2016

Comparing Fountas and Pinnell's Reading Levels to Reading Scores on the Criterion Referenced Competency Test

Shunda F. Walker
Walden University

Walden University

COLLEGE OF EDUCATION

This is to certify that the doctoral study by

Shunda Walker

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Jose Otaola, Committee Chairperson, Education Faculty
Dr. Edward Williams, Committee Member, Education Faculty
Dr. Lisa Reason, University Reviewer, Education Faculty

Chief Academic Officer

Eric Riedel, Ph.D.

Walden University
2016

Abstract

Comparing Fountas and Pinnell's Reading Levels to Reading Scores on the Criterion

Referenced Competency Test

by

Shunda Walker

MS, Walden University, 2005

BS, Troy State University, 2002

Doctoral Study Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Education

Walden University

February 2016

Abstract

Reading competency is related to individuals' success at school and in their careers. Students who experience significant problems with reading may be at risk of long-term academic and social problems. High-quality measures that determine student progress toward curricular goals are needed for early identification and interventions to improve reading abilities and ultimately prevent subsequent failure in reading. The purpose of this quantitative nonexperimental ex post facto research study was to determine whether a correlation existed amongst student achievement scores on the Fountas and Pinnell Reading Benchmark Assessment and reading comprehension scores on the Criterion Reference Competency Test (CRCT). The item response theory served as the conceptual framework for examining whether a relationship exists between Fountas and Pinnell Benchmark Instructional Reading Levels and the reading comprehension scores on the CRCT of students in Grades 3, 4, and 5 in the year 2013-2014. Archival data for 329 students in Grades 3-5 were collected and analyzed through Spearman's rank-order correlation. The results showed positive relationships between the scores. The findings promote positive social change by supporting the use of benchmark assessment data to identify at-risk reading students early.

Comparing Fountas and Pinnell's Reading Levels to Reading Scores on the Criterion

Referenced Competency Test

by

Shunda Walker

MS, Walden University, 2005

BS, Troy State University, 2002

Doctoral Study Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Education

Walden University

February 2016

Dedication

My dissertation and the many hours put into it are dedicated to my wonderful

husband, DeAnte Walker, my two beautiful daughters, Taja and Jyra, my parents, Louis

and Mary Foster, and more importantly to my most gracious heavenly father who guided

me every step of the way. My husband's love, strength, patience, and never-ending

encouragement helped me reach my long time goal, and for this, I will be forever

thankful. Taja and Jyra, this dissertation represents a major investment of time and

energy that should have been spent with you over the past seven years of your life. Know

that I love you more than words can describe, and that I did this for you. I want to express

my deepest appreciation to my parents for the many sacrifices made toward my education

and instilling its value within me.

Acknowledgments

There are many people that have supported me on this journey. Their support, guidance, and words of encouragements made the completion of this project possible. I would like to begin by thanking my heavenly father. He left his word as a source of inspiration through dark hours. The scriptures kept me humbled, focused, and inspired to persevere throughout this great work. I could not have completed one page without him. I continue to pray that he will help me use this great work to continue to bring honor to him.

I am also especially thankful for my family, especially my husband and two daughters. DeAnte, Taja and Jyra you believed in me and supported me throughout my graduate training. You pushed me to greater heights, just by believing in me. Your expectations, words of praise, hugs, and inspirational vitamins were appreciated. I would also like to acknowledge the efforts of my mom and dad. Mom and dad to you I will forever be grateful. You have followed me all the way. You gave up so much for me to reach my dreams and not once have you asked for anything in return. Your prayers, long conversations, friendship, and love have carried me for the past seven years.

I appreciate all of the individual and collective contributions and support from my committee members and co-chairs. Dr. Otaola and Dr. Williams provided me with wise and insightful feedback that helped me frame this work. Their critical eyes on my dissertation provided me with opportunities to become a better writer.

Table of Contents

ii

iii

List of Figures

Section 1: Introduction to the Study

The ability to read and comprehend is vital to students' success in school, as well as to their achievements beyond school life. Reading, which is the gateway to all other academic achievement, is the most important skill learned by children (Jordan, Snow, and Porsche, 2000). Despite this critical importance, many children in the United States are failing to learn to read (National Center for Education Statistics, 2013).

Educational accountability has become a popular word in education (Madaus and Russell, 2010). Districts, schools, educators, and students are held responsible for student performance. There has become an expectation for schools to make certain that students pass the state test (Madaus and Russell, 2010). Ravitch (2010) stated that assessments have become the crux on which the destiny of students and the notoriety and futures of their teachers, principals, and schools lies. Formative and summative assessments play different roles in education accountability and remain a concern for school districts (Black & Wiliam, 2009). Summative high-stakes assessments influence administrative decisions and are used to judge the quality of instruction in schools (Black & Wiliam, 2009). On the contrary, formative assessments are used to improve instructional practices and focus on the learning needs of students (Black & Wiliam, 2009).

Students come to school with factors that may present challenges to teachers. Effective teachers believe that they can teach all students to read and that these factors are not excuses (Begeny, Krouse, Ross, and Mitchell, 2009). A determined attitude alone is not enough for those teachers to be successful. Teachers need effective teaching

techniques and curriculum that have a proven track record of success to meet the challenge of teaching students to read (Powell & Kalina, 2009). As students' progress through grade levels in school struggling to read, the appropriate interventions that provide instruction to close the widening gap between their grade level peers becomes ever more challenging (Roberts, Torgesen, Boardman, & Scammacca, 2008). Reading skills that are critical for students' success can improve with explicit small group instruction (Begeny, Krouse, Ross, and Mitchell, 2009).

This study was undertaken to determine the relationship between the Fountas and Pinnell (F and P) Benchmark Assessment Instructional Reading Levels and the reading comprehension scores on the Criterion Reference Competency Test (CRCT) in grades three, four, and five in the year 2013-2014. This first section provided the background, states the problem, describes its significance, and outlines the theoretical framework used in the study. The section concluded with a discussion of the definitions of terms utilized in the study, limitations, and delimitations of the study.

Problem Statement

Educators need a method to determine if a correlation exists between students F and P instructional reading levels and reading CRCT scores. For several years, the Fountas and Pinnell Reading Benchmark Assessment has been the primary benchmark assessment used in the school system being studied (Henry District Public School System, 2013). There is no current data showing the degree of correlation between the Fountas and Pinnell instructional levels and the reading CRCT scores. Since the CRCT is

a high-stakes test used to measure the academic progress of schools, it is necessary that teachers utilize reliable tools that will ensure that students are making progress in reading and prepared for the state test (Brown-Chidsey & Steege, 2010). Students at risk for academic failure as a result, of reading problems must be identified early enough for appropriate interventions to be provided (Brown-Chidsey & Steege, 2010). Additionally, the district in this study desires students to meet or exceed on the reading part of the CRCT to demonstrate compliance in meeting the requirements of the NCLB legislation (Henry District Public School System, 2013). Currently, the Fountas and Pinnell is the only reading benchmark assessment used to measure students reading performance before the state test (Henry District Public School System, 2013). However, the extent to which the Fountas and Pinnell Reading Benchmark instructional levels correlated with reading scores on the Georgia CRCT was unknown. Educators needed benchmark data that would enable them to make educated decisions about classroom practices and assessment uses. Furthermore, educators needed to determine whether a statistically significant correlation existed amongst student achievement scores on the Fountas and Pinnell Reading Benchmark Assessment and reading comprehension scores on the Criterion Reference Competency Test (CRCT). Specific explanations concerning the expectations and measurements of the assessments will be discussed in Section 2.

Reading is a necessity in life. It is a way for people to function and participate in the world around them. Learning to read can be a struggle but if teachers are aware of their learners interventions can be set up for students to succeed. There are always

children who find it difficult to learn to read (Fountas & Pinnell, 2012). Children who have difficulty in reading often face the struggle throughout their lives (Fountas & Pinnell, 2012), thus, motivating educational professionals to find strategies and interventions to assist with this challenging problem. Fountas and Pinnell (2012) contend that teachers are responsible for raising all readers to the level of achievement they need to function and succeed in daily life. Because there are students who struggle with reading and fall further behind each year, schools need to become proactive in the intervention movement.

Nature of the Study

Nonexperimental quantitative ex post facto design was selected to determine if a correlation existed between student Fountas and Pinnell Instructional Levels and student achievement scores on the CRCT. The nonexperimental, quantitative, correlational design was chosen based on Creswell's (2013) guidelines. Creswell's guidelines indicate that a correlational design is selected when researchers want to relate similarities between two variables (Creswell, 2013). The Fountas and Pinnell Instructional Levels were treated as the predictor variables and CRCT test scores were treated as the outcome (criterion) variables. This design utilized descriptive and inferential statistics and evaluation research to describe the population. Determining if a predictive relationship existed between the Fountas and Pinnell Benchmark assessment levels and the CRCT was the focus of this study.

Research Questions

Based on the extant literature that presented in Section 2, research questions were developed to help guide this study. This question provides the foundation for understanding the predictive validity of the Fountas and Pinnell Benchmark assessment. Because the CRCT is a high-stakes test that is used to measure the academic progress of schools, it is critical that educators make use of the most appropriate assessment tools to ensure students are prepared for state assessments. The following questions will be explored in this study:

RQ1: What is the relationship between the third grade spring Fountas and Pinnell Reading Instructional Levels and the third grade CRCT Reading Comprehension Scores?

RQ2: What is the relationship between the fourth grade spring Fountas and Pinnell Reading Instructional Levels and the fourth grade CRCT Reading Comprehension Scores?

RQ3: What is the relationship between the fifth grade spring Fountas and Pinnell Reading Instructional Levels and the fifth grade CRCT Reading Comprehension Scores? The hypotheses are as follows:

H_0: There is no relationship between the spring Fountas and Pinnell Benchmark Assessment Instructional Reading Levels and the reading comprehension scores on the Criterion Reference Competency Test (CRCT) in grades three, four, and five in the year 2013-2014.

H$_a$: There is a relationship between the spring Fountas and Pinnell Benchmark

Assessment Instructional Reading Levels and the reading comprehension scores on

the Criterion Reference Competency Test (CRCT) in grades three, four, and five in

the year 2013-2014.

Purpose of the Study

The purpose of this nonexperimental, ex post facto, quantitative research study

was to determine whether a correlation exists amongst student achievement scores on the

Fountas and Pinnell Reading Benchmark Assessment and reading comprehension scores

on the Criterion Reference Competency Test (CRCT). Finding a relationship between

these two scores helped determine whether data gathered from district-wide administered

Fountas and Pinnell Benchmark Assessments could serve as a tool for predicting

students' performance on the reading part of the CRCT, or other state tests. It may serve

educators to be able to decide if this time consuming assessment is worth the invested

time (Popham, 2009). The assessment may also serve as an accurate indicator of

performance on standards-based state assessments. Research studies indicate that

formative and benchmark assessments have the potential to be used to predict student

performance on state tests (Merino & Beckman, 2010). Early identification of students at

risk of failure is necessary to target and implement interventions to improve student

performance.

There is a renewed effort by a school district in Georgia to use benchmark

assessments to measure student progress; stakeholders are interested in the relationship

between the Fountas and Pinnell Instructional Levels and the reading CRCT scores. Benchmark assessments are becoming popular tools to assist educators in screening, making instructional decisions, and monitoring student progress (Black & Wiliam, 2009). Christ et al. (2010) demonstrated that teachers who utilize formative assessment data are apt to rely more heavily on data to guide instruction. Benchmark assessments by educators can also be connected to the extensive amount of research that link student learning to formative assessments (Black & Wiliam, 2009). Unfortunately, there are a limited number of studies that compare reading benchmark assessment scores and high-stakes reading scores. This study represents a step toward understanding and validating the potential use of the Fountas and Pinnell Assessment.

The study's results are important, considering the impact assessments have on all stakeholders. The findings add to the literature and knowledge of instructional practice by validating the relationship between Fountas and Pinnell Instructional Levels and reading scores on the CRCT. These findings, as a result, allow school districts to access information that can help implement change in their curriculum, schools, and classrooms.

Theoretical Framework

The Item Response Theory (IRT) provided a framework for this study. Item Response Theory, also known as the latent trait theory, is a common statistical theory about how performance relates to the competence that is measured by the items on the test (Hambleton, Swaminathan, & Rogers, 1991). This study will examine an IRT linking method to link Fountas and Pinnell instructional levels and CRCT reading scores.

Linking is generally used to describe the comparison of results from two or more separate assessments (Embretson, 1996). The IRT provides ways to approach test linking. This approach will be used to find the instructional levels on the Fountas and Pinnell test that corresponds to the summed scores on the various reading CRCT domains. This theory also attempts to explain a person's response to test items (Lord, 1984). IRT implies that if a student has a high ability in reading, he or she will probably get easy items correct (Baker, 2001). Conversely, if a student has a low reading ability and the items are difficult, he or she will potentially get the items wrong (Baker, 2001). In this regard, the IRT serves as a guide for assessing the academic variables, such as Fountas and Pinnell instructional levels and CRCT reading scores, to determine if a relationship exists.

Due to latent traits' unobservable nature, researchers have to measure them in the form of a test or survey with the assumption that the latent traits would influence a person's response to items (Hambleton, Swaminathan, & Rogers, 1991). The beliefs of IRT are established on two fundamental assumptions. First, a person with higher ability should have a greater probability of success on assessment items than a person with lower ability (Hambleton, Swaminathan, & Rogers, 1991). Secondly, any person should always be more likely to do better on an easier item than on a more difficult one (Hambleton, Swaminathan, & Rogers, 1991). In the case of this study, students with higher reading abilities should have a greater probability of success on the Fountas and Pinnell and CRCT assessments. IRT also assumes that every examinee has some true location on the continuum of the specific latent trait (Hambleton, Swaminathan, & Rogers, 1991). It is

this location that influences the examinees response to any item(s) on a test or survey (Hambleton, Swaminathan, & Rogers, 1991). Despite the fact that measurement tools may vary for the same latent trait, the examinees location on the continuum should be consistent across test formats (Hambleton, Swaminathan, & Rogers, 1991). Thus, it is assumed that the examinee achievement on the Fountas and Pinnell can be linked to achievement on the CRCT by examining reading abilities.

In test linking studies, such as this one, the critical question is: Can different tests, which measure the same construct, yield comparable scores? The IRT suggests that alternate forms are balanced in terms of equivalent test information functions (TIF) (Embretson, 1996). IRT is noted as having a built in linking mechanism (Embretson, 1996) which supports the notion that through IRT two sets of summed scores from two different tests can be easily calibrated using linking items. In the current study, an IRT model will be used to place the Fountas and Pinnell and CRCT assessment results on a common scale. This will allow the researcher to link the scores of different reading assessments and determine if a relationship exists.

In many educational assessment situations, there is an underlying variable of interest (Baker, 2001). In this study, the variable of interest is reading ability. Reading ability is what psychometricians refer to as an unobservable, or latent, trait (Baker, 2001). Although the attributes of reading ability can be listed, it cannot be measured directly. Reading ability is a concept rather than a physical dimension (Baker, 2001). The Fountas and Pinnell and CRCT assessments were developed and administered to determine how

much reading ability students possessed. The items on each test measures some aspect of reading ability. Under item response theory, the linking task was used to determine corresponding Fountas and Pinnell instructional reading levels and CRCT reading scale scores. This study looks to determine if a relationship exists between the Fountas and Pinnell instructional levels and reading CRCT scale scores.

In the context of this research study, the IRT offers mathematical functions that characterize the relationship between a students' reading ability as being measured by two instruments and offers an alternative to traditional linking methods (Embretson, 1996). Utilizing this framework, estimates of reading abilities are balanced across the subsets of items that make up different test forms. According to Hambleton (1989), test takers taking two different test forms can be plotted on a single common scale of ability. In this quantitative correlational study, students' scores on the Fountas and Pinnell and Reading CRCT are plotted on a single, common scale to determine if a relationship exists.

The development of standardized assessments was to provide educators with data that identifies students' strengths and weaknesses in different subject areas (Schelppenbach, 2010). The Fountas and Pinnell Reading Benchmark assessment and the CRCT are two such assessments that have this goal (Henry District Public School System, 2013). Both tests were created to measure students' progress so teachers can assure students' learning needs are addressed (U.S. Department of Education, 2010). Item Response Theory models are used to estimate the abilities of students taking both tests.

Standardized assessments are expected to provide diagnostic information to assist students and educators in identifying strengths and weaknesses in learning (Schelppenbach, 2010). These tests measure student achievement by collecting and analyzing information about student learning so that teachers and students can interpret and utilize the information to improve student learning (Schelppenbach, 2010).

The Common Core State Standards consist of 10 reading anchor standards (Georgia Department of Education, 2009). The reading standards are divided between reading for literature standards and reading for informational standards (Georgia Department of Education, 2009). Anchor reading standards focus on key ideas and details, craft and structure, integration of knowledge and ideas, range of reading and text complexity, and response to literature (Georgia Department of Education, 2009). . Both the CRCT and Fountas and Pinnell assessments measure how well students are able to closely read and analyze a text (key ideas and details), students attention to the language and structure of a text (craft and structure), critical thinking skills (integration of knowledge and skills), and their ability to read complex literary and informational text (Georgia Department of Education, 2009) . Data collected from both assessments allows teachers and other stakeholders to make informed academic decisions. Item Response Theory methods were used to link the Fountas and Pinnell instructional levels to the CRCT reading scores.

In summary, an IRT method was used to link two different assessments. The IRT supports the notion that alternate test forms are balanced in terms of equivalent test

information functions (TIF), and examinee achievement can be predicted from one or

more abilities (Baker, 2001). Additionally, the theory of latent traits suggests that in

testing situations examinee performance on a test can be predicted (Baker, 2001). IRT

specifies the relationship between observable student test performance (scores) and the

unobservable traits (reading ability). Educators at the school in this study use students'

past performance levels on the Fountas and Pinnell assessment to predict (or explain)

reading scores on the CRCT (Henry District Public School System, 2013). They also

interpret information about students' reading understanding before the end of the year

state assessment to improve student learning and prevent poor performance from

occurring (Henry District Public School System, 2013).

Operational Definitions

The following terms are defined as used within this study:

Actuation: The process of analyzing data and using result interpretations to

modify instructional practices (Halverson, 2010).

Adequate Yearly Progress (AYP): Adequate Yearly Progress is an annual measure

of student participation and achievement of statewide assessments and other academic

indicators (Georgia Department of Education, 2009).

Assessment: The process of collecting, synthesizing, and making use of

information to assist in decision-making (Airasian, 1994).

Common Core Georgia Performance Standards (CCGPS): A set of core

standards in English language arts, mathematics in kindergarten through high school, and

literacy in science, history/social studies, and technical subjects in grades six to

12(Georgia Department of Education, 2009).

Criterion reference competency test (CRCT): An assessment designed to measure

how well students acquire the skills and knowledge described in the Georgia Performance

Standards (Georgia Department of Education, 2009).

Formative assessment: Formal and informal assessments that are used during the

learning process to measure a student's progress toward meeting the learning standards

and to improve a students' growth toward mastery (Georgia Department of Education,

2009).

Fountas and Pinnell Benchmark Assessment System (F&P): "A series of texts that

can be used to identify a student's current reading level and progress along a gradient of

text levels over time" (Fountas & Pinnell, 2010, p.1).

Interim/benchmark assessment: A standardized medium cycle assessment

administered three or four times a year. The assessment is used to evaluate students'

knowledge and skills as they relate to predetermined curricular outcomes on summative

assessments and to inform instructional decision at the classroom, school and district

level (Li, Marion, Perie, & Gong, 2010;Marshall, 2008; Popham, 2008).

*Instructional Reading Level***:** The instructional reading level is made up of 26

levels, arranged from A to Z Level A is the easiest level to read and level Z is the hardest

level. To determine levels A to K, you must find the level at which a student reads the

text with 90 to 94 % accuracy and has excellent or satisfactory comprehension. If they

have limited comprehension, they must have 95% accuracy. To determine level L to Z, you must find the level at which a student reads the text with 95% to 97% accuracy and excellent or satisfactory comprehension. If they have limited comprehension, they must score 98% or higher (Fountas & Pinnell, 2010).

Predictive validity: The utility and accuracy of a screen for predicting performance on a future outcome measure (Roehrig et al., 2007).

Response to intervention (RTI): An assessment and intervention model that integrates high-quality teaching and assessment methods using systematic data-based activities (Brown-Chidsey & Steege, 2010).

Screener: A brief measure of skills designed to predict future academic performance (Jenkins, Hudson, & Johnson, 2007).

Summative assessment: Summative assessment occurs at the end of a course or unit of study to provide a single measure of achievement at that point (Georgia Department of Education, 2009).

Limitations and Assumptions

This study was conducted with the assumption that students tried their best when taking both assessments. That is, students did not intentionally perform poorly on the benchmark assessment or standardized assessments for reasons other than not mastering the content. It was also assumed that teachers will utilize equally effective teaching strategies while implementing the Common Core Georgia Performance Standards

(CCGPS). Additionally, it was assumed that the targeted school has followed state guidelines for testing and information recorded in database is accurate.

This study was limited to third, fourth, and fifth grade students who were administered both the Fountas and Pinnell Benchmark Assessments and CRCT test during the 2013-2014 school year. As a result, the findings are general and applicable to third through fifth grade students with similar demographic characteristics, achievement levels, and learning abilities. Another limitation is the design of the study. There is no guarantee that the generalized results of the convenience sample apply to other groups given the nonexperimental ex post facto design. This study is also limited to an elementary school in a rural district of Atlanta, Georgia. This study is not applicable to other locations with different demographic and academic characteristics. Generalizations of results to other states should be cautioned as the composition of the reading tests might vary.

