behaviors in the collaborative meetings were focused on sharing. Data were shared, best practices were shared, weaknesses were shared, and a plan of action was jointly developed to impact instruction.

Frequent short classroom visits, whether called classroom walk-through or evaluations, were essential to monitoring the practices and the instruction for fidelity and quality. Principals thought that visibility in the classroom and school, to both teachers and students, was helpful to furthering the collaborative atmosphere and the knowledge

necessary to lead the school. The principal as participant rather than observer was an important concept to the characterization of the effective principal.

Finally, the focus on professional development and the conveyance of its importance was paramount to the practice of the effective principal. Principals recognized the need for professional development to enhance teacher skill and build quality instructional programs. All principals focused on the data to identify the training focus. In many cases, when data indicated school-wide skill weakness, principals facilitated professional development that identified research based strategies to address the skill and trained all teachers to implement the strategy in their content areas. Those principals that delivered the training developed an even more powerful collaborative relationship with teachers by not only building credibility, but by putting themselves in front of teachers doing what teachers do on a daily basis, opening themselves to judgment and ridicule, but at the same time solidifying the team relationship.

Implications and Recommendations for Practice

The data in this study supported the research regarding the effectiveness of formative assessments in instruction and prediction by documenting the strong correlation between the target district's formative assessment in reading, DBMRA, and the FCAT Reading scale score. Findings suggested that formative assessments, when developed to meet the criteria documented in research, were effective tools to gain data on student learning and modify instruction in reading. Formative assessment data in reading was able to be used to identify student weaknesses for the purpose of remediation

and intervention to positively impact student achievement. Additionally, the data indicated that prediction of the standardized test (FCAT) performance was possible with the formative assessment (DBMRA) score for students at all high schools and in all subgroups.

Though improvement of reading was important, perhaps a more important contribution of this study was the documentation that reading skill and achievement had a strong, positive correlation to student achievement in mathematics. The study added to limited research available showing that reading achievement on the formative assessment (DBMRA) could predict mathematics achievement (FCAT). When scores were raised on the formative assessment, corresponding increases in standardized mathematics scores could be documented for students at all high schools and in all subgroups.

The implications of these findings are two-fold. First, the data would imply that in order to raise reading achievement, formative assessments should be used consistently in intensive reading and language arts classes. Data from these assessments should be used to prescribe and implement interventions, thus raising formative assessment scores that would then translate to standardized test score increase. Second, it may be possible that reading skill proficiency will manifest itself in skill proficiency in mathematics. Reading across content areas should be incorporated into the instructional program in mathematics and teachers should receive training on literacy strategies and how to build them into mathematics lessons.

Another important contribution and implication for educators of this study were the findings that there were no differences in the strength of correlation between schools

or between subgroups. This was important because lack of achievement has often been justified by subgroup or school location, administration, size, age, or other factors. Data in this study suggested that there was no difference in schools or subgroups in terms of the strength of formative assessments (DBMRA) as predictors of the standardized test (FCAT).

Principal interviews offered several suggestions to administrators in terms of best practices. It would appear that the first role of the effective principal would be to establish, build, cultivate, and nurture the culture of the organization that thrives on collaborative and participative efforts. Several practices have contributed to that environment. Among the strongest are (a) knowing the data and being comfortable with it, (b) participating and leading data or collaborative meetings, and (c) scheduling or delivering professional development designed in response to data. Most important in principal best practices was the concept that the practice was consistent and pervasive throughout the school, monitored through classroom visits and ongoing conversations, thereby making it part of the cultural norm.

Recommendations for Future Research

Further research is suggested in the following areas:

1. A similar study needs to be conducted in other districts in Florida where formative assessments exist to determine if the relationship between formative assessments and FCAT is consistent.

2. Studies should be conducted measuring the relationship of reading skill and achievement to skill and achievement in other content areas such as science and social studies at the high school level.
3. Similar studies should be conducted at the middle school level to determine the strength of relationship between formative assessments and FCAT.
4. Studies should be conducted investigating existing formative assessments that show a strong correlation to FCAT scores to determine the most effective format, question content, length, and administration time frame for the formative assessments.
5. Similar studies need to be conducted to identify a rubric for creating an effective formative assessment that is strongly correlated to FCAT achievement.
6. Further research needs to be conducted to identify administrative best practices at all levels that strongly relate to increased student achievement across content areas.
7. Further study is recommended to determine teachers' use of formative assessment data to change and differentiate instruction.
8. A study is recommended to investigate students' perceptions of the value of formative assessments and use of the assessment in the students' learning.