Scope and Delimitations

This study is delimited to test scores from a reading assessment, one benchmark assessment, a population of third through fifth grade students in a rural district of Atlanta, and the specific time frame of study. The data to test the single hypothesis was obtained from students that were in third, fourth, and fifth grade in the 2013 school year. I worked with one benchmark assessment, the Fountas and Pinnell Benchmark assessment. With a limited time frame to conduct the study, only data from one school was used.

Because this study looked to determine if a correlation existed between student Fountas and Pinnell Instructional Levels and student achievement scores on the CRCT, the study included students who had completed both tests. Therefore, this study was delimited to 2012-2013 third, fourth, and fifth graders who had completed the Fountas and Pinnell and the CRCT. Finding a correlation between Fountas and Pinnell Instructional Reading level and reading comprehension CRCT scores was the scope of this study.

Significance of the Study

Leaders in education face the difficult task of educating all children and providing teachers with the most current scientific-based interventions. Improving student learning is the ultimate goal of the nation's education system (Bambrick-Santoyo, 2010). The significance of this research study was to look for a relationship between the instructional reading levels and reading CRCT comprehension scores. There was an urgent need for educators to know if benchmark assessment scores, such as Fountas and Pinnell scores, were comparable to state test scores, such as CRCT scores, and could reliably forecast students' performance on state standardized assessments.

Parents, legislators, and educators are seeking assessments that can accurately measure student progress and predict the achievement of students on high stake standardized tests. Adopting new programs that focus on measuring students' performance can be a difficult task for administrators and teachers. Schools located in rural South Georgia are in need of assessments that offer prompt and correct information.

The Fountas and Pinnell assessment and CRCT provide data that can be used to measures the effectiveness of classroom instruction at the school and school district system level. Both assessments also provide diagnostic information to assist students and teachers in identifying learning strengths and weaknesses.

This study is useful as it investigates a benchmark formative assessment that offers prompt and accurate information with the possibility of predicting future success on state assessments. Since this study succeeded in finding a correlation, teachers have an instrument that can help them to know in advance how their students are doing in relation to the final CRCT test. This information will help them maximize student success on standardized tests. Furthermore, the vast amount of money and time invested in assessments is justified by the findings. Since findings indicate that the Fountas and Pinnell Benchmark assessment scores are comparable to the CRCT scores or similar state test scores; the investments for Fountas and Pinnell resources are supported. Additionally, the findings are valued because the results provide information that is helpful to educators when using benchmark data to make decisions related to student deficiencies. The results are critical in data-driven school settings where teachers are expected to use achievement data to adjust instruction early to increase the potential for student success on summative assessments. This study also adds new knowledge to the field of education in the area of student assessment, and determines whether the Fountas and Pinnell assessment is worth the money and time required to implement it. It is known how or to what extent the data from the local district-wide benchmark assessments

inform educators in individual schools about preparing students to succeed on the high-stakes summative assessment in reading.

The No Child Left Behind Act (NCLB) (2001) mandates that all students obtain grade level performance on the CRCT by the year 2014. Consequently, school districts must investigate all factors that could impede them from reaching the goals outlined in NCLB. Furthermore, the percentages of students not meeting the state standards on the third and fifth-grade Georgia CRCT in reading can also hinder school systems from gaining the maximum amount of points on the College and Career Readiness Performance Index (CCRPI) (Barge, 2012). Because the aforementioned issues are critical to the success of the school and the academic progress of the student, the relationship of reading on academic achievement is an important topic of interest to parents, teachers, school officials, students, and community members.

Summary

The rationale behind the development of this research study is to support schools and educators in determining if a relationship exists between the instructional reading levels and reading CRCT comprehension scores. This section included information that provided an understanding of the examined problem to. The purpose of this nonexperimental quantitative research study was to examine whether a statistically significant correlation exists amongst student achievement scores on the Fountas and Pinnell Reading Benchmark Assessment and reading comprehension scores on the

CRCT. The theoretical perspective, significance of the study, research question, definition of important terms, and delimitations were also presented in Section 1.

Further investigation will be done to explore literature and research on the importance of utilizing benchmark assessment data to maximize student performance on standardized test. Section 2 contains a review of the literature that supports this topic. The literature in Section 2 supports the research purpose, research methodology, and research design by reviewing previous literature on the topic or similar variables. In Section 3, the methodology of the study is described by including the composition of the sample, the instrumentation utilized, and procedures that will be followed. Section 4 will contain the data analysis and findings. Finally, Section 5 will include a summary, conclusions, recommendations, and implications.

Section 2: Literature Review

Reading is an integral function of everyday life. It is a skill essential to academic

success. It is also the responsibility of educators to educate all students appropriately in

an effort to prepare them for college and careers. As a result of federal and state

initiatives (No Child Left Behind [NCLB], 2001), schools and districts are now required

to set high standards for student growth and achievement (Henry District Public School

System, 2013). Georgia currently uses the CRCT to measure statewide achievement

(Henry District Public School System, 2013). It is not best practice to make daily

decisions based on one end of the year assessment (Jenkins, Deno, & Mirkin, 1979) so

benchmark assessments are used in some schools to collect assessment data (Bambrick-

Santoyo, 2010). Assessment data has become a critical element in determining students'

knowledge and progress (Bambrick-Santoyo, 2010). Education systems are using

benchmark assessments to collect valuable information to assist them in making

instructional decisions that improve learning for all students prior to the high-stakes state

test (Bambrick-Santoyo, 2010). Educators need precise information when making

decisions about assessment systems. Providing literature that focuses on the development

of benchmark programs and their predictive nature will help administrators and educators

better understand how well a specific type of benchmark assessment might predict

performance on a high-stakes state test.

The overall focus of this study is to determine if a correlation exists between

students' Fountas and Pinnell instructional levels and the reading CRCT scores. Due to

the stress related to the pressure of meeting the requirements of NCLB, it is imperative that educators know if a relationship exists between the Fountas and Pinnell benchmark scores and CRCT scores. This information may inform administrators and teachers on the connection between instructional data use and improved student achievement. This chapter presents a summary of the literature that frames this investigation. Research indicates that the elements that determine a successful reader from an unsuccessful reader are foundational skill knowledge (Kaminski & Good, 1996). The literature review begins with a discussion of the elements of reading necessary to ensure later success. Next, descriptions of the reading common core standards and different approaches to assessing comprehension are given. Also, in this review there is an overview of the item response theory (IRT), a rationale for data-driven decision making in schools, and literature related to the definition and purpose of high-stake standardized and benchmark reading assessments. The literature related to data-driven decision making can help educators determine if the proper conditions and best practices related to data-driven decision making are being implemented in the study school. The section will conclude with a discussion of predictive value of benchmark assessments, an overview of the Fountas and Pinnell assessment and what it is intended to measure are included.

Educators in this study utilize the Fountas and Pinnell assessment to determine students reading competencies and limitations in an effort to determine their instructional reading levels (Fountas & Pinnell, 2010). Determining accurate instructional levels is a critical aspect of planning appropriate interventions to support students (Fountas &

Pinnell, 2010). The Leveled Literacy Intervention (*LLI)* program provides an opportunity for teachers to give ongoing formative assessments, and collect information that can inform instructional decision making (Fountas & Pinnell, 2010). Assessments as well as early interventions for struggling readers are critical for future academic success.

The databases searched for this literature review include CARET, EBSCOhost, ERIC, GoogleScholar, ProQuest, and Questia. Terms used for research were *assessment, predictive validity, standardized assessment, formative assessments, benchmark assessments, reading assessments, universal screeners, curriculum based measure*, and *summative assessments*.

Reading Elements

The ultimate goal of reading is comprehension (Durkin, 1980). While the elements of phonological awareness, phonemic awareness, fluency, and vocabulary are fundamental to the reading process, "if there is no comprehension, there is no reading" (Durkin, 1980, p.191). The aforementioned skills, which are involved in early reading acquisition and comprehension, prove to be most problematic for students with reading difficulties (Stanovich and Stanovich, 1995). Most forms of personal learning, intellectual growth, and educational attainment depend on the ability to read and extract meaning from text (Durkin, 1980). For this reason, students' success in school has been linked to their ability to comprehend what they read (Durkin, 1980). Reading comprehension is a multifaceted job that involves cognitive skills working together.

Reading comprehension assessments provide an indication of how well all of the reading cognitive skills are working together (Durkin, 1980).

Phonological awareness provides the basis for phonics (Fountas and Pinnell (2010). It is the ability to manipulate either individual or groups of sound (Fountas and Pinnell (2010). Phonemic awareness, a subcomponent of phonological awareness, is the ability to hear and manipulate sounds in words. A student that is phonemically aware is able to isolate (Fountas and Pinnell (2010). Lundberg (1991) indicated that phonemic awareness and literacy achievement are very reliant upon one another. Fountas and Pinnell (2010) defined phonemic awareness as the ability to hear, recognize, distinguish, and manipulate individual sounds or phonemes in spoken words. Furthermore, Fountas and Pinnell (2010) stated that research indicated that instruction in phonemic awareness is effective in preventing or remediating reading difficulties. Phonemic awareness is an essential skill that all readers should acknowledge. It is the basis of reading and often the deciding factor in students becoming a proficient or struggling reader (Fountas & Pinnell, 2010).

Although fluency is not the only indicator of deep comprehension, it is definitely a strong indicator of comprehension (Fountas & Pinnell, 2010). According to Fountas and Pinnell (2010), reading fluency is defined as decoding quickly without mindful attention, using expression, and voice, and comprehending. Fluency plays a critical role in reading success. It is the bridge between decoding and comprehension—two critical tasks that successful readers must perform (Fountas & Pinnell, 2010). When a student

automatically identifies words, he or she is able to comprehend text more completely (Fountas & Pinnell, 2010). Fluency is compromised when readers fail to decode words automatically. Developing fluency will allow a student to transition from learning to read to reading to learn.

Vocabulary and oral language development are foundations for literacy and comprehension (Clay 2010). Students need to know the meaning of words that they are reading in order to make sense of text. Reading and talking with children play a critical role in developing their vocabulary and oral language development. Research indicates that language development is directly related to a child's environment (Clay 2010; Dewey 2012). Other research studies indicate that a correlation exists between a child's vocabulary and oral language development and literacy (Callaghan & Madelaine, 2012; Hill & Launder, 2010). Vocabulary and oral language skills play a pivotal role in developing reading.

The ultimate goal of all reading instruction is targeted at helping a reader comprehend text. Comprehension is combining reading strategies together to understand and interpret text. Fountas and Pinnell (2010) identified decoding words, monitoring and self-correcting, gathering, predicting, maintaining fluency, and adjusting reading rate as key strategies for comprehending text effectively. The aforementioned skills need to be performed effectively for a reader to comprehend text. Furthermore, developing reading strategies such as connecting, inferring, summarizing, synthesizing, and critiquing helps students expand the meaning of text and become proficient readers (Fountas & Pinnell,

2010). Comprehending connects the skills and strategies of reading to make meaning for the reader.

Fountas and Pinnell (2010) noted that a proficient reader thinks within the text, beyond the text, and about the text. Thinking within the text involves actions such as determining words, monitoring and self-correcting, searching for and using information, summarizing, maintaining fluency, and adjusting (Fountas & Pinnell, 2010). Thinking beyond the text includes predicting, making connections, synthesizing, and inferring. Thinking about the text includes analyzing and critiquing the text. The crucial aspect for the actions is the reader's ability to initiate the necessary tasks to gain self -control of these complicated behaviors (Fountas & Pinnell, 2010). However, some readers fail to think within the text, beyond the text, and/or about the text. Thus, an important matter in education today is the failure of students to read effectively and comprehend what they read. The failure of children in academic content areas that require reading and comprehending skills can be linked directly to the inability to access and use metacognitive strategies. There are many different kinds of struggling readers. Some at - risk readers have high accuracy, decoding almost every word given to them, and low comprehension, where they are unable to remember and understand what they have previously read; these students are known as word callers(Meisinger, Bradley, Schwanenflugel, Kuhn, & Morris, 2009). Studies have estimated that approximately one third of all struggling readers consist of word callers (Cartwright, 2010). Even though there are risk factors that are genetic and environmentally based, the preponderance of

learning problems occur not within the child but in the inability of the system to find a way to teach the child (Fountas & Pinnell, 2010). If students are to be successful in reading comprehension, a variety of methods must be used to educate the students, including instruction in metacognitive skills. Additionally, early identification of reading problems is critical to prevent the academic struggles associated with reading deficits.

Georgia Common Core ELA Standards

The Georgia Common Core State Standards (CCSS) were developed in an effort to provide a consistent framework to prepare students for success in college and beyond. Georgia, along with 47 other states, developed this set of core standards for students in kindergarten through 12th grade in mathematics and English language arts (CCST Initiatives, 2013). The English Language Arts (ELA) standards were used as a general guideline for ELA instruction. At each grade level, The Georgia ELA Common Core State Standards have defined the tasks that students should understand and be able to accomplish by the end of each grade level. English language arts instruction is broken up into four strands: Reading, Writing, Speaking and Listening, and Language (CCSS Initiative, 2013). Each strand is further broken down into a strand-specific set of anchor standards that are identical across all grades and content areas. The anchor standards are the fundamental skills that students should have in order to be college and career ready. Each grade level consists of 10 reading anchor standards, 10 writing anchor standards, six speaking and listening anchor standards, and six language anchor standards (CCSS Initiative, 2013.

Grade K-5 reading anchor standards are further broken down into the categories of literature, informational texts, and foundational skills. The 10 anchor standards for literature and informational text were developed around four categories (CCSS Initiative, 2013). The four categories are: key ideas and details, craft and structure, integration of knowledge and ideas, and range of reading and level of text complexity (CCSS Initiative, 2013). The Foundational skill anchor standard is also developed around four categories. The four categories are print concepts, phonological awareness, phonics and word recognition, and fluency. Third through fifth grade only focus on phonics and word recognition and fluency.

Similarly, the common core writing standard also has 10 anchor standards that establish the cross disciplinary expectations for writing. The ELA writing standards are divided into four categories: text types and purpose, production and distribution of writing, research to build and present knowledge, and range of writing (CCSS Initiative, 2013). Text types and purpose include the first three writing standards: argument, informative, and narrative writing. The next three writing standards are included under the production and distribution of writing. Research to build and present knowledge encompass the third three standards. The 10th standard grouped under range of writing requires students to write routinely over a period of time. The writing standards focus on students being able to construct valid arguments by supporting a point of view with clear reasons and evidence, write narratives to develop real or imagined experiences, use technology to produce and publish writing, conduct research projects that build

knowledge about topics, and to write routinely over extended time frames. Unlike the reading and writing anchor standards, there are only six speaking and listening anchor standards.

The six speaking and listening anchor standards are broken down into two categories. The two categories are comprehension and collaboration and presentation of knowledge and ideas (CCST Initiative, 2013). Comprehension and collaboration encompass the first three standards and presentation of knowledge and ideas encompass the last three standards. The speaking and listening standards are critical standards that are often ignored. These standards include coming to discussions prepared, building on the ideas of others and expressing their own ideas clearly and persuasively. The students should also be able to report on a topic with appropriate facts, while speaking clearly at an understandable pace. Like the speaking and listening anchor standards, there are six language anchor standards.

Anchor standards in language are separated into the following three groups: conventions of standard English, knowledge of language, and vocabulary acquisition and use. The contents of the language domain are mainly focused around grammar and vocabulary. The first two standards are included under conventions of standard English. There are several elements under this category. The second category, knowledge of language, includes one standard. Vocabulary acquisition includes the last three standards. Students are expected to demonstrate command of the conventions of standard English grammar when speaking or writing, which includes students being able to explain the

functions of nouns, pronouns, adverbs, conjunctions, prepositions, and interjections (CCST Initiative, 2013). When writing, students should also be able to demonstrate command of the conventions of standard English capitalization, punctuation, and spelling. Students will use knowledge of the language and its conventions to choose words, phrases, and punctuation to convey ideas precisely. Lastly, the vocabulary acquisition and use standard focuses on students' ability to determine the meaning of unknown and multiple meaning words and phrases (CCST Initiative, 2013). Although the ELA anchor standards described above are identical across all grades and content areas, each anchor standard has an accompanying grade-specific standard that translates the broader standard into grade-appropriate end-of-year expectations.

My study focuses on reading comprehension so the reading literacy and reading informational standards need a deeper explanation. The skills necessary for reading literature and informational texts are the same. The reading literature and reading informational text share the same 10 anchor standards. Sometimes the anchor standards are the same for each, and sometimes they are slightly different. Sharing the same 10 anchor standards deepens reading skills for students. Students are able to apply the same skills to various texts. In third, fourth, and fifth grade the reading literacy and informational text standards are divided into four categories (CCSS Initiatives, 2013). Key ideas and details is the first college/career readiness standard in the reading strand. There are multiple standards within this strand. Core reading is at the center of this anchor standard (CCST Initiative, 2013). Students are expected to closely read a variety

of complex texts and have rich discussions related to the texts. The third grade standard specifically states that students should ask and answer questions referring explicitly to the text as the basis for answers (CCST Initiative, 2013). Similarly, this standard in the fourth grade focuses on referring to details and examples in text when explaining what the text states (CCST Initiative, 2013). The fifth grade requires students to accurately quote from a text when explaining the meaning of the text (CCST Initiative, 2013). This standard also requires third, fourth, and fifth graders to determine and summarize the theme of a variety of stories (CCST Initiative, 2013). The other element of this study expects students to describe characters in a story. Fourth and fifth graders have to describe characters using more details and compare and contrast strategies.

Craft and structure is the second college/career readiness reading strand. This strand includes several standards. The standards require students to determine the meaning of words and phrases that are used in the grade level texts (CCST Initiative, 2013). They are also expected to distinguish literal from nonliteral language. Students should explain how parts or series of stories fit together or build on one another. Another element included under this strand is comparing and contrasting points of views in stories (CCST Initiative, 2013).

Integration of Knowledge and Ideas is the third college/career readiness strand. Under this strand students have to explain how specific aspects of texts (visual, oral presentation, and multimedia) contribute to what is conveyed in a text. Stories in the same genre, topic, or theme are compared and contrasted. In the final college/career

readiness strand range of reading and level of text complexity is addressed. In all three grade levels students read and comprehend literature and informational text in the grade level above them complexity band (CCST Initiative, 2013). Integration of Knowledge and Ideas and all of the other reading standards are assessed based upon a variety of indicators.

Assessing Comprehension

Effective assessment is a crucial component of early identification. Consequently, evaluation of comprehension at the classroom level should chiefly be used to provide an indicator of how well all of the sub processes of reading are working together. The assessment data should be used to help teachers make decisions that will help meet the reading needs of every student. There is no general agreement on how to best measure comprehension, thus it remains a controversial topic (McKenna & Stahl, 2003). Clay (2011) determined through her research that reading assessments should be child specific, consisting of recording what the student does when processing texts of specified difficulty. The assessment should refer to the student's skill strengths and weaknesses and literacy moves made while processing the text. The results should be compared with a model of similar behaviors used by children who make satisfactory progress in reading.

There are three popular approaches to assessing reading comprehension. Questioning is one of the traditional approaches to measuring reading comprehension (McKenna & Stahl, 2003). Teachers evaluate students' reading comprehension by asking questions at various levels of thinking. Three common levels of questions used to

evaluate reading comprehension are literal questions, inferential questions, and critical questions (McKenna & Stahl, 2003). Literal questions require students to recall specific facts and details that are explicitly stated in a text. Inferential questions require the readers to use prior knowledge and information from the text to make connections (McKenna & Stahl, 2003). Finally, critical thinking questions leave room for discussion and require the readers to form value judgement about the text (McKenna & Stahl, 2003). A second approach to assessing reading comprehension is cloze testing. This testing method systematically deletes certain words from portions of text, and students are asked to replace the missing words. McKenna & Stahl (2003) stated that a student's ability to provide logical word replacement indicates their ability to comprehend the text. Oral retelling can also be used to measure students reading comprehension. Teachers assess how well students comprehend what they read by having students read out loud then tell what happened in the story (McKenna & Stahl, 2003). The retell is typically evaluated using a checklist. Evaluators use general comprehensibility to gauge comprehension (McKenna & Stahl, 2003). Both the Fountas and Pinnell and CRCT measure students' comprehension through questioning.

Data-Driven Decision Making

Collecting and analyzing data to guide educational decisions is part of the accountability process of NCLB. The idea that drives NCLB is that educational leaders and teachers will utilize student data to inform decision making. Educational institutions use data-based decision making as an attempt to analyze data and use the results to

inform decisions and improve instruction. If schools are to rely on data collected from assessing students with instruments such as the Fountas and Pinnell, it is imperative that researchers and educators take the time to closely examine the strengths and possible limitations in regard to decision-making for high-stakes tests such as the CRCT. Research related to the transformation of theory into practice is essential when trying to determine the efficacy of utilizing data to inform instruction and improve student performance (Wilkins & Shin, 2011). This research will provide an understanding of how practitioners translate reading assessment data into action. Research indicates that educators have varying ideas about what data- based decision making is and how it can be used to improve student learning (Jennings 2012; Weiss, 2012). Schildkamp & Kuiper (2010) describe data-based decision making as an organized method that uses data to modify teaching and learning in the classroom. The data-based decision making process uses analyzed data to identify student and school's strengths and weaknesses. Hess and Fullerton (2009) stated that data-driven decision making does not simply require good data; it also requires good decisions.

The Data-Driven Decision Making Process

The data-driven decision making process also describes the multifaceted task of utilizing data to inform educational decisions (Mandinach et al., 2008). Mandinach (2012) labeled this as pedagogical data literacy. In regards to this study, data-driven decision makes reference to the procedures which teachers and educational leaders use to systematically obtain, organize, and analyze reading assessment data to inform

educational decisions (Datnow, 2011). These processes give stakeholders a better understanding of how practitioners evaluate reading benchmark assessments and translate reading assessment data into action. Stakeholders must ensure that benchmark assessment use lines up with the intended purpose of the assessment. Then they can evaluate the impact of the benchmark assessment, and use the assessment results to make instructional decisions that improve student learning.

Data use and data-based decision making have become popular educational phrases and are used in many ways in schools. Educators place tremendous belief in utilizing data to transform education (Wilkins & Shin, 2011). Education systems use the data driven decision making approach to analyze assessment data and use the results to improve and or predict student performance (Wilkins & Shin, 2011). The current school in this study utilizes reading benchmark data to predict performance on the end of the year standardized reading assessment. Continuous school improvement involves more than utilizing data from a single assessment, such as the CRCT. Likewise, districts concerned with students' performance on the end-of-the year state test increasingly administer benchmark assessments (Sawchuk, 2009). Li et al. (2010) suggest that benchmark assessment data is primarily used to identify the instructional needs of students, in an effort to inform classroom instructional decisions. Lai and McNaughton (2009) indicated that, as it relates to data, the concept is that teachers will use achievement data to determine students' zone of proximal development and modify instructional strategies accordingly.

The data –driven decision making process involves collecting, analyzing, and using data to make decisions (Bernhardt, 2009). Several researchers outline the circumstances that would best serve data-driven decision making practices (Simmons, 2012, Flowers & Carpenter, 2009, Hamilton et al., 2009). Understanding these conditions can help educators determine if the proper conditions and best practices, related to data-driven decision making, are being implemented in educational institutions. Simmons (2012) implied that districts could support data-driven decision making by having the following things arranged: provide stakeholders with access to greater quantity and quality of data; regular collaborative planning time for data to be analyzed and discussed; professional development to increase stakeholders data literacy, and strategies that address schools deficiencies. Similarly Flowers & Carpenter (2009) suggested five simple steps for data-decision making in schools. These steps consist of the following: Carefully examine the school's improvement plan, decide how data will be utilized, recognize pertinent data, analyze and discuss data, establish goals and measure progress. Hamilton et al. (2009) present a data use practice guide that provides a framework for using student achievement data to serve as a foundation for instructional decision making. Hamilton et al. (2009) recommend various decisions that can be made using data such as altering lessons, regrouping of student groups, targeting specific students and specific needs, individualizing and responding to the needs of individual students, and using instructional time efficiently. The practice guide summarizes five suggestions for educators. The five suggestions are: make data a continuous cycle of instructional

improvement, create self-directed learners where students evaluate their own data and set goals, develop a distinct vision for data use within the school, provide supports that promote the development of a data-driven culture within schools, establish and preserve a district- wide data system. Despite the fact, that the procedures described in the literature seems to be easy to follow, data-driven decision making can be complicated when it comes to using relevant data to make informed decisions.

Implementing Data-Driven Decision Making in School Systems

The literature described the best conditions for implementing data-driven decision making in school systems. However, the process of converting data into useful information for educators is what will impact student learning. One of the most challenging steps in the data-driven decision making process is turning data into useful information about student performance (Moore, 2011). Many schools appear to be data rich and information poor (Wilhelm, 2011). Sharratt and Fullan (2012) explained that many times educators are overwhelmed by the excessive amounts of information obtained from assessment data. The initial exhaustion of the abundance of data makes it difficult to connect data and instructional changes (Moore, 2011). Limited research is available on how teachers act on data. There is literature that indicates principals' role in assisting educators. It is important to note that principals who lack data literacy face many obstacles when trying to lead in data-driven decision making (Wu, 2009). These are the leaders who are unsure of what the data means and how it should be used. When administrators clearly communicate the expectations of data use in the school, data use

becomes non-threatening to teachers (Deike, 2009). Principals play a key role in easing anxieties and providing ongoing learning opportunities for teachers to analyze, discuss, and make decisions related to data (Ward-Roberts, 2009; Anfara & Donhost, 2010). They must ensure that the professional learning and collaborative planning time is focused on data use (Butler, 2009). Godreau Cimma (2011) conducted a qualitative case study in Connecticut. His study indicated that the principal was responsible for establishing the tone for data-driven decision making in a school. Empowering educators to develop their instructional decisions off of data can lead to continual improvements in students' instruction (Cosner, 2011).

While various sources of data are collected, evaluated, and monitored to support school efforts (Smith, Johnson, & Thompson, 2012), educators still have a difficult time converting data into useful information. Anfara & Donhost (2010) established five stages in the data-driven decision making implementation process that assists educators with converting data into useful information. Stage one is organizing data for success, stage two is developing assessment literacy, stage three involves aligning data systems, and the final stage is modifying instruction (Anfara & Donhost, 2010). Student data, such as Fountas and Pinnell benchmark data, is an essential component of the professional learning experience at the study school. Fountas and Pinnell assessment data is organized for reading success within the school. Teachers are trained on the assessment system. Assessments are given three times a year, teachers meet to discuss data results, and adjust instruction accordingly.

As data become a critical element in improving organizations, school leaders must examine some of the literature that points to data-driven decision making as a key element in improving student achievement. Schools and school districts collect many forms of ongoing data. Schelppenbach (2010) pointed out the different types of data that schools manage during the school year. Achievement data is one of the most crucial forms of data managed (Schelppenbach, 2010). Achievement data is derived from annual state assessments, curriculum based assessments, and benchmark tests that are administered periodically to monitor student progress (Schelppenbach, 2010). Despite the different types of data available, schools typically focus on achievement data as their primary source to make decisions (Schelppenbach, 2010). The types of data collected and used in schools depend on the discoveries leaders are trying to make. Summative assessments are given at the end of the school year, and unable to provide information that allows for immediate instructional changes (Schelppenbach, 2010). Therefore, educational leaders prefer to use local benchmark assessments. Dunn and Mulvenon (2009) reported that benchmark assessments can play many different roles in education. They can be used to diagnose weaknesses, evaluate teaching and learning, and to predict student performance on summative assessments. Research indicates that benchmark assessments provide data on learning and allows teachers to modify instruction appropriately (Cauley & McMillan, 2009). Lim and Roger (2010), and Brundage and Hancock (2010) investigated formative assessments. The Georgia CRCT was used to measure increased academic performance at the end of the school year. The data

suggested that student academic growth could be predicted using every 1 point increase on the formative assessment. This data supports the use of benchmark assessments to predict performance on summative assessments.

Application Of Data-Driven Decision Making On Student Achievement At An Elementary School

Bernhardt (2009) investigated the application of data-driven decision making on student achievement at an elementary school. He noted that, for two years, students in different grade levels and content areas improved. Other studies noted that data-driven decision making was key to success within the school (Levin, Datnow, & Carrier, 2012; Foley & Sigler, 2009). Supovitz & Taylor (2003) conducted a study on a standards-based district reform in Duval District, Florida. They found that data-driven decision making caused improvement in student achievement throughout the four year study. A similar study conducted by Togneri & Anderson (2003) investigated five high poverty school districts that had shown academic improvement across grade levels, and races. After a thorough examination, these researchers found that the common factor amongst the five districts was that they were all engaged in data-driven decision making. Using benchmark assessments, such as the Fountas and Pinnell, has been shown to be reliable for making instructional decisions and has the potential to improve student learning (Davidson & Frohbieter, 2011). While state summative assessments are also able to document student performance, they are unable to provide meaningful results in a timely manner Anfara (2010).

Although most of the literature highlights the areas in which data-driven decision making supports improvements in academic achievement, there is very little literature that connects the process of data use with the outcome of data use. More research is needed in this area. Coburn & Turner (2012) determined that there was little literature that connected data use with data outcomes. He explained that most people just assumed that positive educational outcomes are the results of educators having correct data and procedures for using data. Educational leaders are responsible for preparing students for 21st century learning. Politicians and other stakeholders express the need for students to be college and career ready when they graduate from high school. Popham (2009) reported that instructional leaders are expected to get more instructional mileage out of the assessment data at hand. Data-driven decision making is the key to measuring the skills necessary for today's learners.

Standardized Assessments

Assessments have been a critical component of the American education system for more than two centuries. Although teachers are required to use different types of assessments to measure students' achievement, high-stakes tests ultimately evaluate teacher and student performance. Standardized reading comprehension assessments can be useful in identifying students with poor comprehension. Teachers also indicated that high stakes testing influenced their teaching methods (Au, 2011). Wright (2012) states that the primary purpose of today's state and educational policy makers is to assess and monitor the academic success of at risk students through standard based reforms. Turner

(2009) indicated that American standardized testing began with the growth of the Elementary and Secondary Education Act of 1965 (ESEA). These educational reforms have relied on standardized testing as a method of measuring student achievement. However, with the passage of No Child Left Behind, emphasis has been placed on results. This accountability movement holds teachers responsible for students producing "good test results" (Lashway, 1999). Therefore, an accurate predictor of student reading performance on these high stakes test is necessary.

Assessments available to measure student performance in reading are plentiful, but NCLB has emphasized the standardized assessments (Guilfoyle, 2006). In the 1800s, these assessments were used sparingly to monitor progression through a course of study. Madaus and Russell (2010/2011) noted that assessments during that time were primarily oral. Assessments were used periodically to determine competency and to evaluate the curriculum. According to Madaus and Russell (2010/2011), Mann is mainly responsible for exposing the idea of standardized tests. Mann believed that oral assessments should be replaced with essay type tests. After the invention of scanners, multiple choice assessments became popular. Multiple choice assessments, such as the CRCT, became reputable sources for gathering data in schools primarily due to the cost and efficiency needed to administer these assessments on a large scale (Berns & Sandler, 2009; Colburn, 2009).

In 1983, the National Commission on Excellence in Education released the report A Nation at Risk. This publication called for improved teaching by way of benchmarks,

standards, and high stake's tests. This widely publicized report introduced the standards and high-stakes testing movement (Au, 2009). At the start of the 21st century, Iowa was the only state without a statewide assessment (Au, 2009). District and state assessment data is currently the driving force for education systems (Oberg, 2009). Consequently, every state now uses high-stakes testing as a reputable source for gathering data related to student achievement. (Baker & Johnston, 2010).

Unlike low-stakes standardized tests, such as benchmark tests, high-stakes standardized tests are connected to substantial consequences (Turner, 2009). These assessments are used to judge schools, teachers and students (Berns & Sandler, 2009). As a result of funding associated with student performance, schools are expected to show improvements on high-stakes test (Upadhyay, 2009). Behrent (2009) noted that teachers are forced to aim their attention at preparing students to do well on the end of the year test. The use of high-stakes data, as it relates to accountability, has been prevalent at the elementary level. In 2002, the federal No Child Left Behind law demanded that all students be proficient in reading and math. High-stakes testing became the primary method in which states began to measure student performance (Amrein-Beardsley, 2009; Rothstein, 2009; Viadero, 2009). Standard based high-stakes assessment results are still used annually to measure school performance (Anderson, Leithwood, & Strauss, 2010). Wright (2012) explained that the state of Georgia's revised promotion and retention policy requires a child to demonstrate mastery on the CRCT in order to be promoted to the next grade. This increased accountability evoked great interest in the relationship

between benchmark assessments and state standardized tests. Godwin (2014) asserts that the emphasis on formative data, or benchmark assessments along with summative assessment would prevent distortion of performance data.

Benchmark Assessment Use and Purpose

The prevalence of benchmark assessments has grown significantly over the past few years (Lai and McNaughton, 2009; Shepard, 2010). Benchmark assessments are state aligned tests that are administered to students at multiple intervals throughout the year (Donhost, 2010). Benchmark assessments are now implemented in many urban school districts. Burch (2010) reported that 82% of urban schools expressed that they had implemented some form of benchmark assessment. Li, Marion, Perie, and Gong (2010) states that the primary purpose of benchmark assessments is to identify the instructional needs of students so that teachers can make informed instructional decisions. The goal is to be able to judge events for indications of other events, so that schools can prepare in all cases for the coming of what is anticipated. Reading benchmark assessments provide teachers with ongoing information throughout the school year regarding student progress in reading. Teachers are able to utilize the data to adjust instruction or curriculum to meet student needs. District and schools are counting on reading benchmark assessments data to predict student achievement. Consequently, they are willing to spend a substantial amount of money, time and resources on the implementation and use of benchmark assessments (Burch, 2010). Benchmark assessments can be summarized into three

general classes of purposes. These three purposes can be classified as instructional, evaluative, and predictive.

The instructional purpose of a benchmark assessment is to provide results that enable educators to adjust instruction and curriculum to better meet the needs of students. Instruction is modified based on the results of these assessments while keeping in mind, the goal of helping students reach learning targets. Benchmark assessments are implemented for instructional purpose with the idea that the results can be used similar to the results of formative assessments, to produce increased student achievement (Moss & Brookhart, 2009; Li et al., 2010). Similar to formative assessments, benchmark assessments provide information about a program's effectiveness, accountability, and provide evidence of student progress to inform daily decisions and practices (Chappuis, 2009). The benchmark assessments are administered between classroom formative assessments, and the end of the year standardized test (Bulkley, Olah, & Blanc, 2010). These assessments are used to connect formative classroom assessments to statewide summative assessments.

Time period of Benchmark Assessments

Benchmark assessments, such as the Fountas and Pinnell Assessments, are used periodically through the year to assess students' strengths and weaknesses. Since there is no uniform consensus as to when they are administered, it is assumed that they are administered every few months (Popham, 2011). They inform students and teachers on what students know and on what they still need to learn (Torgesen & Miller, 2009). Data

derived from benchmark assessments, such as the Fountas and Pinnell assessment, assist educators in monitoring students' progress toward state standards. Assessment data also provides the means for educators to tailor instruction so that the standards that are being taught can be assessed (Nichols, Meyers, & Burling, 2009). Teachers use these assessments to monitor student's progress and inform instruction with the goal of improving student learning. Chappuis, Chappuis, & Stiggins (2009), noted five criteria that assessments must meet in order to inform effective decision making. Assessments must have a clear purpose, clear learning targets, sound assessment designs, effective communication of results, and student involvement in the assessment process.

Finally, if the instructional purposes are geared at providing feedback to students about their learning, assignments should be created to engage and challenge students. Halverson (2010) provided a modern context for the use of benchmark assessments as a component of a feedback system. The information obtained from test data, such as Fountas and Pinnell data, is a logical basis for instructional decision making. However, it is critical that policy makers ensure that the information obtained is worthwhile and accurate. When teachers are able to translate worthwhile, accurate assessment data into practical information and use it to modify instruction, student improvement occurs. Benchmark assessments provide information on learning and provide teachers with opportunities to make pertinent adjustments to their instruction (Cauley & McMillan, 2009).

Evaluative Benchmark Assessments

Contrary to instructional uses, evaluative benchmark assessments provide evaluative information about the curriculum or instruction (Marshall, 2008; Popham, 2009). Davidson and Frohbieter (2011) noted that 80% of the district and school leaders declared that they used benchmark assessments for evaluative purposes. Evaluative assessment systems must provide detailed information about the curricular units. When comparing the effectiveness of different instructional programs, evaluative purpose benchmark assessments are used. These assessments can be administered at different points in the year to measure growth. The ultimate goal would be to evaluate the effectiveness of a program, strategy, or teacher. On a much smaller scale, the assessments could be used to identify concepts that students grasped or did not grasp.

Unlike instructional and evaluative assessments, predictive assessments are created to decide the likelihood of each student meeting some criterion score on the end of the year assessment. As the NCLB goals continue to rise, the predictive purposes of benchmark assessments may increase. The Fountas and Pinnell assessment can serve all of these purposes, but for the sake of this study the predictive purpose will be investigated.

Both data –driven decision making principles and formative assessment practices are integrated to form the sole purpose of benchmark assessments. Sawchuk and Cain (2009) explained that benchmark assessments can be used to examine student understanding of content material and to predict student performance on summative state

tests. These assessments are administered to students at multiple intervals throughout the year (Donhost, 2010). Educators are able to use the assessment results to measure students' knowledge of skills taught and to adjust the course of learning in preparation for the end of the year test (Moss & Brookhart, 2009). Benchmark assessments ability to lead teaching practices and inform instructional practices is what places it in the category of a formative assessment. Since benchmark assessments are typically administered two to three times annually and used to measure student's progress toward mastery of state standards, it could also be classified as summative. Schools acknowledge the requirement for data that is both summative and formative.

The data-based decision making aspect of benchmark assessments is also an area explored in this section. Data driven decision making, a complex social process, is a common component of the accountability process and how benchmark assessments improve student achievement (Wilkins & Shin, 2011). Frohbieter, Greenwald, Stecher and Schwartz (2011) recognized three elements of benchmark assessments and how each enhances teachers' knowledge and use of data: These included purpose, cycle of use, and planned use for instruction. Benchmark assessments are given with the purpose of assessing strengths and weaknesses. They are administered three times a year annually. The assessment data from these benchmark assessments are used to plan and support instruction. Marshall (2008) suggested that benchmark assessments, if handled well, constitute the most effective single initiative that a principal can implement.

Another study found that the absence of instructional decision-making process, related to benchmark assessment data, caused teachers to make unproductive instructional decisions (Olah, Lawrence, and Riggan, 2010). Similarly, Olah, Lawrence, and Riggan (2010) declared that benchmark assessments alone only identify student learning needs. These researchers discovered that the absence of benchmark assessment data to inform the instructional decision making process led to ineffective instructional decisions by teachers.

Use of Benchmark Data in Schools

Hess and Mehta (2013) distinguished four issues with the use of data in schools. Hess and Mehta (2013) noted how teachers often lack the professional development needed to develop an understanding of how to use data. They also pointed out the political influence of data use in schools. The third issue involved educators not understanding the types of data to use or the purpose for using the data. Hess and Mehta (2013) noted the fourth issue as educators and leaders lacking the training necessary for them to analyze, interpret, and respond to data.

Halverson (2010) suggested that schools establish programs that accomplish the three functions of intervention, assessment, and actuation. Interventions consist of two tiers, district policies, and school based practices. Assessments provide data to educators so that they can determine what students have learned. Actuation is the process of analyzing data and using results interpretations to modify instructional practices. Recent data also suggests that when benchmark assessments, such as the Fountas and Pinnell, are

used to assess current understandings, student achievement is increased (Black &

Wiliam, 2009; Clark, 2011). However, before decisions can be made based on data,

schools must agree on the value of the data (Young & Kim, 2010). This study seeks to

determine if a relationship exists between the benchmark data and state data. The results

of this study will assist stakeholders in establishing the value of the benchmark data

currently being used.

Shepard (2008) argued that benchmark assessments should depict important

learning goals, be connected to instructional units, be consistent with the pacing of the

curriculum, and provide information that is unavailable from other sources. Schools are

increasingly administering benchmark tests, created to serve as an early warning system

for the state accountability test, in addition to the annual tests administered as part of

NCLB (Bambrick-Santoyo, 2010; Borja, 2006; Honawar, 2006; Zehr, 2006). The

benchmark assessment is an integral element in connecting classroom practices with

high-stakes accountability tests. The benchmark assessment acts as a bridge between the

formative and summative assessments. It depicts learning goals, is connected to

instructional units, and provides teachers with data that allows them to track student

progress throughout a school year. Achievement data obtained from benchmark

assessments is used to determine the quality of instruction and curriculum being

implemented. The connection can only be made if teachers use the data and make data-

based instructional decisions in the classroom. Schools must "link everyday classroom

practices with school wide outcome" and "develop data-driven practices" if they hope to meet accountability targets (Halverson, 2010).

Predictive Validity of Benchmark Assessments

Despite the fact that there is limited evidence on the effectiveness of benchmark assessments as a measurement tool for increased student achievement, districts continue to increase their use (Shepard, 2010). Considering the amount of time, money, and resources allocated to benchmark assessments, it is critical that further research is conducted to identify key features of benchmark assessment use that contribute to student learning improvements. Educators give various benchmark assessments in an effort to collect data. The data is used for various purposes such as to drive instruction, monitor students' progress, to predict what level a student is performing at, and to make other educational decisions (Merino & Beckman, 2010). A few studies have emphasized the need to make data important within educational systems (Wayman, Cho, & Shaw, 2009). Other studies provide evidence of data use leading to improvement in student achievement (Carlson, Borman & Robinson, 2011; Slavin, Holmes, Madden, Chamberlain & Cheung, 2011, Lai et al., 2009; Faria et al., 2012; Campbell & Levin, 2009). Two studies initiated by the Center for Data-Driven Reform in Education (CDDRE) noted results from investigations of a benchmark assessment initiative. The benchmark assessment initiative was implemented in 59 districts across seven states. Quarterly benchmark assessments were developed and administered. Consultants developed and administered the assessment, and trained the faculty members on how to

use the data to identify weaknesses. After 12 months of the in initiative Carlson et al. (2011) revealed that the program had a small, positive impact on math achievement. The study also indicated that there was a positive impact on reading achievement, yet not significant. Slavin et al. (2011) found greater effects on elementary reading and math achievement after four years of the CDDRE intervention.