Summary

This study was intended to be of value to the targeted district by validating the district assessment as a viable formative assessment that could predict FCAT reading scores for 10th-grade students. Additionally, the study was intended to identify the administrative best practices that resulted in growth between the district assessment score, predicted FCAT score, and the actual FCAT score. Data from this study should be useful to educational systems in general that are attempting to build a program of formative assessments for the purpose of state test score prediction and student achievement gains as specified by NCLB. The administrative best practices that have been documented can be replicated and modeled by other administrators in high schools.

APPENDIX A
SCHOOL DISTRICT APPROVAL



June 25, 2009

Guidance & Testing

215 S.E. 6th Street • Ocala, FL 34471
(352) 671-7157 • Fax (352) 671-7587
F R S (800) 955-8770 (voice) • (800) 955-8771 (TTY)

To Whom It May Concern:

Ms. Marilyn Underwood has proposed to conduct research utilizing data from the Marion County School District in partial fulfillment of requirements for a Doctoral degree at the University of Central Florida. As Executive Director of Staff Development of the Marion County Public Schools, she has a legitimate educational interest in these data, and she understands statutory and ethical requirements regarding the use of confidential student data.

The student data sets that Ms. Underwood proposes to analyze include the following:

- Marion County Benchmark Reading Assessment student achievement data for grades 3-10
- FCAT Reading student achievement data for grades 3-10

In addition, aggregate school and district reports include the following:

- District attendance data
- District and school FCAT Reading data
- District and school AYP data
- District and school "START" reports (comprehensive reports of multiple educational data elements)

Also, Ms. Underwood plans to conduct interviews with selected school principals to glean additional information regarding use of assessment data.

Please consider this letter as conditional approval for Ms. Underwood to conduct her research project in the Marion County School District. This approval is contingent on receipt of approval of her project from the Institutional Review Board of the University of Central Florida.

Ms. Underwood has our best wishes for a successful project.

Yours truly,

Janet Weldon
Director of Guidance and Assessment

APPENDIX B
DISTRICT BENCHMARK READING ASSESSMENT (DMBRA)
SUMMARY REPORT

View Test Summary : **Print Test Summary :****Test name :** 10RE-DBMA-06**Subject :** Language Arts**Framework:** Marion County Curriculum/Florida
SSS**Grade/Course :** Grade 10**Description :****Test Summary**

Code	Description	Items	Total Pts	Required Pts
FLLA10-LA.A.1.4.2	Selects and uses strategies to understand words and text, and to make and confirm inferences from wh	11	11	8.8
FLLA10-IA.A.2.2.7	Recognizes the use of comparison and contrast in a text. (FCAT - Gr. 3-5 Bench)	2	2	1.6
FLLA10-LA.A.2.4.1	Determines the main idea and identifies relevant details, methods of development, and their effectiv	9	9	7.2
FLLA10-LA.A.2.4.2	Determines the author purpose and point of view and their effects on the text. (FCAT - Includes LA	6	6	4.8
FLLA10-LA.A.2.4.4	Locates, gathers, analyzes, and evaluates written information for a variety of purposes, including r	5	5	4.0
FLLA10-LA.A.2.4.7	Analyzes the validity and reliability of primary source information and uses the information appropri	4	4	3.2
FLLA10-LA.A.2.4.8	Synthesizes information from multiple sources to draw conclusions. (FCAT)	3	3	2.4
FLLA10-IA.E.2.2.1	Recognizes cause and effect relationships in literary texts. (FCAT - Gr. 3-5 Bench)	6	6	4.8
FLLA10-IA.E.2.4.1	Analyzes the effectiveness of complex elements of plot, such as setting, major events, problems, con	4	4	3.2

Answer Key

Q. No.	Correct Option(s)	Type	Standard(s) Code	Standard Description
1	4	Single Choice Single Select	FLLA10-LA.A.1.4.2	Selects and uses strategies to understand words and text, and to make and confirm inferences from wh
2	1	Single Choice Single Select	FLLA10-LA.A.1.4.2	Selects and uses strategies to understand words and text, and to make and confirm inferences from wh
		Single Choice Single	FLLA10-	Selects and uses strategies to understand words and