A similar study conducted by Lai et al. (2009), investigates data use as an intervention to improve student achievement. The quasi-experimental study occurred in New Zealand over a five year period. The results of the study indicated that when teachers and school leaders use data to enhance their teaching methods, there is a significant increase in student achievement. Campbell & Levin (2009) claimed that effective use of assessment and data to support positive outcomes for educators requires careful attention to building capacity to access, understand and apply data. Faria et al. (2012) conducted a similar study that examined the relationship between teachers' and schools' use of benchmark test data and student academic improvement in reading and math. Grades 4, 5, 7, and 8 participated in the study. The researchers found a slightly significant effect for teachers' use of benchmark assessments on student achievement.

The study by Deno et al. (2009) examined the relationship between student performance on a maze universal screener, used as a benchmark test, and a standardized reading test. The two reading assessments yielded measures of .61 to .77. The results of this study provided support for using CBM benchmark data to predict performance on state assessments.

Pearce and Gayle (2009) conducted a study at a Reading First school in South Dakota to determine if the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency scores could predict subsequent reading scores on the Dakota State Test of Educational Proficiency. Pearce and Gayle found that the benchmark assessment predicted comprehension on the state test. It specifically indicated students who might score proficient, opposed to students who would fail. Hefflin (2009) conducted a similar study to determine if benchmark scores could be used to increase student performance on the Pennsylvania System School Assessment (PSSA). He found that the benchmark and high-stakes testing scores were linked. Hefflin (2009) noted that students' involvement in data analysis contributed to the improved performance.

Another study conducted by Keller-Margulis, Shapiro, and Hintze (2008) also investigated the relationship between benchmark assessments and a statewide achievement test. The reading group consisted of 1,461 students, and the math group consisted of 1,477 students. Participants were taken from six elementary schools and included students from grades 1 through 5. AIMS web probes were used to measure students' oral reading proficiency. Scores were compared to the Pennsylvania System of School Assessment (PSSA), and to the Terra Nova Achievement Test-Second Edition. Results suggested that the CBM data were moderately correlated to statewide achievement data. This study also supports the use of CBM to predict performance on state accountability tests.

Brown and Coughlin (2007) conducted a study on the predictive validity of selected benchmark assessments used in the Mid-Atlantic Region. Company generated benchmarks were used in their study. The MAP, STAR, and TerraNova showed some validity, but the TerraNova provided the most appropriate predictive information for one state. Similarly, Feng et al. (2008) article Can an Intelligent Tutoring System Predict Math Proficiency as Well as a Standardized Test, indicated that these assessments give some data, but not enough to explore trends.

All of the studies were conducted with different purposes in mind. However, each one resulted in reading benchmark assessments being strong indicators for overall proficiency on state assessments. These studies provide promising evidence that proper data use can lead to increased student achievement, resulting in school improvement.

Item Response Theory

Item response theory, also known as latent trait theory, attempts to model an association between an examinee's responses to items and an underlying latent trait that is measured by the items. This theory has been utilized frequently in the development of standardized tests. The basis of the Item Response Theory is Louis Thurstone's paper entitled," A Method of Scaling Psychological and Educational Testing." Thurstone's work provided an approach for placing the items of the Binet and Simon test of children's mental development on an age appropriated scale. IRT calibration and scale linking methodologies can be used to place benchmark assessments items on the same scale as the reading CRCT test items. Additional pioneering work related to IRT is attributed to

Lord and Wingersky. Lord and Wingersky (1984) introduced a procedure to estimate the

IRT possibility for each summed score on a test. Multiple IRT estimations may be used to

place the summed scores on the same scale. One application would involve calibrating

each test, X and Y, and then use a method similar to the Stocking –Lord Transformation

to link the IRT scales (Stocking & Lord, 1983). Another application is to concurrently

calibrate the items in a single run, which enables the scores from both tests to be used

interchangeably. Calibrating two different tests for a specific purpose can be achieved

using an IRT approach (Linn, 1993).

Fountas and Pinnell Benchmark Assessment System

Educators determine students reading competencies and limitation in an effort to

determine their instructional reading levels. Determining accurate instructional levels is a

critical aspect of planning appropriate interventions to support students. It is imperative

that schools correctly monitor the progress of readers and provide appropriate

interventions that will improve reading ability. Greenstein (2012) stated that 21st century

evaluations must be obvious, adaptable, and sensitive to learners' needs, educational, and

combined with teaching and learning. A precise measurement tool that complies with the

above standard and is used to measure students' progress in reading before the state test

is the Fountas and Pinnell Benchmark Assessment System. This assessment presents

educators with information to improve teaching and learning in reading. Teachers

administer the assessment in approximately 20 to 30 minutes in a one on one teacher

student conference type session.

The Fountas and Pinnell Benchmark Assessment is administered three times a year. The assessment is time consuming; therefore, it is imperative to determine if this assessment collects data that aids in the preparation for the CRCT. The purpose of this nonexperimental ex post facto quantitative research study is to examine whether a statistically significant correlation existed amongst student achievement scores on the Fountas and Pinnell Reading Benchmark Assessment and reading comprehension scores on the Criterion Reference Competency Test (CRCT). The researcher seeks to determine if the Fountas and Pinnell Instructional Reading Levels predict reading scores on the CRCT. Measurement systems that are capable of accurately measuring reading skills and provide the necessary information for teachers to adjust instruction are critical to the success of students. In an effort to meet the higher standards of education, and promote students with the skills needed to be successful in school, college, and careers, stakeholders are seeking assessments that offer diagnostic information that assist educators in preparing students for standardized assessments.

There are three parts to the Fountas and Pinnell assessment. There is an oral reading component, a questioning component, and a written assessment part. To carry out the assessment, a student reads aloud and talks about a leveled book. The teacher writes down the accuracy rate, oral reading fluency rate, and comprehension assessment. The comprehension conversation assessment is given shortly after the oral reading assessment. Students can obtain a score from 0-3 in the following categories: Within The Text, Beyond The Text, and About The Text. Summed scores are determined and the

following rankings are given: Excellent, Satisfactory, Limited, Unsatisfactory (Fountas, 2010). The teacher uses the assessment data to assign students an instructional text level. The text levels range from A to Z. A is the lowest level and Z is the highest. The Fountas and Pinnell Assessment assess students reading comprehension, or ability to construct meaning while reading. Reading deficiencies prohibit students from being successful in other educational task, including performance on state tests. The theory of latent traits assumes that in testing situations, examinee performance on a test can be predicted by defining traits. Because the Fountas and Pinnell Test Preparation Items are aligned with the Common Core State Standards, teachers should be able to use the results of the Fountas and Pinnell to predict performance on the CRCT. Educators should also be able to use the result of the Fountas and Pinnell assessment to guide instruction in an effort to improve performance on the CRCT.

Research Methodology

This research study that examines the relationship between a reading benchmark assessment and a high-stake test incorporated a quantitative nonexperimental ex post facto correlational design approach. Correlational research design is a nonexperimental methodology used to determine if a predictive relationship exists between two variables (Creswell, 2013). According to Creswell (2013) quantitative research is used to collect statistically- analyzed numerical data, and to investigate relationships between known variables. Creswell (2013) noted that a quantitative method is imperative when testing objective theories by exploring relationships among variables

Experimental or nonexperimental designs are the two most common types of quantitative research (Lodico, Spaulding, & Voegtle, 2010). Numerical data is used to determine cause–effect relationships when conducting experimental quantitative research, whereas nonexperimental quantitative research uses numerical data to determine whether a relationship exists between variables or describe preexisting groups (Lodico, Spaulding, & Voegtle, 2010). Causal –comparative and correlational are the two types of nonexperimental designs. Correlational nonexperimental designs try to determine the extent to which relationships exist between variables (Lodico, Spaulding, & Voegtle, 2010). A Spearman's Rank-Order Correlation will be used to determine the relationship between the two assessment test scores. The researcher does not control or manipulate the variables in the study (Lodico, Spaulding, & Voegtle, 2010).

Correlational research design is a nonexperimental methodology used to determine if a relationship exists between two variables (Creswell, 2013). Since this study's variables cannot be manipulated, nonexperimental correlational research design is appropriate. Experimental research generally looks to see if one variable has an impact on another variable, while nonexperimental research determines whether or not there is a relationship between variables (Lodico, Spaulding, & Voegtle, 2010). Nonexperimental means that the independent variable will not be manipulated and random sampling will not occur since data has already been recorded. Examining relationships between the variables will be the focus of this study. The two tests are not scored in the same way. A passing score on the CRCT is 800. The numerical instructional level for meeting

standards varies by grade level. The two variables of interest in this study are the 2013-2014 Spring Fountas and Pinnell instructional reading levels and the 2013-2014 Spring CRCT scores. This design will use descriptive and inferential statistics. Correlational research does not determine causation, but allows for conclusions to be formed about the relationship of the Fountas and Pinnell Assessment and the CRCT.

A non-parametric version of Pearson Correlation (where data is not assumed to be normally distributed) - a Spearman's Rank-Order Correlation will be used to determine the relationship between the two assessment test scores (Salkind, 2008). The Spearman's rho statistics measure the rank-order association between two scale or ordinal variables. They work regardless of the distributions of the variables. The measures of rank order are handy for discovering whether there is any kind of association between two variables, but when they find an association it's a good idea to find a transformation that makes the relationship linear. This is because there are more predictive models available for linear relationships, and the linear models are generally easier to implement and interpret. A visual representation of the data through the use of a scatter plot will further assist the researcher in determining whether or not a relationship exists between the two variables (Salkind, 2008). The Spearman's rank-order correlation coefficient will be the correlational technique used as it provides more robust results in categorical scales as compared to other bivariate correlations such as Pearson's correlation coefficient and Kendall's tau-b.

Since this study examines the extent to which instructional reading levels, on a reading benchmark assessment, are related to reading scores on the CRCT, a quantitative nonexperimental ex post facto research design was selected over qualitative or experimental, designs. This method was appropriate for accomplishing the goal of collecting numeric data to examine the relationship between the Fountas and Pinnell Benchmark assessment levels and the state CRCT scores. Several researchers have established the validity of using benchmark assessments to predict the performance on state reading tests (Hintze & Silberglitt, 2005; Stage Jacobsen, 2001).

Most of the studies investigated oral reading or the maze. Very few studies investigated the relationship between benchmark assessments and the CRCT. An ex post facto design will also be used to investigate the F and P benchmark assessment and the CRCT. The research design will determine if a relationship exists between the two variables. The phrase ex post facto actually means "after the fact". Archival data for this study will be collected ex post facto, meaning the assessments were administered and scored before the research study began (Howell, 2011). Using an ex post facto design will allow the researcher to examine the students by looking at the facts (data) that already exists. The researcher has no control over the independent variables since the events have already occurred (Lohmeier, 2010). The quantitative nonexperimental post facto research design was considered acceptable because archival data will be collected, and no variable will be manipulated. Quantitative methods will be used to find possible correlations between the Fountas and Pinnell Instructional Levels and CRCT results.

Summary

States are now required to administer yearly reading and math assessments to grades three through eight, as a result, of No Child Left Behind Act. In an effort to detect deficiencies early and improve student learning and performance on statewide tests, various supplemental assessments are being administered. Benchmark assessments are being designed and implemented to serve as an early warning system for the state summative test. There is a significant amount of evidence in the literature review that show that benchmark assessments are effective tools for monitoring student progress and predicting performance in different content areas (Cusumano, 2007; Deno 2003; Fewster Macmillan, 2002). Furthermore, the literature review presented an overview of the benchmark assessments, and a substantial amount of literature on standardized assessments, and data-based decision making. This literature review aimed to investigate the predictive validity of a formative benchmark assessment as it relates to summative state assessments. State summative assessments are poor tools for guiding learning at the classroom level but are beneficial for informing accountability decisions. Therefore, it is imperative that the predictive validity of Fountas and Pinnell, which is used at the study school, is investigated. The investigation will determine if utilizing this assessment maximizes and predicts performance on the state CRCT assessment.

The review of the literature indicates that reading interventions when used early, effectively, and intensely can prevent readers from further failure (Torgesen, 2009). The idea is to follow the Skill Deficit Theory and intervene before a child fails or falls so far

behind that it seems impossible for them to catch up. Children need to be taught directly and intensively to learn the skills needed to read. Intervention programs can work. The focus of this review was to show that early intervention plays a key role when children struggle with learning to read. The longer the struggle the harder it is for children to recover. Experiences and development encourage the process of reading, but many skills need to be taught explicitly. Learning needs to be done in a scaffolding approach and build on what children already know to help get them to what they do not know yet (Fountas & Pinnell, 2010). The *LLI* program is one approach that incorporates a multi-faceted approach to learning. It is a supplementary tool that has shown it can make progress for some readers. When the *LLI* program was implemented, results from Fountas and Pinnell (2010) noted an increase in some children who were able to get back to grade level in reading.

In conclusion, summative assessments are commonly used as instruments for summarizing student learning and for ranking students and schools. Formative benchmark assessments take a different approach and may be able to transform both teaching and learning. Furthermore, since empirical research on the use of benchmark assessments is sparse, there is a need for more empirical research to determine whether formative benchmark assessments can predict student performance on state reading assessments.

The next section, Section 3, contains a description of the research methodology that includes the research design, population, sample, implementation procedures, and

instrumentation. The results of the study will be presented in Section 4. The final section

will present the interpretation of the results, implications for social change,

recommendations for action, and recommendations for future research.

Section 3: Methodology

This section begins with an introduction of the study's purpose and research questions. An explanation of the research design, setting, and sample follows. A discussion of instrumentation and materials will proceed. Then, procedures for data collection and analysis will be presented. The section concludes with a description of the measures taken to ensure that participants' privacy rights are protected.

The purpose of this nonexperimental ex post facto quantitative research study was to examine whether a statistically significant correlation exists between student achievement scores on the Fountas and Pinnell Reading Benchmark Assessment and reading comprehension scores on the Criterion Reference Competency Test (CRCT). The reading portion of the CRCT was more often the cause of failure that resulted in retention, making the reading section the target of this study. There could be many reasons why some students fail the reading portion of the CRCT; nonetheless, one question that has not been answered is are the reading Fountas and Pinnell Instructional Levels correlated to the reading scores on the CRCT? If so, can the Fountas and Pinnell predict how well students will do on the CRCT? A correlation is hypothesized between student instructional reading level on the Fountas and Pinnell assessment and students' performance on the reading section of the Georgia CRCT. Currently, no studies address the relationship between the Fountas and Pinnell Comprehension assessment and scores on the CRCT. Teachers in the system sometimes questioned the validity of the Fountas and Pinnell and speculated as to whether the test was aligned to the standards and if the

students' results on the Fountas and Pinnell related to students' results on the CRCT. This nonexperimental ex post facto study research design was needed to determine if the Fountas and Pinnell instructional levels are related to the reading scores on the CRCT.

Research Design and Approach

A quantitative, nonexperimental ex post facto, correlational research design was used to test the single hypothesis. Correlational research design is a nonexperimental methodology used to determine if a predictive relationship exists between two variables (Creswell, 2013). Since this study's variables cannot be manipulated, correlational research design is appropriate. Experimental research generally looks to see if one variable has an impact on another variable, while nonexperimental research determines whether or not there is a relationship between variables (Lodico, Spaulding, & Voegtle, 2010). Nonexperimental means that the independent variable will not be manipulated and random sampling will not occur since data has already been recorded. Examining relationships between the variables will be the focus of this study. The two tests are not scored in the same way. A passing score on the CRCT is 800. The numerical instructional level for meeting the standard varies by grade level. The two variables of interest in this study are the 2013-2014 Spring Fountas and Pinnell instructional reading levels and the 2013-2014 Spring CRCT scores. This design will use descriptive and inferential statistics. Correlational research does not determine causation, but allows for conclusions to be formed about the relationship of the Fountas and Pinnell Assessment and the CRCT.

A nonparametric version of Pearson Correlation (where data is not assumed to be normally distributed) a Spearman's Rank-Order Correlation was used to determine the relationship between the two assessment test scores (Salkind, 2008). The Spearman's rho statistics measure the rank-order association between two scale or ordinal variables. They work regardless of the distributions of the variables. The measures of rank order are useful for discovering whether there is any kind of association between two variables, but when they find an association it is beneficial to find a transformation that makes the relationship linear (Salkind, 2008). This is because there are more predictive models available for linear relationships, and the linear models are generally easier to implement and interpret. A visual representation of the data through the use of a scatter plot further assisted the researcher in determining whether or not a relationship exists between the two variables (Salkind, 2008). The Spearman's rank-order correlation coefficient was the correlational technique used as it provided more robust results in categorical scales as compared to other bivariate correlations such as Pearson's correlation coefficient and Kendall's tau-b. According to Jackson (2009), studies use correlation coefficients when one or more of the variables measured are on an ordinal scale. The Spearman's Rank-Order Correlation was used to measure the correlation between Fountas and Pinnell instructional levels (scores) and CRCT assessment scores. Spearman's Rank-Order Correlation was selected to analyze the data in this study since it will reveal the direction of the relationship between the two variables. Since a relationship existed between the

two variables, in the future a linear regression model will be needed to determine

predictive ability.

According to Drew, Hardman, and Hosp (2008), a correlational design using the

Spearman rank order correlation requires a sample size greater than 10. If there is a small

sample, less than 10, Kendall's tau is preferred. The basis of this number is on the size

needed for statistical procedures so that the sample is likely to be a good estimate of the

characteristics of the population (Creswell, 2013). The Spearman rank order determines

the relationship between two ranked variables, instead of interval or ratio variables

(Neutens & Rubinson, 2010). Nonparametric tests and t-tests that compare independent

group differences were not appropriate because they do not express associative

relationships between the variables in terms of correlation. The design of the Spearman

rank order correlation is for nonparametric, rank-ordered data represented by ratings on

Likert-type survey scales. The Spearman rank order correlation coefficient was

appropriate because the survey in this study will be collected through rank ordered data

using Likert-type scales.

A quantitative design was used to support the theory that students lacking the

prior knowledge to perform well on prerequisite tests, such as the Fountas and Pinnell,

will not perform well on subsequent tests, such as the CRCT. Furthermore, the results

from standard based assessments can be used for decision making (Price & Kortez,

2005). This theory indicates that the assessment data from the Fountas and Pinnell

assessment can be used for decision making that may result in improved student performance.

Setting and Sample

The research took place at an elementary school located in a suburb of Atlanta, Georgia. In the 2011-2012 school year, the enrollment size was 664, while, in 2012-2013, it slightly increased to 674. The school serves students in grades K-5 in the Henry District School District. Currently, there are approximately 300 third, fourth, and fifth grade students at the research setting. The school wide demographics consisted of 44% African American, 5% Hispanic, 45% white, and 6% other. Fifty-six percent of the students that attend this school receive free or reduced lunch.

Population and Eligibility Criteria

The population for this study consisted of all students' spring Fountas and Pinnell Instructional Levels and CRCT scores during the 2013-2014 school years in the Georgia schools. The students must have been in grades 35 during this school year. There were approximately 329 students in the ABC school that fit this criterion (Henry District Public School System, 2013).

The ideal participants for this study were third, fourth, and fifth grade students. Of the 750 students at the elementary school in this study, 330 are third, fourth, or fifth grade students. Considering the availability of the targeted subject group, I used one criterion to exclude students from the sample. To be included in the study, students must have taken both assessments (Fountas and Pinnell and CRCT) during the 2013-2014 school year.

Students with incomplete data sets will not be included in the study. The entire population fit this criterion and was studied. Sex and socioeconomic status was an inclusion criterion. Random sampling would allow me to generalize the findings of this study back to the entire population from which the sample will be drawn (Lodico et al., 2010). However, I used nonrandom, convenience sampling since archival data has been collected and is available, with permission.

Instrumentation

Currently, numerous assessments exist that reliably and accurately measure reading comprehension skill deficits in students. The Fountas and Pinnell Assessment was created to provide reliable diagnostic information related to student comprehension ability (Fountas & Pinnell, 2010). Fountas and Pinnell Assessment consist in a series of texts that can be used to identify a student's current reading level. The assessment consists of an oral reading component, a conversation component, and can include a written component. The study school does not use the written component. Fountas and Pinnell scores are created on an ordinal scale. For the purpose of performing statistical analysis, the Fountas and Pinnell Instructional Levels (variables) have been coded. Each instructional reading level (A-Z) is assigned a numerical value (1-26) with consistent intervals between all measurement points. This benchmark test is a district wide universal screener assessment designed to measure kindergarten through fifth grade students' reading proficiency. Henry County School District began administering this test in 2010-2011.

Fountas collaborated with Pinnell to create the Fountas and Pinnell Benchmark Assessment System. The assessment system can be used as a benchmark assessment as well as a progress monitoring tool to assess student's reading skills. This measure is used to assess phonological awareness, phonics, vocabulary, fluency, and comprehension. Test preparation items in the Fountas and Pinnell align with the Common Core State Standards for English Language Arts. The Fountas and Pinnell assessment takes approximately 20 to 30 minutes to administer in a one to one student teacher conference style. This assessment is comprised of three parts. There is an oral reading component, a questioning component, and a written part. A comprehension conversation (questioning component) follows the reading of a text and allows students to demonstrate literacy skills. The psychometric properties of the Fountas and Pinnell test were found to be reliable where internal consistency of items yielded $r = 0.97$. School districts in Georgia administer the Fountas and Pinnell assessment three times per school year. To measure the skills students should have mastered the previous year; the first test is given at the beginning of the school year in the month of September. The second test is administered in December and the final Fountas and Pinnell assessment is administered in March.

In the study, instructional reading levels were used. A student's instructional reading level is the level in which they can read 90 -95 % of the words in a text with satisfactory or excellent comprehension. Levels A and B are generally for kindergarten and first graders. These students are considered emergent readers. Students in levels B through H (2-8 coded) are considered early readers. They need to spend more time

focusing on fluency. Levels H-M (8-13 coded) are your transitional readers. Educators are faced with the challenge of developing transitional readers into advanced readers. The Fountas and Pinnell scores are reported using the following scale based on instructional levels: (Does Not Meet Expectations, Approaches Expectations, Meets Expectations, and Exceeds Expectations). The scaled scores will be compared to the scaled scores on the CRCT: 1=(does not meet), 2=(meets), or 3=(exceeds).

The second instrument, the CRCT, is an assessment used to test students in grades 2 - 8 in the state of Georgia. The CRCT is designed to measure student mastery of the Common Core Georgia Performance Standards (CCGPS). Students are assessed in the areas of mathematics, language arts, reading, social studies, and science. The reading portion of the CRCT was used for this study. Students' achievement on the reading part of the CRCT is reported in two forms: scale scores and achievement levels. The assessment is designed to reveal individual strengths and weaknesses as it relates to the curriculum. The CRCT scores are reported using a scale based on cut scores for performance levels based on scaled scores. The scaled scores range from 800 to 920 and are divided into three categories called achievement levels with one being the lowest level(does not meet 799 or less) of achievement to three, which is the highest achievement level (exceeds, 850 or higher). A scale score of 2(between 800 and 849) would mean a student meets the standards (Georgia Department of Education, 2012). For comparison, CRCT reading scale scores (interval data) were collected.

The Reading CRCT assessment consists of 50 items designed to measure student achievement of the fundamental skills and knowledge of the state's content standards. The assessment is composed of 40 operational items and 10 field-test items. The operational items contribute to the student's score, the field-test items do not. Three domains make up the Reading CRCT. The domains are: Literary Comprehension, Reading for Information (3rd), or Information and Media Literacy (4th and 5th), and Reading Skills and Vocabulary Acquisition. Literary Comprehension requires students to comprehend literary works. The Reading for Information domain for third graders and the Information and Media Literacy domain for fourth and fifth grade require students to comprehend, recall, and analyze informational text. Reading Skills and Vocabulary Acquisition refer to skills required to acquire new vocabulary. These domains assess the CCGPS reading standards, which are divided into the following categories: Key Ideas and Details; Craft and Structure; Integration of Knowledge and Ideas; Range of Reading Levels of Text Complexity, and Foundational skills.

Professional content specialists developed the items on the CRCT and allowed Georgia educators and curriculum specialists to review these items (Georgia Department of Education, 2010). In 2004, it was found that the mathematics CRCT total test reliability ranged from 0.89 to 0.90 (Georgia Department of Education, 2012). Reliability indicates whether the same measurement gives the same or comparable result for the same student every time (Georgia Department of Education, 2012). The Georgia Technical Advisory Committee (TAC), an independent group of experts in the area of

educational measurement, convened with the Georgia Department of Education's Testing Division quarterly to ensure that the CRCT was a high quality test (Georgia Department of Education, 2012). The CRCT is administered to students each April. School districts' yearly progress is determined based on the results of the CRCT.

The CRCT scoring system reports interval scale scores. The Lowest Obtainable scale score on the CRCT (LOSS) on the 2012 and 2013 Reading CRCT is 650, and the Highest Obtainable Scale Score (HOSS) is 920. Scores of 850 or above indicate that students have exceeded the state's standards; scores from 800-849 met the standards, and scores below 800 did not meet the standards. A relationship between the Fountas and Pinnell instructional levels and CRCT in the area of reading in grades 3, 4, and 5 in the year 2013-2014 may exist.

Both instruments (CRCT and Fountas and Pinnell) were designed to measure how well students obtain the reading skills and knowledge described in the Reading Common Core Georgia Performance Standards (CCGPS) for his or her grade level. The Common Core English/language arts standards emphasize an integrated approach to language arts. The standards are divided into reading, writing, speaking and listening, and language standards. At the elementary level each of these standards is connected to a set of College and Career Readiness (CCR) Anchor Standards and then unpacked into grade level standards. There are 10 reading CCR Anchor Standards which are divided into the following categories: Key Ideas and Details; Craft and Structure; Integration of Knowledge and Ideas; and Range of Reading Levels of Text Complexity. The 10 CCR

Anchor Standards are further broken down into grade level standards for reading

literature and reading informational text. Furthermore, in the elementary grades, there are

also standards for foundational reading skills: print concepts, phonological awareness,

phonics and word recognition, and fluency. The foundational skill standards do not have

CCR Anchor Standards; they candidly make certain that students obtain the foundational

knowledge required for students to learn to read.

Data Collection

With granted permission to conduct this study, a single data file containing the

Fountas and Pinnell Spring Reading levels for 2013-2014 and CRCT assessment scores

for 2013-2014 was retrieved from the school system's central data warehouse. The

researcher collaborated with the testing director to extract student's data from the POINT

database during the winter of 2014. The data included students' Fountas and Pinnell

instructional levels and CRCT test results for the 2013-2014 school year for third, fourth

and fifth grade students. Demographical information was removed from the downloaded

data to protect students' identity. This included gender, race, age, economic status, and

students with disabilities.

The data was transferred to a Statistical Package for the Social Sciences (SPSS)

spreadsheet where it will be organized by grade level. The Spearman correlation was

utilized in this study since one of the variables consists of categorical ranked data

(Gravetter & Wallnau, 2008). This study measured the relationship between students'

achievement on the reading portion of the CRCT, which will be represented by a

categorical score and the instructional reading levels of the Fountas and Pinnell Assessment. Below level, instructional levels will be assigned a numerical value of 0, on level instructional levels will be assigned a numerical value of 1, and above level instructional levels will be assigned a value of 2.

The instructional text level, predictor variable, consists of 26 levels measured A to Z. Level A is the easiest and level Z is the most difficult. Instructional text levels A-K is determined by finding the level at which the students reads the text with 90 to 94% accuracy and has satisfactory or excellent comprehension. Furthermore, 95% or higher accuracy and limited comprehension is sufficient for determining A-K instructional level. Instructional levels L-Z is determined by finding the level at which the students reads the text with 95-97% accuracy and has satisfactory or excellent comprehension, or 98% accuracy and limited comprehension (Fountas & Pinnell, 2010). The comprehension conversation assessment is used to determine students' instructional level (Fountas & Pinnell, 2010). Students earn 0-3 in three comprehension question categories. The three categories are within the text, beyond text, and about the text. Scores are totaled, and rankings of excellent, satisfactory, limited, and unsatisfactory are given. This comprehension score is one factor used to determine the instructional text level. Accuracy rate is a second factor used to determine instruction level. Accuracy rate is the percentage of words that students can correctly read. The teachers note errors that students make while reading on the assessment form. Errors are added up, and the accuracy chart is used to determine the percentage of words that the student read

correctly. Both comprehension scores and accuracy percentages are used to determine a students' instructional reading level based on the scales noted above. CRCT scores, predictor variable, were also reported as a numerical score.

One research question was developed to guide this study based on the limited literature on Fountas and Pinnell instructional levels and their correlation (if any) with CRCT scores. The research question is as follows:

Research Question (RQ): Is there a relationship between student achievement scores on the Fountas and Pinnell Reading Benchmark Assessment and student achievement scores on the reading portion of the CRCT?

H_0: There is no relationship between the spring Fountas and Pinnell Benchmark Assessment Instructional Reading Levels and the reading comprehension scores on the Criterion Reference Competency Test (CRCT) in grades three, four, and five in the year 2013-2014.

H_A: There is no relationship between the spring Fountas and Pinnell Benchmark Assessment Instructional Reading Levels and the reading comprehension scores on the Criterion Reference Competency Test (CRCT) in grades three, four, and five in the year 2013-2014.

Procedure and Data Analysis

The data analysis included descriptive statistics where applicable to provide a profile of the participants in the study and find if there is a correlation between the two variables of interest. A Spearsman's Rho correlation was used to test the single

hypothesis. The outcome of a correlational analysis is a correlation coefficient used to specify whether a linear relationship exists between two variables measured on the same subject (Leedy & Ormrod, 2010). The Spearsman's (rho) rank correlation coefficient was able to provide a statistical method to determine the strength of the relationship between both variables, the reading performance on the Fountas and Pinnell and reading results on the CRCT.

Threats to Validity

Convenience sampling does not offer generalizable data in the universal sense of the meaning. Giorgi asserts that obtaining universality when discussing human experience is not obtainable as context dominates all human experience. In order to generalize, a context-similar situation for the phenomena under study would have to exist.

One potential threat to the validity of this research study was the use of only one school to look at correlation of scores. Utilizing the scores of more schools in the district with similar demographics in the future would strengthen the study. The large number of students in the study compensated the mobility of students in and out of the school. The other limitation was related to the design of the study. Convenience samples would supply the researcher with the scores used. There is no guarantee that the generalized results of the convenience sample will apply to other groups given the non- experimental design. Extending the research to other populations may decrease the convenience sample threat.

Protection of Participants

The researcher followed all appropriate procedures of IRB (10-07-15-0098912) prior to conducting the proposed quantitative research study. Once permission to conduct this study was obtained from the Institutional Review Board (IRB), an informed consent letter was sent to the Henry District administration requesting the 2013-2014 Fountas and Pinnell and CRCT scores (Appendix A). The informed consent letter was also sent to the district's superintendent. The researcher obtained approval from the Henry District administration and superintendent. Ethical standards were practiced and adhered to as specified by IRB. To protect students' rights of privacy, data was stored in a password protected computer on the school grounds in researcher's classroom. The data will remain there for approximately five years. To further protect students' names, responses were assigned numerical codes. The researcher, school principal, and assistant principal were the only people who had access to the data. Consent from Henry District Board of Education Research Department for research and Data Use Agreement was obtained and included in the appendix.

Role of the Researcher

As a member of the faculty in the study's targeted school, I was at the site where data was collected. I am a third grade teacher who is a grade level chair and part of the school's leadership team. In a professional manner, I analyzed the data that was available through the school district's website.

Summary

With attention on student achievement, teachers must take into account the role of assessments in learning. Classroom assessment practices are currently changing (Frey & Schmitt, 2007; McMillan, 2007; Popham, 2006, Tierney, 2006). In the state of Georgia, school districts are continuing to use CRCT for accountability purposes and to determine if schools make adequate progress. If schools perform poorly on this assessment and do not make adequate progress, sanctions are placed on the schools. This quantitative, correlational study seeks to determine the extent to which Fountas and Pinnell instructional reading levels and reading CRCT scores are related. The findings may provide evidence needed to determine if Fountas and Pinnell assessment scores indicate how students might perform on future state assessments. The design and methodology that was used to conduct the research study was discussed in this section. An interpretation and explanation of the results of the study will be presented in Section 4. The conclusion for this quantitative research study will then be provided in Section 5. The conclusion will include implications for social change, recommendation for action, and recommendation for future research.

Section 4: Results

The purpose of this nonexperimental ex post facto, correlational research design was to determine the relationship between the spring Fountas and Pinnell Benchmark Assessment Instructional Reading Levels and the reading comprehension scores on the Criterion Reference Competency Test (CRCT) in grades 3, 4, and 5 in the year 2013-2014. I used regression analysis to test the null and alternate hypothesis. The participants consisted of third, fourth, and fifth grade students' spring Fountas and Pinnell Instructional Levels and CRCT scores during the 2013-2014 school year in the Georgia schools (archival data). This section provides the data analyses and findings of the research study.

Research Tools

The research tools used to conduct this study were Fountas and Pinnell and Reading CRCT. Fountas and Pinnell and Reading archival test data was collected for 329 3rd-5th grade students with complete data sets. A single data file containing the Fountas and Pinnell Spring Reading levels for 2013-2014 and CRCT assessment scores for 2013-2014 was retrieved from the school system's central data warehouse. After the data was collected the data was transferred to a Statistical Package for the Social Sciences (SPSS) software where it was organized by grade level. Regression analyses were used to test the null and alternative hypothesis.

Data Analysis Procedure

Inferential statistics were used to draw conclusions from the sample tested. The Statistical Package for the Social Sciences (SPSS) was used to code and tabulate scores collected and provide summarized values where applicable including the mean, standard deviation, central tendencies and variance. Regression analyses were used to evaluate the three research questions. The research questions were:

RQ1: What is the relationship between the 3rd grade spring Fountas and Pinnell Reading Instructional Levels and the third grade CRCT Reading Comprehension Scores?

RQ2: What is the relationship between the 4th grade spring Fountas and Pinnell Reading Instructional Levels and the fourth grade CRCT Reading Comprehension Scores?

RQ3: What is the relationship between the 5th grade spring Fountas and Pinnell Reading Instructional Levels and the fifth grade CRCT Reading Comprehension Scores?

Prior to analyzing the research question, data cleaning and data screening were undertaken to ensure the variables of interest met appropriate statistical assumptions. Thus, the following analyses were assessed using an analytic strategy in that the variables were first evaluated for missing data, univariate outliers, reliability, normality, linearity, and homoscedasticity. Subsequently, regression analyses were run to test the research questions. Table 1displays a summary of the variables and analyses used to evaluate the four research questions.

Table 1- Variables and Statistical Tests Used to Evaluate Research Questions 1-3

Research Question	Criterion Variable	Predictor Variable	Analysis
1	3rd Grade CRCT Reading Comprehension	3rd Grade Instructional Reading Level	Regression
2	4th Grade CRCT Reading Comprehension	4th Grade Instructional Reading Level	Regression
3	5th Grade CRCT Reading Comprehension	5th Grade Instructional Reading Level	Regression

Demographics

Data were collected from a population of 329 3rd – 5th grade students from

Georgia during the 2013-2014 school years. Specifically, there were 105 students in the

third grade ($N = 105$), 116 were in the fourth grade ($N = 116$), and 108 were in the fifth

grade ($N = 108$). Additionally, 52% of all students were female ($n = 171$), 47.1% were

male ($n = 155$) and three participants did not provide their gender (0.9%, $n = 3$). Table 2

is a presentation of the frequency and percent statistics of participants' gender by grade

levels.

Table 2- Frequency and Percent Statistics of Participants' Gender by Grade Levels

Gender	3rd Grade		4th Grade		5th Grade		Total	
	Frequency (n)	Percent (%)	Frequency (n)	Percent (%)	Frequency (n)	Percent (%)	Frequency (n)	Percent (%)
Female	56	53.3	61	52.6	54	51.4	171	52.0
Male	48	45.7	54	46.6	53	49.1	155	47.1
Missing	1	1.0	1	0.9	1	0.9	3	0.9
Total	105	100.0	116	100.0	108	100.0	329	100.0

Note. Total $N = 329$

As summarized in Table 2, there were 0 Asian students in third grade (*N=0*), three

Asian students in fourth grade (*N=3*), and one Asian student in fifth grade (*N=1*). There

were 47 African American students in third grade (*N=47*), 47 African American students

in fourth grade (*N=47*), and 53 African American students in fifth grade (*N-53*). Two

third grade students were Hispanic (*N=2*), 13 fourth grade students were Hispanic

(*N=14*), and 18 fifth grade students were Hispanic. There were 49 Caucasian third

graders (N=*49*), 48 Caucasian fourth graders, and 44 Caucasian fifth graders (*N=44*). Six

students in third grade were biracial (*N=6*), four fourth grade students were biracial

(*N=4*), and two fifth grade students were biracial (*N=2*). Data was missing from one third

grader (*N=1*), one fourth grader (*N=1*), and one fifth grader (*N=1*).

Table 3- Frequency and Percent Statistics of Participants' Ethnicity by Grade Levels

Ethnicity	3rd Grade		4th Grade		5th Grade		Total	
	Frequency (*n*)	Percent (%)	Frequency (*n*)	Percent (%)	Frequency (*n*)	Percent (%)	Frequency (*n*)	Percent (%)
Asian	0	0.0	3	2.6	1	0.9	4	1.2
African American	47	44.8	47	40.5	53	49.1	147	44.7
Hispanic	2	1.9	13	11.2	3	2.8	18	5.5
Other	0	0.0	0	0.0	0	0.0	0	0.0
Caucasian	49	46.7	48	41.4	48	44.4	145	44.1
Biracial	6	5.7	4	3.4	2	1.9	12	3.6
Missing	1	1.0	1	0.9	1	0.9	3	0.9
Total	105	100.0	116	100.0	108	100.0	329	100.0

Note. Total *N* = 329

Analyses of Research Question 1-3

Research questions 1-3 were evaluated using regression analyses to determine if

any significant relationships existed between 3rd - 5th grade spring Fountas and Pinnell

Reading Instructional Levels and the third grade CRCT Reading Comprehension Scores.

The criterion variables for research questions 1-3 were 3rd – 5th participants' reading

comprehension scores (respectively) as measured by the CRCT. Specifically, two

subscales were used to evaluate the CRCT including Reading (REA) and English Language Arts (ELA). Response parameters were measured on an interval scale were scores less than 799 indicated that students' did not meet standards, scores between 800 and 849 indicated that they did meet standards, and scores 850 or greater indicated they exceeded the standards. The predictor variable was participants' instructional reading levels as measured by the Fountas and Pinnell Assessment. Response parameters were measured on an interval scale ranging between 1 and 26 were $1 = A$, $2 = B$, $3 = C$… $24 = X$, $25 = Y$ and $26 = Z$. That is, higher scores indicated greater levels of reading achievement.

Data Cleaning

Data were collected from a valid population of 329 3rd – 5th grade students from Georgia during the 2013-2014 school years (3rd graders $N = 105$, 4th graders $N = 116$, and 5th graders $N = 108$). Before the research questions were evaluated, the data were screened for missing data and univariate outliers. Missing data were investigated using frequency counts and several cases existed. Specifically, missing cases included two third grade students, five fourth grade students, and seven fifth grade students. The aforementioned cases with missing scores were removed from the analyses of research questions 1-3. The data were screened for univariate outliers and three cases with univariate outliers were found and were removed from the analyses. Thus, between 101 and 110 valid responses from participants were received and were evaluated by the regression analyses for research question 1 ($N = 102$), research question 2 ($N = 110$), and

research question 3 ($N = 101$). Descriptive statistics of the criterion and predictor

variables are displayed in Table 4 by grade levels.

Table 4- Descriptive Statistics of the Criterion and Predictor Variables by Grade Levels

Variables by Grade Levels	N	Min	Max	Mean	Std. Deviation	Skewness	Kurtosis
3rd grade							
Instructional Reading Level	102	9	25	17.314	3.172	0.177	-0.267
REA	102	767	896	845.755	28.852	-0.077	-0.485
ELA	102	788	911	839.559	25.963	0.255	-0.135
4th grade							
Instructional Reading Level	110	13	26	20.182	2.743	-0.015	-0.372
REA	110	781	920	853.246	25.641	0.078	0.387
ELA	110	773	903	836.309	25.758	-0.058	-0.103
5th grade							
Instructional Reading Level	101	13	26	21.970	2.321	-0.501	1.210
REA	101	796	920	838.030	23.604	0.789	1.131
ELA	101	793	888	843.713	20.533	0.307	0.077

Note. 3rd grade total $N = 102$, 4th grade total $N = 110$, and 5th grade total $N = 101$

Test of Normality

Before the research question was analyzed, basic parametric assumptions were

assessed. That is, for the criterion (3rd-5th grade REA and ELA subscale scores) and

predictor variables (3rd-5th grade instructional reading level), assumptions of normality,

linearity and homoscedasticity were tested. Linearity and homoscedasticity were

evaluated using scatterplots and no violations were observed. To test if the distributions

were normally distributed the skew and kurtosis coefficients were divided by the

skew/kurtosis standard errors, resulting in z-skew/z-kurtosis coefficients. This technique

was recommended by Tabachnick and Fidell (2007). Specifically, z-skew/z-kurtosis

coefficients exceeding the critical range between -3.29 and +3.29 ($p < .001$) may indicate

non-normality. Thus, based on the evaluation of the z-skew/z-kurtosis coefficients, no

variables exceeded the critical range. The variables were considered normally distributed

and the assumption of normality was not violated. Displayed in Table 5 are skewness and

kurtosis statistics of the criterion and predictor variables.