3	4	Select	LA.A.1.4.2	text, and to make and confirm inferences from wh
4	2	Single Choice Single Select	FLA10-LA.A.2.4.1	Determines the main idea and identifies relevant details, methods of development, and their effectiv
5	2	Single Choice Single Select	FLA10-LA.A.2.4.2	Determines the author purpose and point of view and their effects on the text. (FCAT - Includes LA
6	3	Single Choice Single Select	FLA10-LA.A.2.4.4	Locates, gathers, analyzes, and evaluates written information for a variety of purposes, including r
7	4	Single Choice Single Select	FLA10-LA.A.2.4.8	Synthesizes information from multiple sources to draw conclusions. (FCAT)
8	1	Single Choice Single Select	FLA10-LA.E.2.2.1	Recognizes cause-and-effect relationships in literary texts. (FCAT - Gr. 3-5 Bench)
9	1	Single Choice Single Select	FLA10-LA.A.2.4.7	Analyzes the validity and reliability of primary source information and uses the information appropr
10	1	Single Choice Single Select	FLA10-LA.A.1.4.2	Selects and uses strategies to understand words and text, and to make and confirm inferences from wh
11	3	Single Choice Single Select	FLA10-LA.A.2.4.1	Determines the main idea and identifies relevant details, methods of development, and their effectiv
12	1	Single Choice Single Select	FLA10-LA.A.1.4.2	Selects and uses strategies to understand words and text, and to make and confirm inferences from wh
13	4	Single Choice Single Select	FLA10-LA.E.2.2.1	Recognizes cause-and-effect relationships in literary texts. (FCAT - Gr. 3-5 Bench)
14	1	Single Choice Single Select	FLA10-LA.A.2.4.1	Determines the main idea and identifies relevant details, methods of development, and their effectiv
15	1	Single Choice Single Select	FLA10-LA.E.2.4.1	Analyzes the effectiveness of complex elements of plot, such as setting, major events, problems, con
16	4	Single Choice Single Select	FLA10-LA.A.1.4.2	Selects and uses strategies to understand words and text, and to make and confirm inferences from wh
17	3	Single Choice Single Select	FLA10-LA.A.2.2.7	Recognizes the use of comparison and contrast in a text. (FCAT - Gr. 3-5 Bench)
18	3, 1	Multiple Choice Single Select	FLA10-LA.A.2.4.2	Determines the author purpose and point of view and their effects on the text. (FCAT - Includes LA
19	1	Single Choice Single Select	FLA10-LA.A.1.4.2	Selects and uses strategies to understand words and text, and to make and confirm inferences from wh
20	4	Single Choice Single Select	FLA10-LA.A.2.4.1	Determines the main idea and identifies relevant details, methods of development, and their effectiv
21	3	Single Choice Single Select	FLA10-LA.A.2.4.2	Determines the author purpose and point of view and their effects on the text. (FCAT - Includes LA
22	2	Single Choice Single Select	FLA10-LA.A.2.4.1	Determines the main idea and identifies relevant details, methods of development, and their effectiv
23	1	Single Choice Single Select	FLA10-LA.A.2.4.4	Locates, gathers, analyzes, and evaluates written information for a variety of purposes, including r
24	3, 1	Multiple Choice Single Select	FLA10-LA.A.2.4.7	Analyzes the validity and reliability of primary source information and uses the information appropr
25	3	Single Choice Single Select	FLA10-LA.E.2.2.1	Recognizes cause-and-effect relationships in literary texts. (FCAT - Gr. 3-5 Bench)

26	2	Single Choice Single Select	FLA10-LA.E.2.2.1	Recognizes cause-and-effect relationships in literary texts. (FCAT - Gr. 3-5 Bench)
27	3	Single Choice Single Select	FLA10-LA.A.2.4.4	Locates, gathers, analyzes, and evaluates written information for a variety of purposes, including r
28	2	Single Choice Single Select	FLA10-LA.A.1.4.2	Selects and uses strategies to understand words and text, and to make and confirm inferences from wh
29	3	Single Choice Single Select	FLA10-LA.A.2.4.8	Synthesizes information from multiple sources to draw conclusions. (FCAT)
30	3	Single Choice Single Select	FLA10-LA.A.2.4.2	Determines the author purpose and point of view and their effects on the text. (FCAT - Includes LA
31	4	Single Choice Single Select	FLA10-LA.E.2.4.1	Analyzes the effectiveness of complex elements of plot, such as setting, major events, problems, con
32	3	Single Choice Single Select	FLA10-LA.E.2.4.1	Analyzes the effectiveness of complex elements of plot, such as setting, major events, problems, con
33	2	Single Choice Single Select	FLA10-LA.A.2.4.7	Analyzes the validity and reliability of primary source information and uses the information appropr
34	1	Single Choice Single Select	FLA10-LA.A.2.4.8	Synthesizes information from multiple sources to draw conclusions. (FCAT)
35	3	Single Choice Single Select	FLA10-LA.E.2.2.1	Recognizes cause-and-effect relationships in literary texts. (FCAT - Gr. 3-5 Bench)
36	3	Single Choice Single Select	FLA10-LA.E.2.4.1	Analyzes the effectiveness of complex elements of plot, such as setting, major events, problems, con
37	4	Single Choice Single Select	FLA10-LA.A.1.4.2	Selects and uses strategies to understand words and text, and to make and confirm inferences from wh
38	2	Single Choice Single Select	FLA10-LA.A.2.4.7	Analyzes the validity and reliability of primary source information and uses the information appropr
39	3	Single Choice Single Select	FLA10-LA.A.2.4.4	Locates, gathers, analyzes, and evaluates written information for a variety of purposes, including r
40	4	Single Choice Single Select	FLA10-LA.A.2.4.1	Determines the main idea and identifies relevant details, methods of development, and their effectiv
41	3	Single Choice Single Select	FLA10-LA.A.2.2.7	Recognizes the use of comparison and contrast in a text. (FCAT - Gr. 3-5 Bench)
42	2	Single Choice Single Select	FLA10-LA.A.1.4.2	Selects and uses strategies to understand words and text, and to make and confirm inferences from wh
43	2	Single Choice Single Select	FLA10-LA.A.2.4.1	Determines the main idea and identifies relevant details, methods of development, and their effectiv
44	4	Single Choice Single Select	FLA10-LA.A.2.4.2	Determines the author purpose and point of view and their effects on the text. (FCAT - Includes LA
45	3	Single Choice Single Select	FLA10-LA.A.2.4.4	Locates, gathers, analyzes, and evaluates written information for a variety of purposes, including r
46	3	Single Choice Single Select	FLA10-LA.A.2.4.2	Determines the author purpose and point of view and their effects on the text. (FCAT - Includes LA
47	4	Single Choice Single Select	FLA10-LA.A.1.4.2	Selects and uses strategies to understand words and text, and to make and confirm inferences from wh
48	2	Single Choice Single	FLA10-	Determines the main idea and identifies relevant