Table 5-Shewness and Kurtosis Statistics of the Criterion and Predictor Variables

Variables by Grade Levels	N	Skewness	Skew Std. Error	z-skew	Kurtosis	Kurtosis Std. Error	z-kurtosis
3rd grade							
Instructional Reading Level	102	0.177	0.239	0.741	-0.267	0.474	-0.563
REA	102	-0.077	0.239	-0.322	-0.485	0.474	-1.023
ELA	102	0.255	0.239	1.067	-0.135	0.474	-0.285
4th grade							
Instructional Reading Level	110	-0.015	0.230	-0.065	-0.372	0.457	-0.814
REA	110	0.078	0.230	0.339	0.387	0.457	0.847
ELA	110	-0.058	0.230	-0.252	-0.103	0.457	-0.225
5th grade							
Instructional Reading Level	101	-0.501	0.240	-2.088	1.210	0.476	2.542
REA	101	0.789	0.240	3.288	1.131	0.476	2.376
ELA	101	0.307	0.240	1.279	0.077	0.476	0.162

Note. Note. 3rd grade total $N = 102$, 4th grade total $N = 110$, and 5th grade total $N = 101$

Results of Research Question 1

Null Hypothesis 1 (H1$_0$): There is no relationship between the spring Fountas and

Pinnell Benchmark Assessment Instructional Reading Levels and the reading

comprehension scores on the Criterion Reference Competency Test (REA and ELA

subscale scores) in grade 3 in the year 2013-2014.

Alternative Hypothesis 1 (H1$_A$): There is a significant relationship between the

spring Fountas and Pinnell Benchmark Assessment Instructional Reading Levels and the

reading comprehension scores on the Criterion Reference Competency Test (REA and

ELA subscale scores) in grade 3 in the year 2013-2014.

Using SPSS 23.0, hypothesis 1 was evaluated using two regression analyses to

determine if a significant relationship existed between 3rd grade students' instructional

reading levels and their scores on the two CRCT subscales (REA and ELA). Results from

the first regression analysis indicated that a significant relationship did exist between the

criterion (REA subscale) and predictor variables (instructional reading level), $R = .703$,

$R^2 = .494$, $F(1, 100) = 97.510$, $p < .001$. That is, 49.4% ($R^2 = .494$) of the variance

observed in the criterion variable (REA) was due to the predictor variable (instructional

reading level). Displayed in Table 6, are summary statistics of the regression analysis.

Table 6-Summary of Regression Analysis for Hypothesis 1 using 3rd Grade REA

Subscale and Instructional Reading Level

Source	R	R^2	Standard Error	F	df1	df2	Sig. (p)
Model 1	0.703	0.494	2.268	97.510	1	100	< .001

	Unstandardized Coefficients			Standardized Coefficients			
Source	B	Std. Error		Beta	T	Sig. (p)	Part Correlation
(Constant)	-48.011	6.619			-7.253	< .001	
Instructional Reading Level	0.077	0.008		0.703	9.875	< .001	0.703

Note. Criterion variable = REA subscale. Total $N = 102$

Results from the second regression analysis indicated that a significant

relationship did exist between the criterion (ELA subscale) and predictor variables

(instructional reading level), $R = .680$, $R^2 = .462$, $F(1, 100) = 85.825$, $p < .001$. That is,

46.2% ($R^2 = .462$) of the variance observed in the criterion variable (ELA subscale) was

due to the predictor variable (instructional reading level). Since significant relationships

were found in both regression analyses, the null hypothesis for research question 1 was

rejected in favor of the alternative hypothesis. Displayed in Table 7, are summary

statistics of the regression analysis. Furthermore, displayed in Figures 1 and 2 in

Appendix D is a scatterplot of participants' scores on the criterion (REA and ELA

subscales) and predictor variables by grade levels.

Table 7- Summary of Regression Analysis for Hypothesis 1 using 3rd Grade ELA

Subscale and Instructional Reading Level

Source	R	R^2	Standard Error	F	df1	df2	Sig. (p)
Model 1	0.680	0.462	2.338	85.825	1	100	$< .001$

Source	Unstandardized Coefficients		Standardized Coefficients	T	Sig. (p)	Part Correlation
	B	Std. Error	Beta			
(Constant)	-52.387	7.527		-6.960	$< .001$	
Instructional Reading Level	0.083	0.009	0.680	9.264	$< .001$	0.680

Note. Criterion variable = ELA subscale. Total $N = 102$

Results of Research Question 2

Null Hypothesis 2 (H2$_0$): There is no relationship between the spring Fountas and

Pinnell Benchmark Assessment Instructional Reading Levels and the reading

comprehension scores on the Criterion Reference Competency Test (REA and ELA

subscale scores) in grade 4 in the year 2013-2014.

Alternative Hypothesis 2 (H2$_A$): There is a significant relationship between the spring Fountas and Pinnell Benchmark Assessment Instructional Reading Levels and the reading comprehension scores on the Criterion Reference Competency Test (REA and ELA subscale scores) in grade 4 in the year 2013-2014.

Hypothesis 2 was evaluated using two regression analyses to determine if a significant relationship existed between 4th grade students' instructional reading levels and their CRCT subscale scores (REA and ELA). Results from the first regression analysis indicated that a significant relationship did exist between the criterion (REA subscale) and predictor variables (instructional reading level), $R = .607$, $R^2 = .368$, $F(1, 108) = 62.918$, $p < .001$. That is, 36.8% ($R^2 = .368$) of the variance observed in the criterion variable (REA) was due to the predictor variable (instructional reading level). Displayed in Table 8, are summary statistics of the regression analysis.

Table 8- Summary of Regression Analysis for Hypothesis 2 using 4th Grade REA Subscale and Instructional Reading Level

Source	R	R^2	Standard Error	F	df1	df2	Sig. (p)
Model 1	0.607	0.368	2.191	62.918	1	108	< .001

Source	Unstandardized Coefficients			Standardized Coefficients	T	Sig. (p)	Part Correlation
	B	Std. Error		Beta			
(Constant)	-35.207	6.986			-5.040	< .001	
Instructional Reading Level	0.065	0.008		0.607	7.932	< .001	0.607

Note. Criterion variable = REA subscale. Total $N = 110$

Results from the second regression analysis indicated that a significant relationship did exist between the criterion (ELA subscale) and predictor variables (instructional reading level), $R = .640$, $R^2 = .409$, $F(1, 108) = 74.788$, $p < .001$. That is, 40.9% ($R^2 = .409$) of the variance observed in the criterion variable (ELA subscale) was due to the predictor variable (instructional reading level). Since significant relationships were found in both regression analyses, the null hypothesis for research question 2 was rejected in favor of the alternative hypothesis. Displayed in Table 9, are summary statistics of the regression analysis. Figures 1 and 2 in Appendix D display a scatterplot of participants' scores on the criterion (REA and ELA subscales) and predictor variables by grade levels.

Table 9- Summary of Regression Analysis for Hypothesis 2 using 4th ELA Subscale and Instructional Reading Level

Source	R	R^2	Standard Error	F	df1	df2	Sig. (p)
Model 1	0.640	0.409	2.119	74.788	1	108	< .001

Source	Unstandardized Coefficients			Standardized Coefficients	T	Sig. (p)	Part Correlation
	B	Std. Error		Beta			
(Constant)	-36.794	6.591			-5.582	< .001	
Reading Instructional Level	0.068	0.008		0.640	8.648	< .001	0.640

Note. Criterion variable = ELA subscale. Total $N = 110$

Results of Research Question 3

Null Hypothesis 3 (H3$_0$): There is no relationship between the spring Fountas and Pinnell Benchmark Assessment Instructional Reading Levels and the reading comprehension scores on the Criterion Reference Competency Test (REA and ELA subscale scores) in grade 5 in the year 2013-2014.

Alternative Hypothesis 3 (H3$_A$): There is a significant relationship between the spring Fountas and Pinnell Benchmark Assessment Instructional Reading Levels and the reading comprehension scores on the Criterion Reference Competency Test (REA and ELA subscale scores) in grade 5 in the year 2013-2014.

Hypothesis 3 was evaluated using two regression analyses to determine if a significant relationship existed between 5th grade students' instructional reading levels and their CRCT subscale scores (REA and ELA). As with the correlation conducted for the second hypothesis, the results from the first regression analysis indicated that a significant relationship did exist between the criterion (5th grade REA subscale) and predictor variables (instructional reading level), $R = .589$, $R^2 = .346$, $F(1, 99) = 652.469$, $p < .001$. That is, 34.6% ($R^2 = .346$) of the variance observed in the criterion variable (REA subscale) was due to the predictor variable (instructional reading level). A high positive trend existed between the 5th grade students' Fountas and Pinnell reading instructional levels and their CRCT reading subscale scores. Displayed in Table 10, are summary statistics of the regression analysis.

Table 10- Summary of Regression Analysis for Hypothesis 3 using 5th Grade REA

Subscale and Instructional Reading Level

Source	R	R^2	Standard Error	F	df1	df2	Sig. (p)
Model 1	0.589	0.346	1.886	52.469	1	99	< .001

Source	Unstandardized Coefficients			Standardized Coefficients			
	B	Std. Error		Beta	T	Sig. (p)	Part Correlation
(Constant)	-26.538	6.699			-3.961	< .001	
Reading Instructional Level	0.058	0.008		0.589	7.244	< .001	0.589

Note. Criterion variable = REA subscale. Total N = 101

Results from the second regression analysis indicated that a significant

relationship did exist between the criterion (5th grade ELA subscale) and predictor

variables (instructional reading level), R = .640, R^2 = .409, $F(1, 108)$ = 74.788, p < .001.

That is, 40.9% (R^2 = .409) of the variance observed in the criterion variable (5th grade

ELA subscale) was due to the predictor variable (instructional reading level). Thus, since

significant relationships were found in both regression analyses, the null hypothesis for

research question 3 was rejected in favor of the alternative hypothesis. The alternative

hypothesis suggests that there is a significant relationship between the spring Fountas and

Pinnell Benchmark Assessment Instructional Reading Levels and the reading

comprehension scores on the Criterion Reference Competency Test (REA and ELA

subscale scores) in grade 5 in the year 2013-2014. Displayed in Table 11, are summary

statistics of the regression analysis. Figures1 and 2 in Appendix D is a scatterplot of

participants' scores on the criterion (REA and ELA subscales) and predictor variables by grade levels.

Table 11- Summary of Regression Analysis for Hypothesis 3 using 5th Grade ELA Subscale and Instructional Reading Level

Source	R	R^2	Standard Error	F	df1	df2	Sig. (p)
Model 1	0.631	0.980	1.811	65.379	1	99	$< .001$

Source	Unstandardized Coefficients			Standardized Coefficients			
	B	Std. Error		Beta	T	Sig. (p)	Part Correlation
(Constant)	-38.188	7.442			-5.131	$< .001$	
Reading Instructional Level	0.071	0.009		0.631	8.086	$< .001$	0.631

Note. Criterion variable = ELA subscale. Total $N = 101$

Summary

The findings in this study addressed the research question that was explored to determine if a significant relationship between the spring Fountas and Pinnell Benchmark Assessment Instructional Reading Levels and reading comprehension scores on the Criterion Reference Competency Test (REA and ELA subscale scores) in grade 3, 4, and 5 in the year 2013-2014. Results from the regression analysis indicated that scores from the two assessments, the Fountas and Pinnell and Reading CRCT, were significantly correlated. This means that as scores on the reading comprehension assessment increased, instructional reading level scores also increased.

Table 12- Summary of Results for Hypothesis 1-3

Hypothesis	Criterion Variable	Predictor Variable	Analysis	Sig. (p)
1	3rd Grade CRCT Reading Comprehension	3rd Grade Instructional Reading Level	Regression	$< .001$
2	4th Grade CRCT Reading Comprehension	4th Grade Instructional Reading Level	Regression	$< .001$
3	5th Grade CRCT Reading Comprehension	5th Grade Instructional Reading Level	Regression	$< .001$

Note. 3rd grade total $N = 102$, 4th grade total $N = 110$, and 5th grade total $N = 101$

In Section 5, I will present a discussion on the findings of the research study. A detailed discussion of the implications for social change, recommendation for actions, and recommendations for further study will also be presented. A quantitative, nonexperimental ex post facto, correlational research design was used to test the single hypothesis, which examines the relationship between Fountas and Pinnell Instructional levels and CRCT Reading scores. Archival test data was transferred to a Statistical Package for the Social Sciences (SPSS) software where it was organized by grade level. Correlations and other statistical analysis were conducted to determine if a relationship existed between the scores on both reading assessments. The results indicated that there was a strong correlation among Fountas and Pinnell instructional levels and CRCT reading scores. The following research question drove this study: Is there a significant relationship between 3rd, 4th, and 5th grade students Fountas and Pinnell instructional levels and their Reading CRCT scores? Data provide evidence that there is a correlation between Fountas and Pinnell instructional levels and reading CRCT scores.

Section 5: Discussion, Conclusion, and Recommendations

Introduction

The purpose of this research study was to determine whether a correlation existed amongst student Fountas and Pinnell instructional levels and reading comprehension scores on the Criterion Reference Competency Test (CRCT). In this section, I present a summary of the results and conclusions, implications for social change, and recommendation for actions and further study, based on the findings of the study. Most of the previous research identifies a correlation between benchmark assessments and student academic achievement (Li et al., 2010; Moss & Brookhart, 2009). The link between reading benchmark assessments and reading state standardized assessments has been well documented; the link between the Fountas Pinnell benchmark assessment and the Reading CRCT has not been fully represented in previous literature. This study focused on that deficiency. The research took place at an elementary school located in a suburb of Atlanta, Georgia. The population consisted of 329 elementary students (grades 3, 4, and 5) with complete Fountas and Pinnell and CRCT reading data sets. Archival test data was obtained and analyzed using the Statistical Package for the Social Sciences (SPSS) software. The research questions addressed by this study were the following:

RQ1: What is the relationship between the third grade spring Fountas and Pinnell Reading Instructional Levels and the third grade CRCT Reading Comprehension Scores?

RQ2: What is the relationship between the fourth grade spring Fountas and Pinnell Reading Instructional Levels and the fourth grade CRCT Reading Comprehension Scores?

RQ3: What is the relationship between the fifth grade spring Fountas and Pinnell Reading Instructional Levels and the fifth grade CRCT Reading Comprehension Scores?

The data analysis revealed that there were strong relationships between Fountas and Pinnell instructional levels and Reading CRCT scores. A conclusion drawn from this study was that a person with higher ability should have a greater probability of success on assessment items than a person with lower ability. These results provide support for the Item Response Theory, which indicates that students with higher reading abilities should have a greater probability of success on the Fountas and Pinnell and CRCT assessment items. The results also support another fundamental principle of the Item Response Theory which assumes that every examinee has some true location on the continuum of the specific latent trait. It is this location that influences the examinees response to any item(s) on a test or survey (Hambleton, Swaminathan, & Rogers, 1991). Despite the fact that measurement tools may vary for the same latent trait, the examinee's location on the continuum should be consistent across test formats. Thus, the results support the principle that student achievement on the Fountas and Pinnell can be linked to achievement on the CRCT by examining reading abilities.

Interpretation of Findings

This study included the analysis of spring 2013-2014 Fountas and Pinnell reading

levels and Reading CRCT scores of 329 third, fourth, and fifth grade students. Findings

from this study support the idea that Fountas and Pinnell reading levels are positively

related to Reading CRCT scores. My attempt to show that no variation exists between

variables failed. Instead, this study indicates that there was a statistically significant

correlation between Fountas and Pinnell instructional levels and reading CRCT scores.

Pearson correlations showed that success on the reading CRCT is related to success on

the Fountas and Pinnell benchmark assessment. Positive beta scores confirmed that as

scores on the reading comprehension assessment increased, instructional reading level

also increased. The current study found that third grade spring Fountas and Pinnell

instructional levels related to spring reading CRCT scores $R = .703$, $R^2 = .494$, $F(1, 100)$

$= 97.510$, $p < .001$. That is, 49.4% ($R^2 = .494$) of the variance observed in the criterion

variable (REA) was due to the predictor variable (instructional reading level). Likewise,

the fourth grade spring Fountas and Pinnell instructional levels related to the fourth grade

spring reading CRCT $R = .607$, $R^2 = .368$, $F(1, 108) = 62.918$, $p < .001$. That is, 36.8%

($R^2 = .368$) of the variance observed in the criterion variable (REA) was due to the

predictor variable (instructional reading level). Although not as strong, the fifth grade

spring Fountas and Pinnell instructional levels related to the fifth grade spring reading

CRCT scores $R = .589$, $R^2 = .346$, $F(1, 99) = 652.469$, $p < .001$. That is, 34.6% ($R^2 =$

.346) of the variance observed in the criterion variable (REA subscale) was due to the predictor variable (instructional reading level).

Framed by the ideas of item response theory, a person with higher ability should have a greater probability of success on assessment items than a person with lower ability (Hambleton, Swaminathan, & Rogers, 1991). The results of this study support the notion that students with higher reading abilities should have a greater probability of success on the Fountas and Pinnell and CRCT assessments. It also supports the notion that students who lack prior knowledge to perform well on prerequisite tests, such as the Fountas and Pinnell benchmark test, may not do well on subsequent tests, such as the reading CRCT. Assessment data has become a critical element in determining students' knowledge and progress. Brady (2011) stated that remediation and early intervention can reduce the incidence of reading failure. Education systems are using benchmark assessments to collect valuable information to assist them in making instructional decisions that improve learning for all students prior to the high-stakes state test.

In the current study the statistical analyses indicated that Fountas and Pinnell reading instructional levels are significant indicators of scores on the CRCT reading comprehension test. The relationships between Fountas and Pinnell instructional levels and reading comprehension scores tended to be positive for all grade levels, suggesting the higher the Fountas and Pinnell instructional level, the more likely students were to perform well on the reading CRCT. Similar to the findings of past research, this study established a relationship between reading benchmark scores and state reading test scores

(Alonzo, Nese, Park, & Tindal, 2011, Forster & Souvignier, 2011). Unlike previous studies, this research initiative focused on reading achievement measured with the Fountas and Pinnell Benchmark Assessment and the Criterion Reference Competency Reading Test. In this study, a relationship emerged between instructional reading levels and CRCT reading comprehension scores. The results of this study suggest that utilizing benchmark tests as screeners helps educators identify students who may need additional or differentiated instruction (Nese, Park, Alonzo, & Tindal, 2011). According to Bancroft (2010), regularly scheduled benchmark assessments are "utilized as a means to have greater surveillance of teaching and learning, with the primary goal of closing achievement gaps" (p.59).

In general, the findings from this study were consistent with previous research comparing reading benchmark test with statewide reading competency tests. Previous research has indicated that benchmark assessments have the potential to be used to predict students' scores on state tests (Merino & Beckman, 2010). As stated in the review of literature, there is research that explores the comparability of scores from benchmark tests and state summative tests. According to Keller-Margulis, Shapiro, and Hintze (2008), benchmark and state tests are related. The researchers investigated the relationship between benchmark assessments and a statewide achievement test. Results suggested that the benchmark data were moderately correlated to statewide achievement data. This study also supports the notion that benchmark assessment data is correlated to state assessment data. Pearce and Gayle (2009) added to the understanding of the

relationship between benchmark assessments and state test scores. Their study determined that the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency scores could predict subsequent reading scores on the Dakota State Test of Educational Proficiency. Pearce and Gayle found that the benchmark assessment predicted comprehension on the state test. It specifically indicated students who might score proficient, opposed to students who would fail. These findings demonstrate technical soundness for using benchmark assessment data to monitor student growth in reading achievement. Other researchers have also suggested that benchmark assessments and high-stakes testing scores were linked (Hefflin, 2009). Similar to Gebhardt & Shapiro, 2012 study, the findings of this study demonstrated that students who scored higher on the benchmark assessments scored higher on the standardized assessment. The findings of this study indicate that the Fountas and Pinnell Benchmark instructional levels are linked to the CRCT reading comprehension scores.

Implications for Social Change

The current study addresses the field of research that explores the manner in which educators utilize assessment data. The results of this study demonstrate that Fountas and Pinnell instructional levels are positively related to the CRCT reading comprehension scores. This research study will benefit students, teachers, and school systems as it provides evidence that intervention is necessary for students that perform poorly on the Fountas and Pinnel benchmark assessment. Many other positive social change effects could result from establishing a relationship between Fountas and Pinnell

instructional levels and CRCT reading comprehension scores. Ultimately, the results of

this study demonstrate that it is logical to assume that students who fail to succeed on

benchmark assessments will fail to succeed on end of the year state tests. The results of

this study also indicate that the Fountas and Pinnell instructional levels are related to the

Reading CRCT scores. This study provides evidence that benchmark assessments, such

as the Fountas and Pinnell, can be reliable tools for making instructional decisions and

has the potential to improve student learning (Davidson & Frohbieter, 2011). Findings

from this research may improve instructional practices, and assist in the development of

progress monitoring tools to promote student success. Educators can now modify their

instructional practice to improve the teaching and learning cycle. Benchmark assessments

can be incorporated as formative assessments and used to monitor students' progress

toward meeting standards. The data from benchmark assessments, such as the Fountas

and Pinnell, can be used in an ongoing manner to improve student performance on

accountability measures. Teachers have an instrument that can help them to know in

advance how their students are doing in relation to the state test. This information may

help them to maximize student success on standardized tests. Lam (2013) research

indicates that benchmark formative assessment strategies have the potential for actively

involving students in the assessment process, hence motivating them toward mastery of

expected content. Furthermore, the vast amount of money and time invested in the

implementation of the Fountas and Pinnell assessment is justified. The results of the

study indicate that Fountas and Pinnell is an assessment tool that provides an accurate

view of reading skills in students. Data collected from the assessment may assist

educators in more accurately identifying literacy skills lacking in the primary grades. The

benchmark assessments provide evidence of student learning. Teachers use the

benchmark data to diagnose individual learning problems and adjust instructional

strategies to meet the needs of the students. The significant findings of this study may

also lead to positive social change by providing support to the manner in which educators

utilize assessment data.

Recommendations for Actions

The findings of this study indicate that a relationship exists between the Fountas

and Pinnell Benchmark instructional levels and the Reading CRCT scores. This validates

the continued use of the Fountas and Pinnell assessment as a progress monitoring tool.

Educators give various benchmark assessments in an effort to collect data. The data is

used for various purposes such as to drive instruction, monitor students' progress, to

predict what level a student is performing at, and to make other educational decisions

(Merino & Beckman, 2010). Once a relationship has been established between a

benchmark test and performance on a state test, benchmarks results can be used to predict

how well a given student might perform on a state test. Li et al. (2010) states that the

primary purpose of benchmark assessments is to identify the instructional needs of

students so that teachers can make informed instructional decisions. The goal is to be able

to judge events for indications of other events, so that schools can prepare in all cases for

the coming of what is anticipated. Reading benchmark assessments provide teachers with

ongoing information throughout the school year regarding student progress in reading. Given the results of this study, the findings should be disseminated to students, parents, educational practitioners, and other stakeholders that might benefit from it. The research results could be used by the superintendent, principal, and educators who need to obtain a better understanding of the benefits of utilizing the Fountas and Pinnell benchmark data to drive instruction. The results will be shared with the administrators of the participating school in the form of a written summary. After discussing the results with administrators, I will request to share the findings with school personnel. If permission is granted, I will present findings in the form of a power point during a staff meeting or Professional Learning sessions. Classroom teachers that use the Fountas and Pinnell benchmark assessments to progress monitor students reading proficiency should be aware that the instructional levels were significantly linked to the CRCT Reading scores.

Recommendations for Further Action

There is a need for further research regarding the relationship between the Fountas and Pinnell instructional levels and state assessments. The first recommendation I would suggest would be to replicate this study using the new Georgia assessment tool, the Georgia Milestones assessment. Since the CRCT reading assessment has been replaced with the Georgia Milestone assessment, it would be interesting to see if a similar relationship exists. Although readiness indicators were establish to signal where students were relative to performance on the Georgia Milestones, I would suggest this study be replicated using the new assessment. I would also suggest that this study be investigated

at a different school. The findings may differ when applied to different parts of the county. Demographics and statistical differences could influence future results. While the research is specific to this school, it may provide a basis for other similar schools to investigate. Additionally, researchers should consider a qualitative study on the topic wherein experiences of lower performing students are studied in regards to their deficits in reading. The present study focused on quantitative data related to benchmark testing practices. Future research could be used to examine students', teachers', parents', curriculum directors' and administrators' perceptions of the Fountas and Pinnell Benchmark Assessment and its relationship to the state summative assessment. Interviewing these stakeholders about their thoughts related to benchmark testing could provide valuable information to the county. It may also be helpful to examine how benchmark test scores are used by teachers and administrators. A research study examining how the teachers and administrators have used the Fountas and Pinnell Benchmark data to inform instructional practices could provide additional insight into the relationship between the Fountas and Pinnell instructional levels and CRCT reading comprehension scores.

Conclusion

The overall purpose of this study was to determine if a relationship existed between Fountas and Pinnell reading instructional levels and the reading CRCT scores. The ability to read proficiently is critical to a child's success for success in all academic areas and in life. Many students struggle to learn to read proficiently in elementary

school, which puts them at risk for dropping out of high school later. The data in this non experimental ex post facto study strongly suggested that there was a significant relationship between the third, fourth, and fifth grade Fountas and Pinnell instructional levels and Reading CRCT scores. As a result, students who fail to meet appropriate instructional level at specific points in time are in need of interventions aimed at improving their preparation for success on the Georgia Milestone. Results from this study support previous research findings that reading benchmark assessments correlate with state reading tests.

References

Airasian, P. W. (1994). *Classroom assessment*. New York: McGraw-Hill.

Alonzo, J., Nese, J., Park, B., & Tindal, G. (2011). Applied curriculum-based

measurement as a predictor of high-stakes assessment. *Elementary School

Journal, 111*, 608-624. doi:10.1086/659034

Amrein-Beardsley, A. (2009). This is Jeopardy!. *Education Digest: Essential Readings

Condensed for Quick Review, 74*(5), 14-18.

Anderson, S., Leithwood, K., & Strauss, T. (2010). Leading data use in schools:

Organizational conditions and practices at the school and district

levels. *Leadership and Policy in Schools*, *9*(3), 292-327.

Anfara, V. A., & Donhost, M. J. (2010). What Research Says: Data-Driven Decision

Making. *Middle School Journal*, 56-63. Retrieved from

http://ehis.ebscohost.com.ezp.waldenulibrary.org/ehost/pdfviewer/pdfviewer?sid

=0bf3b7bc-5c03-41e4-9b8b- f93d172bb622%40sessionmgr104&vid=7&hid=8

Au, W. (2009). Social studies, social justice: W(h)ither the social studies in high-stakes

testing? *Teacher Education Quarterly*, 43-58. Retrieved January 1, 2015, from

http://files.eric.ed.gov/fulltext/EJ851027.pdf

Au. W., (2011). Teaching under the new Taylorism: High-stakes testing and the

standardization of the 21st century curriculum. *Journal of Curriculum Studies,

43*(1), 25-45.

Baker, M., & Johnston, P., (2010). The impact of socioeconomic status on high stakes

testing reexamined. *Journal of Instructional Psychology, 37*(3). 193-199.

Baker, F. B. (2001). The Basics of Item Response Theory. USA: ERIC Clearinghouse on

    Assessment and Evaluation.

Bancroft, K. (2010). Implementing the mandate: The limitations of benchmark tests.

    Educational Assessment, Evaluation, and Accountability, 22(1): 53-72.

Bambrick-Santoyo, P. (2010). Driven by data: A practical guide to improve instruction.

    San Francisco, CA: Jossey-Bass.

Barge, J. (2012, March). Georgia's college and career ready performance index (CCRPI)

    and ESEA flexibility. Retrieved January 1, 2015, from

    http://www.gadoe.org/Curriculum-Instruction-and-

    Assessment/Accountability/Documents/CCRPI%20for%20Data%20Collections%

    20Conference%2008.21.13.pdf

Begeny, J.C., Krouse, H., E., Ross, S.G., & Mitchell, R.C. (2009). Increasing elementary-

    aged students 'reading fluency with small-group interventions: A comparison of

    repeated reading, listening passage preview, and listening only strategies. *Journal*

    of Behavioral Education, 3, 211-228.

Behrent, M. (2009). Reclaiming our freedom to teach: Education reform in the Obama

    era. *Harvard Educational Review*, *79*(2), 240-246.

Bernhardt, V.L. (2009). Data use: Data-driven decision making takes a big-picture vies of

    the needs of teachers and students. *Journal of Staff Development, 30*(1), 24-27.

Berns, B. B., & Sandler, J. O. (Eds.). (2009). *Making science curriculum matter: Wisdom for the reform road ahead*. Thousand Oaks, CA: Corwin Press.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)*, *21*(1), 5-31.

Borja, R. R. (2006). Voices of experience: District leaders developing school data initiatives can learn from those who have gone before. *Education Week*, *25*(35), 35.

Brady, S. (2011). Efficacy of phonics teaching for reading outcomes: Indicators from post-NRP research. In S. A. Brady, D. Braze, & C. A. Fowler (Eds.), *Explaining individual differences in reading: Theory and evidence* (pp.69-96). New York, NY: Psychology Press.

Brown, R. S., & Coughlin, E. (2007). The Predictive Validity of Selected Benchmark Assessments Used in the Mid-Atlantic Region. Issues & Answers. REL 2007-No. 017. *Regional Educational Laboratory Mid-Atlantic*.

Brown-Chidsey, R., & Steege, M. W. (2010). *Response to intervention: Principles and strategies for effective practice*. New York: Guilford Press.

Brundage, S. & Hancock, A. (2010). Commentaries. *International Journal of Therapy Rehabilitation, 17*(11), 586.

Bulkley, K., Olah, L. N., & Blanc, S. (Eds.) (2010). Special issue on benchmarks for success? Interim assessments as a strategy for educational improvement. *Peabody Journal of Education, 85*(2).

Burch, P. (2010). The bigger picture: Institutional perspectives on interim assessment technologies. *Peabody Journal of Education*, *85*(2), 147-162.doi: 10.1080/01619561003685288.

Callaghan, G., & Madelaine, A. (2012). Leveling the playing field for kindergarten entry: Research implications for preschool early literacy instruction. *Australasian Journal of Early Childhood, 37*(1), 13-23.

Campbell, C., & Levin, B. (2009). Using data to support educational improvement. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)*, *21*(1), 47-65.doi: 10.1007/s11092-008-9063-x.

Carlson, D., Borman, G. D., & Robinson, M. (2011). A multistate district-level cluster randomized trial of the impact of data-driven reform on reading and mathematics achievement. *Educational Evaluation and Policy Analysis*, *33*(3), 378-398. doi: 10.3102/0162373711412765.

Cartwright, K.B. (2010). Word Callers: Small-Group and One –to-One Interventions for Children who "Read" but Don't Understand. Portsmouth, NH: Heinemann.

Cauley, K. M., & McMillan, J. H. (2009). Formative assessment techniques to support student motivation and achievement. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *83*(1), 1-6. doi: 10.1080/00098650903267784.

CCSS Initiative (2013). About the standards. Retrieved from

http://www.corestandards.org/about-the-standards.

Chappuis, J. (2009). Seven strategies of assessment for learning. Portland, OR:

Educational Testing Service.

Chappuis, S., Chappuis, J., & Stiggins, R. (2009). Supporting teacher learning teams.

Educational Leadership, 66(5), 56-60.

Christ, T. J., Silberglitt, B., Yeo, S., & Cormier, D. (2010). Curriculum-based

measurement of oral reading: An evaluation of growth rates and seasonal effects

among students served in general and special education. *School Psychology

Review*, *39*(3), 447.

Clark, I. (2011). Formative assessment: Policy, perspectives and practice. Florida Journal

of Educational Administration & Policy, 4(2), 158-180.

Clay, M.M. (2011). Becoming literate: The construction of inner control. Portsmouth,

NH: Heinemann.

Clay, M. (2010). How very young children explore writing. Portsmouth, NH:

Heinemann.

Coburn, C. E., & Turner, E. O. (2012). The practice of data use: An introduction.

American Journal of Education, *118*(2), 99-111. doi: 10.1086/663272.

Colburn, A. (2009). The Prepared Practitioner. *The Science Teacher, 76*(2), 10-11.

Cosner, S. (2011). Teacher learning, instructional considerations and principal communication: lessons from a longitudinal study of collaborative data use by teachers. *Educational Management Administration & Leadership*, *39*(5), 568-589.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage.

Cusumano, D. L. (2007). Is It Working?: An Overview of Curriculum Based Measurement and Its Uses for Assessing Instructional, Intervention, or Program Effectiveness. *Behavior Analyst Today*, *8*(1), 24-34.

Datnow, A. (2011). Collaboration and contrived collegiality: Revisiting Hargreaves in the age of accountability. *Journal of Educational Change*, *12*(2), 147-158. doi: 10.1007/s10833-011-9154-1.

Davidson, K. L., & Frohbieter, G. (2011). District Adoption and Implementation of Interim and Benchmark Assessments. CRESST Report 806. *National Center for Research on Evaluation, Standards, and Student Testing (CRESST)*.

Deike, M. A. (2009). *The principal as an instructional leader within the context of effective data use*. The University of Texas at Austin.

Deno, S. (2003). Curriculum-based measures: Development and perspectives. *Assessment for Effective Intervention*, *28*(3-4), 3-12.

Deno, S., Reschly, A.L., Lembke, E, Magnussen, D., Callender, S., Windram, H., & Statchel, N. (2009). A school-wide model for progress monitoring. *Psychology in the Schools*, 46, 44-55.

Dewey, J. (2012). Democracy and Education. New York, NY: Simon & Brown.

Donhost, M. (2010). Data-driven decision making. *Middle School Journal, 41* (3), 56-58. doi: 10.1080/00940771.2010.11461758.

Drew, C.J., Hardman, M.L., & Hosp, J.L. (2008). Designing and conducting research in education. Los Angeles: Sage.

Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research & Evaluation,14*(7), 1-11.

Durkin, D. (1980). Reading comprehension instruction in five basal reader series. Reading Research Quarterly. 14(4). 515-544. doi: 10.2307/747314.

Embretson, S. E. (1996). The new rules of measurement. Psychological Assessment, 8, 341-349.

Faria, A., Heppen, J., Li, Y., Stachel, S., Jones, W., Sawyer, K., Palacios, M (2012). The use of interim assessment data in urban schools: Links among data use practices and student achievement. Washington, DC: American Institutes for Research.

Feng, M., Beck, J. E., Heffernan, N. T., & Koedinger, K. R. (2008). Can an Intelligent Tutoring System Predict Math Proficiency as Well as a Standardized Test?. In *EDM* (pp. 107-116).

Fewster, S., & MacMillan, P.D. (2002). School-based evidence for the validity of curriculum-based measurement of reading and writing. *Remedial and Special Education*, *23*(3), 149-156.

Flowers, N., & Carpenter, D. M. (2009). You don't have to be a statistician to use data: A process for data-based decision making in schools. *Phi Delta Kappan*, *91*(2), 64-67. doi: 10.1177/003172170909100214.

Foley, E., & Sigler, D. (2009). Getting smarter: A framework for districts. *Redesigning the "Central Office"*, 5.

Forster, N., & Souvignier, E. (2011). Curriculum-based measurement: Developing a computer-based assessment instrument for measuring student reading progress on multiple indicators. Learning Disabilities: A Contemporary Journal, 9(2), 65-88.

Fountas, I., & Pinnell, G.S. (2012). The Fountas and Pinnell Benchmark Assessment System, Second Edition. Portsmouth, NH: Heinemann.

Fountas, I., & Pinnell, G.S. (2010). The Fountas and Pinnell Benchmark Assessment System, Second Edition. Portsmouth, NH: Heinemann.

Fountas, I. &. (2010). Assessment Guide: A Guide to the Fountus & Pinnell Benchmark Assessment System 1. Portsmouth, NH: Heinemann.

Frey, B. B., & Schmitt, V. L. (2007). Coming to terms with classroom assessment. *Journal of Advanced Academics*, *18*(3), 402-423.

Frohbieter, G., Greenwald, E., Stecher, B., & Schwartz, H. (2011). Knowing and Doing: What Teachers Learn from Formative Assessment and How They Use the Information. CRESST Report 802. *National Center for Research on Evaluation, Standards, and Student Testing (CRESST)*.

Gebhardt, S. N., & Shapiro, E. N. (2012). Comparing computer-adaptive and curriculum-

based measurement methods of assessment. School Psychology Review, 41(4), 295-305.

Georgia Department of Education (GaDOE). (2010). Common core Georgia performance standards. Retrieved from http://www.gadoe.org/CCGPS.aspx

Georgia Department of Education (2009). *Criterion referenced competence tests*. Retrieved from http://www.doe.k12.ga.us/curriculum/testing/crct.asp

Georgia Department of Education. (2012). Georgia Criterion Referenced Competency Tests. Retrieved from http://www.gadoe.org/Curriculum-Instruction-and-Assessment/Assessment/Documents/2702772 sig s12GA%20(2).pdf

Georgia Department of Education. (2013). School Keys: Unlocking excellence through the Georgia school standards. Retrieved from: //www.doe.k12.ga.us/SCHOOL%20KEYS%20FINAL%203-13-12[1].pdf

Godreau Cimma, K. L. (2011). *A Middle School Principal's and Teachers' Perceptions of Leadership Practices in Data-Driven Decision Making*. ProQuest LLC. 789 East Eisenhower Parkway, PO Box 1346, Ann Arbor, MI 48106.

Godwin, B. (2014). Better tests don't guarantee better instruction. Educational Leadership, 71 (6), 78-80.

Gravetter, F., & Wallnau, L. (2008). *Essentials of statistics for the behavioral sciences (6th ed.)*. Belmont, CA: Thomson Wadsworth.

Greenstein, L. (2012). Beyond the Core: Assessing Authentic 21st Century Skills. *Principal Leadership*, *13*(4), 36-42.

Guilfoyle, C. (2006). NCLB Is there life beyond testing Educational Leadership, 64 (3),
8-13.

Halverson, R. (2010). School formative feedback systems. *Peabody Journal of Education*, *85*(2), 130-146. doi: 10.1080/01619561003685270.

Hambleton, R. K. (1989). Principles and selected applications of item response theory. In R. L. Linn (Ed.), Educational measurement (3rd ed.,pp. 147-200). New York: Macmillan.

Hambleton, R.K., & Swaminathan, H., & Rogers, J.H. (1991). Fundamentals of item response theory. New York: Sage publications.

Hamilton, L., Halverson, R., Jackson, S. S., Mandinach, E., Supovitz, J. A., & Wayman, J. C. (2009). Using Student Achievement Data to Support Instructional Decision Making. IES Practice Guide. NCEE 2009-4067. *National Center for Education Evaluation and Regional Assistance*.

Hefflin, P. (2009*). Do benchmark assessments increase student achievement on state standardized tests?* (Doctoral dissertation, Duquesne University, 2009). ProQuest Dissertations and Theses, ID 3371521.

Henry District Public School System (2013). *Superintendent: The District Strategic Plan.* Retrieved from http://www.henry.k12.ga.us.

Hess, F. M. & Mehta, J. (2013). Data: No deus ex machine. *Educational Leadership, 70*(5), 71-75.

Hess, F. M., & Fullerton, J. (2009). The numbers we need: Bringing balanced scorecards to education data. *Phi Delta Kappan, 90*(9), 665–69.

Hill, S., & Launder, N. (2010). Oral language and beginning to read. Australian Journal of Language and Literacy, 33 (3), 240-254.

Hintze, J. M., & Silberglitt, B. (2005). A longitudinal examination of the diagnostic accuracy and predictive validity of R-CBM and high-stakes testing. *School Psychology Review*, *34*(3), 372.

Honawar, V. (2006). Tip of their fingers. Education Week, 25 (35), 38-39.

Howell, D.C. (2011). Fundamental statistics for the behavioral sciences. California: Wadsworth Cengage Learning.

Jackson, S.L. (2009). Research Methods and Statistics: A Critical Thinking Approach 3rd ed. Belmont: CA, Wadsworth.

Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for service delivery in an RTI framework: Candidate measures. *School Psychology Review*, *36*, 582-99.

Jenkins, J.R., Deno, S.L., & Mirkin, P.K. (1979). Pupil progress: Measuring pupil progress toward the least restrictive alternative. *Learning Disability Quarterly, 2*(4), 81-91.

Jennings, J. L. (2012). The effects of accountability system design on teachers' use of test score data. Teachers College Record, 114 (11), 1-23.

Jordan , G.E., Snow, C.E. and Porsche, M.V. (2000). Project EASE: The effect of a

family literacy project on kindergarten students' early literacy skills. Reading

Research Quarterly, 35, 524-546.

Kaminski, R. A., & Good, R. H. (1996). Toward a technology for assessing basic early

literacy skills. School Psychology Review, 25, 215-227.

Keller-Margulis, M. A., Shapiro, E. S., & Hintze, J. M. (2008). Long-Term Diagnostic

Accuracy of Curriculum-Based Measures in Reading and Mathematics. *School*

Psychology Review, *37*(3), 374-390.

Lam, R. (2013). Formative use of summative tests. Use test preparation to promote

performance and self-Regulation. Asia- Pacific Education Researcher, 22(1) 69-

78.

Lashway, L. (1999). Measuring leadership: A guide to assessment for development of

school executives. Retrieved from ERIC database. (ED431209).

Lai, M. K., & McNaughton, S. (2009). Not by Achievement Analysis Alone: How

Inquiry Needs to be Informed by Evidence from Classroom Practices. *New*

Zealand Journal of Educational Studies, *44*(2), 93.

Lai, M. K., McNaughton, S., Timperley, H., & Hsiao, S. (2009). Sustaining continued

acceleration in reading comprehension achievement following an

intervention. *Educational Assessment, Evaluation and Accountability (formerly:*

Journal of Personnel Evaluation in Education), *21*(1), 81-100. doi:

10.1598/RRQ.44.1.2.

Leedy, P.., & Ormrod, J.E. (2010). Practical research: Planning and design (9th ed.). Boston, MA: Pearson.

Levin,B., Datnow, A., & Carrier, N. (2012. Changing school district practices. Boston, MA: Jobs for the Future.

Li, Y., Marion, S., Perie, M., & Gong, B. (2010) An approach for evaluating the technical quality of interim assessments. *Peabody Journal of Education, 85*(2), 163-185. doi:10.1080/01619561003685304.

Lim, S. M. & Roger, S. (2010). The use of interactive formative assessments with first-year occupational therapy students. *International Journal of Therapy and Rehabilitation, 17*(11), 576-586. doi: 10.12968/ijtr.2010.17.11.79538.

Linn, R. L. (1993). Linking results of distinct assessments. Applied Measurement in Education, 6, 83-102.

Lodico, M. G., Spaulding, D. T., & Voegtle, K. H. (2010). *Methods in educational research: From theory to practice* (Vol. 28). John Wiley & Sons.

Lohmeier, J. (2010). Nonexperimental designs. In N. Salkind (Ed.), Encyclopedia of research design. (pp. 911-915). Thousand Oaks, CA: SAGE. Doi: 10.4135/9781412961288.n271.