		Select	LA.A.2.4.1	details, methods of development, and their effectiv
49	4	Single Choice Single Select	FLA10-LA.A.2.4.1	Determines the main idea and identifies relevant details, methods of development, and their effectiv
50	2	Single Choice Single Select	FLA10-LA.E.2.2.1	Recognizes cause-and-effect relationships in literary texts. (FCAT - Gr. 3-5 Bench)

APPENDIX C
ADMINISTRATOR INTERVIEW QUESTIONS

ADMINISTRATOR INTERVIEW QUESTIONS (GRADE 10)

1. To what extent is the District Benchmark Reading Assessment implemented in 10th-grade classrooms with fidelity?
 - Provide examples
 - How do you know?

2. To what extent do you consider the District Benchmark Reading Assessment an important tool in improved student achievement on FCAT Reading at your school?
 - Why do you think that?

3. When the data comes in, what do you do with it?
 - Is it ever used to modify instruction? How?
 - Provide examples.

4. How do you monitor what you told me takes place?

APPENDIX D
INSTITUTIONAL REVIEW BOARD APPROVAL



University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Exempt Human Research

From: **UCF Institutional Review Board #1
FWA00000351, IRB00001138**
To: **Marilyn K. Underwood**
Date: **February 22, 2010**

Dear Researcher:

On 2/22/2010, the IRB approved the following activity as human participant research that is exempt from regulation:

Type of Review:	Exempt Determination
Project Title:	The Relationship of the Tenth Grade District Progress Monitoring Assessment Scores to Florida Comprehensive Assessment Test Scores in Reading and Mathematics for 2009
Investigator:	Marilyn K. Underwood
IRB Number:	SBE-10-06724
Funding Agency:	
Grant Title:	
Research ID:	N/A

This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these changes affect the exempt status of the human research, please contact the IRB. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

In the conduct of this research, you are responsible to follow the requirements of the [Investigator Manual](#).

On behalf of Joseph Bielitzki, DVM, UCF IRB Chair, this letter is signed by:

Signature applied by Joanne Muratori on 02/22/2010 03:06:14 PM EST

IRB Coordinator

APPENDIX E
PRINCIPALS' INFORMED CONSENT

SCHOOL NAME>
<ADDRESS>
<CITY, STATE, ZIP>

Dear Educator:

Thank you for taking the time to participate in this important study about leadership best practices in Marion County, Florida, high schools related to formative assessment data usage for student achievement on FCAT. You are among approximately 10 school level administrators who have been invited to provide input for this research. This study will contribute to understanding of how meaningful best practices regarding data usage can positively impact student achievement and can be replicated in other schools.

I will be available explain this research study to you. Whether or not you take part is up to you. You can agree to take part now and later change your mind. Whatever you decide it will not be held against you. Feel free to ask all the questions you want before you decide.

The study is confidential as are your responses. The research you may take part in will consist of a four question personal interview focused on the practices, policies, and behaviors that you have implemented relative to formative assessment data. Viewing of any personally identifiable information will be limited to me, the researcher. There are no anticipated risks or benefits to participating in this study.

If you have any questions about this study, please contact me at marilyn.underwood@marion.k12.fl.us. My faculty advisor, Dr. Rosemarye Taylor, may be contacted by phone at (407) 823-1469 or by email at rtaylor@mail.ucf.edu. Research at the University of Central Florida involving human participants is carried out under the oversight of the Institutional Review Board (IRB). Questions or concerns about research participants' rights may be directed to the UCF Institutional Review Board Office at the University of Central Florida, Office of Research and Commercialization, 12201 Research Parkway, Suite 501, Orlando, FL 32826-3246. The phone numbers are (407) 823-2901 or (407) 882-2276.

You may also talk to them for any of the following:

- Your questions, concerns, or complaints are not being answered by the researcher.
- You cannot reach the researcher.
- You want to talk to someone besides the researcher.
- You want to get information or provide input about this research.

I will be contacting you to set up a personal interview appointment.
Thank you in advance for taking the time to participate.

Best Regards,

Marilyn K. Underwood
Doctoral Candidate, University of Central Florida
Executive Director, Staff Development

Participant Signature

Date

APPENDIX F
SUMMARY OF RESEARCH



EXPLANATION OF RESEARCH

Title of Project: The Relationship of 10th-Grade District Progress Monitoring Assessment Scores to Florida Comprehensive Assessment Test Scores in Reading and Mathematics for 2009
Principal Investigator: Marilyn K. Underwood
Other Investigators:

Faculty Supervisor: Dr. Rosemarye Taylor

You are being invited to take part in a research study. Whether you take part is up to you.

The purpose of this research is to identify administrative best practices relating to data usage in order to improve student achievement in reading using the District Benchmark Reading Assessment. Participants will be interviewed to determine what actions and best practices they engage in as the instructional leader of the school using data. There will be four interview questions. Interviews will be conducted one-on-one with the researcher coming to the location of the administrators' choice at a mutually acceptable time.

All responses to the interview will be confidential and no identifying information will be used in the study report.

The time needed to complete the interview is approximately 30 minutes. No further time commitment will be required.

Study contact for questions about the study or to report a problem: If you have questions, concerns, or complaints, contact Marilyn K. Underwood, Doctoral Student, Educational Leadership, College of Education, (352) 816-1127 or Dr. Rosemarye Taylor, Faculty Supervisor, Department of Education at (407) 823-2426 or by email at rtaylor@mail.ucf.edu.

IRB contact about your rights in the study or to report a complaint: Research at the University of Central Florida involving human participants is carried out under the oversight of the Institutional Review Board (UCF IRB). This research has been reviewed and approved by the IRB. For information about the rights of people who take part in research, please contact: Institutional Review Board, University of Central Florida, Office of Research & Commercialization, 12201 Research Parkway, Suite 501, Orlando, FL 32826-3246 or by telephone at (407) 823-2901.

APPENDIX G
CONSTANT COMPARISON RESPONSE DETAIL

Question 1: To what extent is the District Benchmark Reading Assessment implemented in 10th grade classrooms with fidelity? Can you provide examples? How do you know?

Chunk	Code
Consistently 100 % Fully implemented Close to 100 %	Consistently
In English class Given in 9 th and 10 th English	Given in 9/10 English
Stop school to administer Given in every class Given every Tuesday Given 9-12	School-wide administration 9-12
Teach the standards tested and then measure Don't do any test prep, just teach standards	Teach the standards and measure
Test early to see where we are A thermometer to gauge instruction	Indicator of where we are
Follow focus calendar Follow curriculum map Use curriculum calendar	Curriculum calendar
Data comes in Data is turned in to scan Teachers are expected to bring the data Inspect what we expect	Data is turned in
Data meetings held to share data Department meetings share data	Data meetings
Classroom visits and walk through Go into the class to see them test Remediation happens regularly Remediate on club day Remediate in reading class	Classroom walk through
Administrative meeting to talk about testing Administrators facilitate data meetings Administrators attend data meetings	Administrators attend and/or run data meetings

To what extent do you consider the District Benchmark Reading Assessment an important tool in improved student achievement on FCAT Reading at your school? Why?

Chunk	Code
Great concept that helps Important Use it and rely on it Validated test with usable results	Use it and rely on it
Could be more credible Results may not be 100%	Results may not be accurate
Teachers have bought into the concept	Strong teacher buy-in
Kids take too many tests We over-test students	Students over-tested
Kids aren't motivated	Students not motivated to do best

Unsure if students actually do their best Not used as a grade so low motivation for kids Why should they do their best if it doesn't matter	
We use the test as a grade to motivate Kids take it seriously because it is graded	Test used as a grade to encourage best
Use it to make adjustments in instruction	Make adjustments in instruction
Tells us where we are Teachers can be held accountable for the learning	Indicator of where we are
Teachers don't understand the skills in the subject Teachers don't understand how the skill is tested	Teachers lack content skills
Teachers give it but don't use the results	Teachers give it but don't use it

When the data comes in, what do you do with it? Is it ever used to modify instruction and how? Can you provide examples?

Chunk	Code
Pull data for the teachers	Collect teacher data
Pull data for the school Pull data for the district	Collect district and school-wide data
Provide the data for the teacher It is easier to just give the teachers the data We provide the data to the teacher so they have it	Give data to teachers
Teachers are required to pull their own data Teachers have to come to meetings with their data Teachers have to have their data	Require teachers to pull data
If poor in an area, they change Re-teach if students did not get it Share data so that they can change if needed	Is used to modify
Expected to meet and share Grade or content areas should talk about what needed	Should be used to modify
Hold data meetings among grades and/or contents Collaborative groups PLCs Collaborative planning time	Collaborative groups
Post results so that all can see Have data walls up and visible	Data walls
Departments share Department meetings	Department meetings
Sit down with teachers one-on-one to review data Have individual meetings Talk with teachers about their data	One-on-one meetings with administrator
Administrators talk to students about results One-on-one student meetings F-CHATS	Student meetings
Administrators analyze the data	Administrator analyzes data
Teachers analyze their data Teachers look at their data Teachers determine class achievement	Teachers analyze data
Data is looked at to compare years Look to see if we did better than last year in the skill Try to see if we are doing better to guess FCAT	Compare years

Meet to share data and lessons Share best practices in groups Share best practices at faculty meetings Look at weakness and get best practices for that area	Share best practices
Look at school weaknesses and make a plan to fix it Present instructional strategies school-wide Train everyone in a strategy to use across contents	School-wide training and skill focus
Teachers analyze the weak areas in the content Teachers re-teach weak areas Teachers develop another lesson to address weakness	Analyze weakness and re-teach
Use the data to celebrate Find something good in the data and celebrate it	Celebrate
Look at the next skill and hand out strategies for that	Focus on next skill

How do you monitor what you told me takes place?

Chunk	Code
Walk through classes Classroom walk through Visit classes to see what is going on In and out of classes on a continual basis	Classroom walk through monitoring
Administrators are assigned a grade to monitor Administrators are assigned a content to monitor We mentor a grade level We supervise a content area or department	Administrator assigned to mentor grade/content
Collaborative groups Hold and attend data meetings Sharing sessions Department meetings to share	Collaborative meetings
Administrators lead the meetings	Administrator led meetings
Hold trainings Set up trainings Train on assigned training days	Hold trainings
Administrators give the training Administrators do research and then give a training Administrators attend a training and then train staff	Administrator delivers training
Teachers have to keep a data notebook	Teachers keep data notebooks
They do it because it is what is expected Inspect what you expect It is the system and they do it I expect them to comply with what is set up	Expectation of compliance
I talk about what is important I always look at data and ask questions Teachers see me excited	Model
Use the district performance evaluation system	Performance evaluations
Informal conversations I talk to teachers about the data	Conversations with teachers

LIST OF REFERENCES

- Ainsworth, L. (2004). *Power standards: Identifying the standards that matter the most*. Englewood, CO: Advanced Learning Press.
- Ainsworth, L., & Viegut, D. (2006). *Common formative assessments: How to connect standards-based instruction and assessment*. Thousand Oaks, CA: Corwin Press.
- Allen, D., Ort, S. W., & Schmidt, J. (2009). Supporting classroom assessment practice: Lessons from a small high school. *Theory Into Practice, 48*(1), 72-80.
- Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. G. (1985). *Becoming a nation of readers: The report of the commission on reading*. Washington: The National Institute of Education.
- Barth, R. S. (2006). Improving relationships within the schoolhouse. *Educational Leadership, 63*(6), 8-13.
- Barton, P. E. (2006). Needed: Higher standards for accountability. *Educational Leadership, 64*(3), 28-31.
- Becker, L. (n.d.). *Effect size*. Retrieved May 5, 2010, from University of Colorado, Colorado Springs, web site:
<http://www.uccs.edu/~faculty/lbecker/es.htm#III.%20Effect%20size%20measures%20for%20two%20dependent>
- Beers, K. (2003). *When kids can't read: What teachers can do*. Portsmouth, NH: Heinemann.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). *Phi Delta Kappan, 86*(1), 9-21.
- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-68.
- Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139-148.
- Bracey, G. W. (1998). The eighth Bracey report on the condition of public education. *Phi Delta Kappan, 80*(2), 112-131.
- Bracey, G. W. (2003). *What you should know about the war against America's public schools*. Boston: Allyn & Bacon.

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- Cotton, K. (2003). *Principals and student achievement: What the research says*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Coutant, C., & Perchemildes, N. (2005). Strategies for teen readers. *Educational Leadership*, 63(2), 42-47.
- Crow, T. (2009). Proof positive: The keys to successful school change are in our grasp. *The Journal of the National Staff Development Council*, 30(5), 12-18.
- Darwin, M., & Fleischman, S. (2005). Fostering adolescent literacy. *Educational Leadership*, 62(7), 85-87.
- Davenport, P., & Anderson, G. (2002). *Closing the achievement gap*. Houston: American Productivity and Quality Center.
- Easton, L. B. (2008). From professional development of professional learning. *Phi Delta Kappan*, 89(10), 755-761.
- Elwood, J. (2006). Formative assessment: Possibilities, boundaries and limitations. *Assessment in Education*, 13(2), 215-232.
- Firestone, W. A. (2009). Accountability nudges districts into changes in culture. *Phi Delta Kappan*, 90(9), 670-676.
- Fisher, D., & Frey, N. (2007). *Checking for understanding: Formative assessment techniques for your classroom*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Florida Department of Education. (2006). *Florida professional development evaluation protocol*. Retrieved April, 2010 from <http://www.fldoe.org/profdev/pdf/1protocoloverview.pdf>
- Florida Department of Education. (2007). *Assessment and accountability briefing book*. Retrieved July 2009, from <http://fcats.fldoe.org/pdf/BriefingBook07web.pdf>
- Florida Department of Education. (2009). *FCAT 2009 results*. Retrieved October 7, 2009, from <http://www.fldoe.org/educators>

- Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring measures in mathematics: A review of the literature. *The Journal of Special Education, 41*(2), 121-139.
- Fuchs, L. S., & Fuchs, D. (1999). Monitoring student progress toward the development of reading competence: A review of three forms of classroom-based assessment. *School Psychology Review, 28*(4), 659-671.
- Fuch, L.S., Fuchs, D., & Zumeta, R. O. (2008). A curricular-sampling approach to progress monitoring: Mathematics concepts and applications. *Assessment for Effective Intervention, 33*(4), 225-233.
- Fullan, M. (2001). *Leading in a culture of change*. San Francisco: Jossey-Bass.
- Fullan, M. (2006). *Turnaround leadership*. San Francisco: Jossey-Bass.
- Fuller, B., Gesicki, K., Kang, E., & Wright, J. (2006). *Is the No Child Left Behind Act working?: The reliability of how states track achievement* (Working Paper 06-1). Berkeley, CA: Policy Analysis for California Education, University of Berkeley (ERIC Document Reproduction Service No. ED492024)
- Fuller, B., Wright, J., Gesicki, K., & Kang, E. (2007). Gauging growth: How to judge No Child Left Behind? *Educational Researcher, 36*(5), 268-278.
- Gaught, W. L. (2008). [Marion County reading and mathematics benchmark assessments: Analysis for six elementary schools in restructuring, 2008]. Unpublished raw data.
- Gaught, W. L. (2009). Personal Interview.
- Goddard, R. D. (2003). Relational networks, social trust, and norms: A social capital perspective on students' chances of academic success. *Educational Evaluation and Policy Analysis, 25*(1), 59-74.
- Gordon, G. (2006). *Building engaged schools: Getting the most out of America's classrooms*. New York: Gallop Press.
- Greene, D. L. (2008). Personal Interview.
- Guilfoyle, C. (2006). NCLB: Is there life beyond testing? *Educational Leadership, 64*(3), 8-13.
- Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practice, 28*(3), 24-31.

- Hintze, J. M., Christ, T. J., & Methe, S. A. (2006). Curriculum-based assessment. *Psychology in the Schools, 43*(1), 45-56.
- International Reading Association and the National Council of Teachers of English, (2010). *Standards for the assessment of reading and writing, Newark, Delaware*, (NO. 776 & 46864). The Joint Task Force on Assessment, Author.
- Irvin, J. L., Meltzer, J., & Dukes, M. (2007). *Taking action on adolescent literacy: An implementation guide for school leaders*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Kelly, R. R., & Gaustad, M. G. (2007). Deaf college students' mathematical skills relative to morphological knowledge, reading level and language proficiency. *Journal of Deaf Studies and Deaf Education, 27*(1), 25-37.
- Lalley, J. P., & Gentile, J. R. (2009). Classroom assessment and grading to assure mastery. *Theory Into Practice, 48*(1) 28-35.
- Leahy, S., Lyon, C., Thompson, M., & William, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership, 63*(3), 18-24.
- Lezotte, L.W. (1992). *Creating the total quality effective school*. Okemos, Michigan: Effective Schools Products, Ltd.
- Linn, R. L. (1994). Performance assessment: Policy promises and technical measurement standards. *Educational Researcher, 23*(9), 4-14.
- Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher, 32*(7), 3-13.
- Lutz-Doeming, C.K. (2007). *An examination of the Pennsylvania 4Sight Benchmark Assessments as predictors of Pennsylvania System of School Assessment performance*. Unpublished doctoral dissertation, Lehigh University.
- Marion County Public Schools. (2007). [District benchmark progress monitoring assessment]. Unpublished assessment.
- Marzano, R. J. (2004). Building background knowledge for academic achievement: Research on what works in schools. Alexandria, VA: Association for Supervision and Curriculum Development.
- Marzano, R. J. (2006). *Classroom assessment and grading that work*. Alexandria, VA: Association for Supervision and Curriculum Development.

- Marzano, R. J., Waters, T., & McNulty, B. A. (2005). *School leadership that works: From research to results*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Moss, C. M., & Brookhart, S. M. (2009). *Advancing formative assessment in every classroom: A guide for instructional leaders*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Moxley, D. E., & Taylor, R. T. (2006). *Literacy Coaching: A handbook for school leaders*. Thousand Oaks, CA: Corwin Press.
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education, 14*(2), 149-170.
- No Child Left Behind Act of 2001, ESEA, 2001, Title I, Part B, Subpart 1, Section 1202(c)(7)9A)(IV)2).
- National Staff Development Council, 2001, retrieved October, 2009 from <http://www.NSDC.org>
- Nichols, P. D., Meyers, J. L., & Burling, K. S. (2009). A framework for evaluating and planning assessments intended to improve student achievement. *Educational Measurement: Issues and Practice, 28*(3), 14-28.
- Pearson Education Inc. (2008). Pearson Benchmark [Computer software]. Saddle River, NJ: Pearson Education.
- Perie, M., Scott, M., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice, 28*(3), 5-13.
- Peterson, K. D., & Deal, T. E. (2002). *The shaping school culture fieldbook*. San Francisco: Jossey-Bass.
- Popham, W. J. (2003). *Test better, teach better: The instructional role of assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Popham, W. J. (2006). Phony formative assessments: Buyer beware! *Educational Leadership, 64*(3), 86-87.
- Popham, W.J. (2008). An unintentional deception. *Educational Leadership, 66*(2), 80-81.
- Priestley, M., & Sime, D. (2005). Formative assessment for all: A whole-school approach to pedagogic change. *The Curriculum Journal, 16*(4), 475-492.

- Quick, E., Holtzman, D. J., & Chaney, K. R. (2009). Professional development and instructional practice: Conceptions and evidence of effectiveness. *Journal of Education for Students Placed at Risk*, 14, 45-71.
- Reeves, D. B. (2004). *Making standards work: How to implement standards-based assessments in the classroom, school, and district*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Schein, E. H. (1986). What you need to know about organizational culture. *Training and Development Journal*, 40(1), 30-33.
- Schlechty, P. C. (1997). *Inventing better schools: An action plan for educational reform*. San Francisco: Jossey-Bass.
- Schmoker, M. (1999). *Results: The key to continuous school improvement*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Schmoker, M. (2006). *Results now*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Senge, P. M. (2006). *The fifth discipline: The art and practice of the learning organization*. New York: Doubleday.
- Shepard, L. A. (2009). Commentary: Evaluating the validity of formative and interim assessment. *Educational Measurement*, 28(3), 32-37.
- Smith, M. S. (2005). Improving student achievement by measuring ability, not content. *T.H.E. Journal*, 32(9), 2-3.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment FOR learning. *Phi Delta Kappan*, 83(10), 758-765.
- Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: The Free Press.
- St. John, S. S. (2009). *Leadership styles and student achievement*. Unpublished doctoral dissertation, Walden University.
- Taylor, R., & Collins, V. D. (2003). *Literacy leadership for grades 5-12*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Thompson, M. (2008). *Leadership, balanced achievement, and accountability: Benchmarking to exemplary practice*. Boone, NC: Learning-Focused.

- Underwood, M.K. (2008). [An analysis of reading progress monitoring assessments in low performing schools: Examining the impact of progress monitoring on FCAT achievement]. Unpublished research
- United States Department of Education. (n.d.). *Guidance on standards, assessments, and accountability*. Retrieved June 23, 2009, from <http://www.ed.gov/print/policy/elsec/guid/standardsassessment/guidance.html>
- United States Department of Education. (2009). *The condition of education 2009* (NCES Publication No. 2009-081). Jessup, MD: Author.
- Visone, J. D. (2009). The validity of standardized testing in science. *American Secondary Education*, 38(1), 46-61.
- Wayman, M. M., Wallace, T., Wiley, H. I., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education*, 41(2). 85-120.
- Weaver, R. (2006). A positive agenda for ESEA. *Educational Leadership*, 64(3), 32-36.
- Whittaker, T. (2003). *What great principals do differently: Fifteen things that matter most*. Columbus, OH: Click Publishing Services.
- Wiggins, G., & McTighe, J. (2005). *Understanding by Design*. Alexandria, VA: Association for Supervision and Curriculum Development.
- William, D. (2006). Formative Assessment: Getting the focus right. *Educational Assessment*, 11(3, 4), 283-289.