Lord, F.M., & Wingersky, M.S. (1984). Comparison of IRT true-score and equipercentile observed-score "equating" Applied Psychological Measurement, 8, 453-461.

Lundberg, I. (1991). Phonemic awareness can be developed without reading instruction. In S. A. Brady & D. P. Shankweiler (Eds.), Phonological Processes in Literacy: A Tribute to Isabelle Y. Liberman (pp.47-53). Hillsdale, NJ: Erlbum.

Madaus, G., & Russell, M. (2010/2011). Paradoxes of high-stakes testing. *Journal of Education, 190* (1,2), 21-30.

Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist*, *47*(2), 71-85. doi: 10.1080/00461520.2012.667064.

Mandinach, E.B., Honey, M., Light, D., & Brunner, C. (2008). A conceptual framework for data-driven decision making. In E.B. Mandinach & M. Honey (Eds.), Data-Driven school improvement: Linking data and learning (pp. 13-31). New York, NY: Teachers College Press.

Marshall, K. (2008). Interim assessments: A user's Guide. *Phi Delta Kappan*,*90*(1), 64-68.

McMillan, J. H. (2007). Formative classroom assessment: The key to improving student achievement. In J. H. McMillan (Ed), Formative classroom assessment: Theory into practice (pp. 1-7). New York: Teachers College Press.

Meisinger, E. B., Bradley, B.A., Schwanenflugel, P.J., Kuhn, M.R. & Morris, R.D. (2009). Myth and reality of the word caller: The relation between teacher nominations and prevalence among elementary school children. School Psychology Quarterly, 24 (3), 147-150.

Merino, K., & Beckman, T. (2010). Using reading curriculum-based measurements as predictors for the measure academic progress map standardized test in Nebraska. *International Journal of Psychology: A Bio psychosocial Approach, 6*, 85-98.

Merriam, S. B. (1998). *Qualitative research and case study applications in education.* San Francisco: Jossey-Bass.

McKenna, M. & Stahl. S. (2003). Assessment for reading instructioin. NY: Guilford Press.

Moore, B. (2011). *I have the data…now what: Analyzing data and making instructional changes.* Larchmont, NY: Eye on Education.

Moss, C. M., & Brookhart, S. M. (2009). *Advancing formative assessment in every classroom: A guide for the instructional leader.* Alexandria, VA: ASCD (Association for Supervision and Curriculum Development).

National Center for Educational Statistics (NCES). (2013). Public school graduates and dropouts from the common core of data. Retrieved from http://nces.ed.gov/pubs2013/2013309.pdf

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), 14-23.

No Child Left Behind Act of 2001, P.L. 107-110,§ 115, Stat 1425 (2002).

Neutens, J.J., & Rubinson, L. (2010). Research techniques for the health sciences, (4[th]ed). San Francisco: Benjamin Cummings.

Oberg, C. (2009). Guiding classroom instruction through performance assessment. *Journal of Case Studies in Accreditation and Assessment*, *1*, 1-11.

Olah, L. N., Lawrence, N. R., & Riggan, M. (2010). Learning to learn from benchmark assessment data: How teachers analyze results. *Peabody Journal of Education*, *85*(2), 226-245.

Pearce, L.R., & Gayle, R. (2009). Oral reading fluency as a predictor of reading comprehension with Amerian Indian and White elementary students. School Psychology Review, 38(3), 419-427.

Popham, W. J. (2006). All about accountability/phony formative assessments: Buyer beware. *Educational Leadership*, *64*(3), 86-87.

Popham, W. J. (2008). Transformative Assessment. Alexandria, VA: Association for supervision and Curriculum Development.

Popham, W. J. (2009). Anchoring down the data. Educational Leadership, 66(4), 85-86.

Popham, W.J. (2011). Formative assessment –A process, not a test. Education Week.

Powell, K.& Kalina, C. (2009). Cognitive and social constructivism: Developing tools for an effective classroom. Education, 130(2), 241-250.

Price, J., & Kortez, D.M. (2005). *Building assessment literacy. In K.P. Boudett, E.A. City, & R.J.Murnane (Eds.). Data wise: A step-by-step guide to using assessment results to improve teaching and learning.* Cambridge, MA: Harvard Education Press.

Ravitch, D. (2010). *The death and life of the great American school system: How testing and choice are undermining education*. New York, NY: Basic Books.

Roberts, G., Torgesen, J.K., Boardman, A., & Scammacca, N. (2008). Evidence-based strategies for reading instruction of older students with learning disabilities. Learning Disabilities Research & Practice, 23, 63-39.

Roehrig, A. D., Petscher, Y., Nettles, S. M., Hudson, R. F., & Torgesen, J. K. (2007). Accuracy of the DIBELS oral reading fluency measure for predicting third grade reading comprehension outcomes. *Journal of School Psychology*,*46*(3), 343-366. doi: 10.1016/j.jsp.2007.06.006.

Rothstein, R (2009). The prospects for no child left behind (Memorandum No. 149). Retrieved from http://www.epi.org/publications/entry/pm149.

Salkind, N.J. (2008). *Statistics for people who think they hate statistics (3rd ed).* Thousand Oaks, CA: Sage.

Sawchuk, S. (2009). Testing Faces Ups and Downs Amid Recession. *Education Week*, *28*(31), 1-16.

Sawchuk, S., & Cain, J. (2009). Testing faces ups and downs amid recession. Education Week, 28 (31), 1-17.

Schelppenbach, M. (2010). How can teachers and schools use data effectively? Retrieved from National Council of Teachers of Mathematics website: http://www.nctm.org/uploadedFiles/Research_News_and_Advocacy/Research/Clips_and_Briefs/Research_brief_16-Schelppenbach(1).pdf.

Schildkamp, K., & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching & Teacher Education, 26* (3), 482-496.

Sharratt, L., & Fullan, M. (2012). *Putting FACES on the Data: What Great Leaders Do!*. SAGE.

Shepard, L. A. (2010). What the marketplace has brought us: Item-by-item teaching with little instructional insight. *Peabody Journal of Education*, *85*(2), 246-257. Doi: 10.1080/0161956100370844.

Shepard, L.A. (2008). *Formative classroom assessment: Caveat emptor. In C. A. Dwyer (Ed.), The future of assessment: Shaping teaching and learning)pp. 279-303).* New York: Lawrence Erlbaum Associates.

Simmons, W. (2012). Data as a level for improving instruction and student achievement. *Teachers College Record*, 114(11).

Slavin, R. E., Holmes, G., Madden, N. A., Chamberlain, A., & Cheung, A. (2011). Effects of a Data-Driven District Reform Model on State Assessment Outcomes. Baltimore: Success for All Foundation.

Smith, R., Johnson, M., & Thompson, K. D. (2012). Data, our GPS. *Educational Leadership, 69*(5), 56-59.

Stage, S. A., & Jacobsen, M. D. (2001). Predicting student success on state-mandated performance-based assessment using oral reading fluency. *School Psychology Review*, *30*, 407-419.

Stanovich, K.E., & Stanovich, P.J. (1995). How research might inform the debate about early reading acquisition. Journal of Research in Reading, 18, 87-105. doi: 10.1111/j.1467-9817.1995.tb00075.x.

Stocking, M.L. & Lord, F.M. (1983). Developing a Common Metric in Item Response Theory. Applied Psychological Measurement, 7(2), 201-210.

Supovitz, J., & Taylor, B. S. (2003). The impact of standards-based reform in Duval District, Florida: 1999-2002. *Philadelphia, PA: Consortium for Policy Research in Education*.

Tabachnick, B., C. & Fidell, L., S. (2007). *Using Multivariate Statistics* (5th edition). Boston, MA: Pearson.

Tierney, R. D. (2006). Changing practices: Influences on classroom assessment. *Assessment in Education, 13*(3), 239-264.

Togneri, W., & Anderson, S. E. (2003). *Beyond islands of excellence: What districts can do to improve instruction and achievement in all schools: A leadership brief*. Learning First Alliance.

Torgesen, J. K., & Miller, D. H. (2009). Assessments to guide adolescent literacy instruction. *Portsmouth, NH: Center on Instruction at RMC Research Corporation*.

Torgesen, J.K. (2009). The response to intervention instructional model: Some outcomes from a largescale implementation in Reading First schools. Child Development Perspectives, 3, 38-40.

Turner, S. L. (2009). Ethical and Appropriate High-Stakes Test Preparation in Middle School: Five Methods that Matter. *Middle School Journal (J3)*, *41*(1), 36-45. doi: 10.1080/00940771.2009.11461702.

Upadhyay, B. (2009). Negotiating identity and science teaching in a high-stakes testing environment: an elementary teacher's perceptions. *Cultural Studies of Science Education*, *4*(3), 569-586.

Viadero, D. (2009). NCES finds states lowered 'Proficiency' bar. Education Week, Retrieved from

http://www.edweek.org/ew/articles/2009/10/29/10nces.h29.html?tkn

Ward-Roberts, V. (2009). *Does data-driven decision making matter for African American students?* University of Southern California.

Wayman, J. C., Cho, V., & Shaw, S. M. (2009). First-year results from an efficacy study of the Acuity data system. *Austin, TX: The University of Texas*.

Weiss, J.A. (2012). Data for improvement, data for accountability. Teachers College Record, 114(11).

Wilhelm, T. (2011). A team approach to using student data. *Leadership, 40*(5), 26-30.

Wilkins, E. A., & Shin, E. K. (2011). Peer Feedback: Who, What, When, Why, and How. *Education Digest: Essential Readings Condensed for Quick Review*,*76*(6), 49-53.

Wright, J. M. (2012). *The Relationship Between Academic Placement and High School Academic Failure in Mathematics* (Doctoral dissertation, WALDEN UNIVERSITY).

Wu, P. (2009). *Do we really understand what we are talking about? A study examining the data literacy capacities and needs of school leaders*. University of Southern California.

Young, V. M., & Kim, D. H. (2010). Using Assessments for Instructional Improvement: A Literature Review. *Education policy analysis archives*, *18*(19), n19.

Zehr, M. A. (2006). Monthly checkups. *Education Week*, *25*(35), 36-37.

Appendix A: Letter to Board of Education & Superintendent

XXXXX District Board of Education
33 North Zack Hinton Parkway
McDonough, Georgia 30253
XXXXX District Board of Education & Superintendent:

I am a student at Walden University working on my doctoral degree in Administrator Leadership for Teaching and Learning. I am conducting a study, "Comparing Fountas and Pinnell Reading Instructional Levels to Reading Scores on the Criterion Reference Competency Test." I would like permission to conduct this study in your school district.

In response to the challenge of improving student's performance in reading I have developed an interest in understanding the role that benchmark assessments play in this process. I am specifically interested in determining if the Fountas and Pinnell Reading Instructional Levels and the Reading scores on the Criterion Reference Competency Test (CRCT)are related. If a relationship exists I would like to determine the predictive validity of this assessment. The goal is to understand and validate the potential use of the Fountas and Pinnell Reading Benchmark Assessment.

The purpose of this quantitative nonexperimental study is to examine whether a spastically significant correlation exists amongst achievement scores on the Fountas and Pinnell Reading Assessment and reading comprehension scores on the Criterion Reference Competency Test. Identifying a relationship between these two scores can help determine whether data gathered from the Fountas and Pinnell Assessment can serve as a tool for predicting students success or failure on the reading part of the CRCT. The research question to be answered is: Is there a relationship between student achievement scores on the Fountas and Pinnell Reading Benchmark Assessment and student achievement scores on the reading portion of the CRCT?

I am requesting your permission to collect Fountas and Pinnell and Reading CRCT 2013-2014 archival data. To protect students' rights of privacy, data will be stored in a password protected computer in the school grounds in researcher's classroom. The data will remain there for approximately five years. To further protect students' names, responses will be assigned numerical codes. I will request that all student identifiers be removed prior to submitting any data for review during research activities. No real names will be used during the research activities to ensure confidentiality. All criteria for Instructional Review Board (IRB) approval will be adhered to for this study.

If there are questions or concerns, please contact Shunda Walker, Ed.D. Candidate and. Ed.D. Dr. Jose Ataola, PhD mentor, Walden University. I will submit a summary of this study to the school district upon completion.

Sincerely,

Shunda Walker

Signature of Participant

_____

Appendix B: Letter of Approval To Conduct Study from the Board of Education

H HENRY
COUNTY
SCHOOLS

September 22, 2015

Shunda Walker
684 Burtons Cove
Hampton, GA 30228

Dear Ms. Walker:

Your request to conduct research in our school system as part of a requirement to obtain your doctoral degree from **Walden University** has been reviewed. Specifically, consideration was given to the description of your research project, proposed data collection procedures, instruments and research timeline.

It is my understanding that you plan to examine the **Correlation between Fountas and Pinnell and Criterion Referenced Competency Reading Comprehension Scores.** Please note that participation in this study is completely voluntary and that all information obtained for this study must be completely confidential. Although you have been approved to conduct research in our district, please note that students, faculty and staff are not required to participate in your study. To preserve the privacy of student and staff information, pseudonyms for teachers, students, schools, and this system must be used in all written reports. This data must be used solely for the purpose articulated in the research application. Please note that our district does not provide contact information for participants you want to include in your sample. You might obtain this information from the school principal, should she/he be willing to participate or from the individual websites.

After considering all of the information submitted, it appears that your research request meets the requirements of Henry County Board of Education policy KEBA, Solicitation of Information. I am, therefore, approving your request to conduct the research in our school system as described in your proposal. I hope that your research project goes well and that the information you obtain will be beneficial to you and the students of Henry County Schools.

Sincerely,

Rodney Bowler,
Superintendent

Copy: Lisa Orr

Appendix C: Data Use Agreement Form

DATA USE AGREEMENT

This Data Use Agreement ("Agreement"), effective as of 10/11/15 ("Effective Date"), is entered into by and between Shunda Walker ("Data Recipient") and Rocky Creek Elementary School ("Data Provider"). The purpose of this Agreement is to provide Data Recipient with access to a Limited Data Set ("LDS") for use in research **in accord with laws and regulations of the governing bodies associated with the Data Provider, Data Recipient, and Data Recipient's educational program.** In the case of a discrepancy among laws, the agreement shall follow whichever law is more strict.

1. <u>Definitions.</u> Due to the study's affiliation with Laureate, a USA-based company, unless otherwise specified in this Agreement, all capitalized terms used in this Agreement not otherwise defined have the meaning established for purposes of the USA "HIPAA Regulations" and/or "FERPA Regulations" codified in the United States Code of Federal Regulations, as amended from time to time.

2. <u>Preparation of the LDS.</u> Data Provider shall prepare and furnish to Data Recipient a LDS in accord with any applicable laws and regulations of the governing bodies associated with the Data Provider, Data Recipient, and Data Recipient's educational program.

3. <u>Data Fields in the LDS.</u> **No direct identifiers such as names may be included in the Limited Data Set (LDS).** In preparing the LDS, Data Provider shall include the **data fields specified as follows,** which are the minimum necessary to accomplish the research: 2013 Fountas and Pinnell Instructional Levels and 2013 Reading CRCT scores for third, fourth, and fifth grade students.

4. <u>Responsibilities of Data Recipient.</u> Data Recipient agrees to:

a. Use or disclose the LDS only as permitted by this Agreement or as required by law;

b. Use appropriate safeguards to prevent use or disclosure of the LDS other than as permitted by this Agreement or required by law;

c. Report to Data Provider any use or disclosure of the LDS of which it becomes aware that is not permitted by this Agreement or required by law;

d. Require any of its subcontractors or agents that receive or have access to the LDS to agree to the same restrictions and conditions on the use and/or disclosure of the LDS that apply to Data Recipient under this Agreement; and

e. Not use the information in the LDS to identify or contact the individuals who are data subjects.

5.   <u>Permitted Uses and Disclosures of the LDS.</u> Data Recipient may use and/or disclose the LDS **for its Research activities only.**

6.   <u>Term and Termination.</u>

a.   <u>Term.</u> The term of this Agreement shall commence as of the Effective Date and shall continue for so long as Data Recipient retains the LDS, unless sooner terminated as set forth in this Agreement.

b.   <u>Termination by Data Recipient.</u> Data Recipient may terminate this agreement at any time by notifying the Data Provider and returning or destroying the LDS.

c.   <u>Termination by Data Provider.</u> Data Provider may terminate this agreement at any time by providing thirty (30) days prior written notice to Data Recipient.

d.   <u>For Breach.</u> Data Provider shall provide written notice to Data Recipient within ten (10) days of any determination that Data Recipient has breached a material term of this Agreement. Data Provider shall afford Data Recipient an opportunity to cure said alleged material breach upon mutually agreeable terms. Failure to agree on mutually agreeable terms for cure within thirty (30) days shall be grounds for the immediate termination of this Agreement by Data Provider.

e.   <u>Effect of Termination.</u> Sections 1, 4, 5, 6(e) and 7 of this Agreement shall survive any termination of this Agreement under subsections c or d.

7.   <u>Miscellaneous.</u>

a.   <u>Change in Law.</u> The parties agree to negotiate in good faith to amend this Agreement to comport with changes in federal law that materially alter either or both parties' obligations under this Agreement. Provided however, that if the parties are unable to agree to mutually acceptable amendment(s) by the compliance date of the change in applicable law or regulations, either Party may terminate this Agreement as provided in section 6.

b.   <u>Construction of Terms.</u> The terms of this Agreement shall be construed to give effect to applicable federal interpretative guidance regarding the HIPAA Regulations.

c.   <u>No Third Party Beneficiaries.</u> Nothing in this Agreement shall confer upon any person other than the parties and their respective successors or assigns, any rights, remedies, obligations, or liabilities whatsoever.

d.   <u>Counterparts.</u> This Agreement may be executed in one or more counterparts, each of which shall be deemed an original, but all of which together shall constitute one and the same instrument.

e.　Headings. The headings and other captions in this Agreement are for convenience and reference only and shall not be used in interpreting, construing or enforcing any of the provisions of this Agreement.

IN WITNESS WHEREOF, each of the undersigned has caused this Agreement to be duly executed in its name and on its behalf.

DATA PROVIDER

Signed: _Jay Fowler_

Print Name: _Jay Fowler_

Print Title: _Principal_

DATA RECIPIENT

Signed: _Shunda Walker_

Print Name: _Shunda Walker_

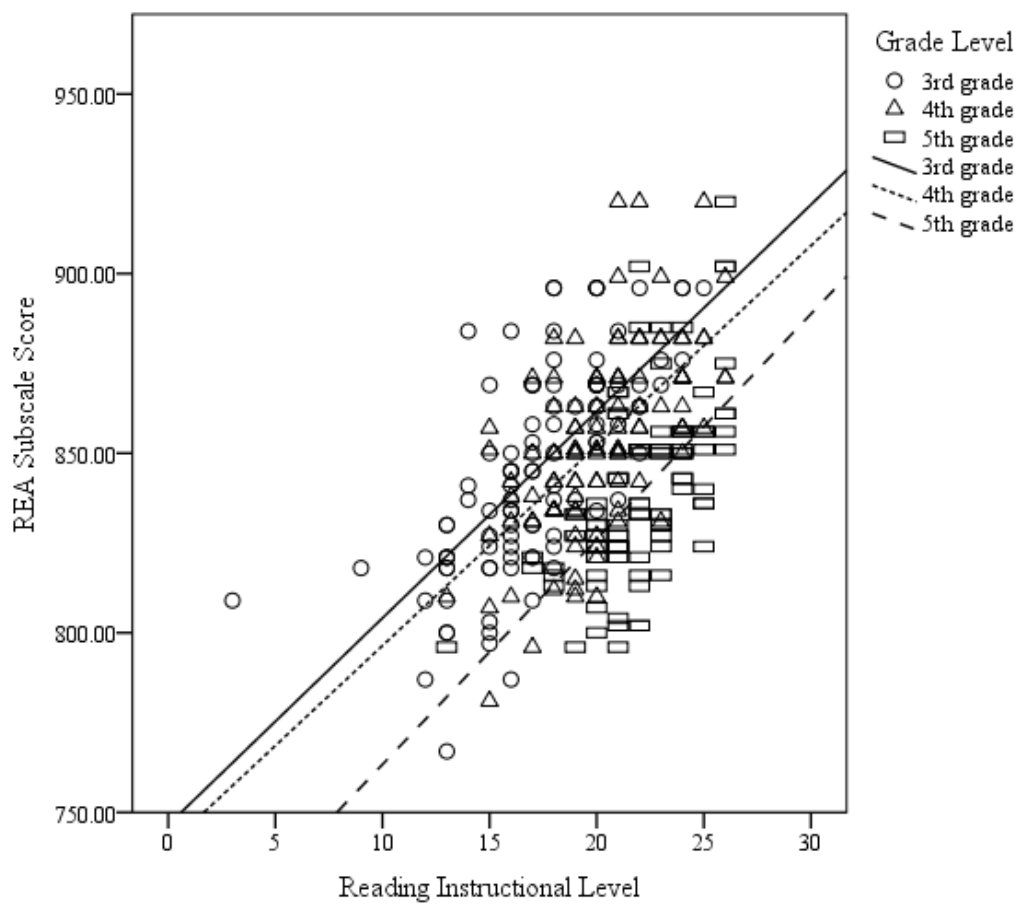Print Title: _3rd Grade Teacher_

Appendix D: Figures

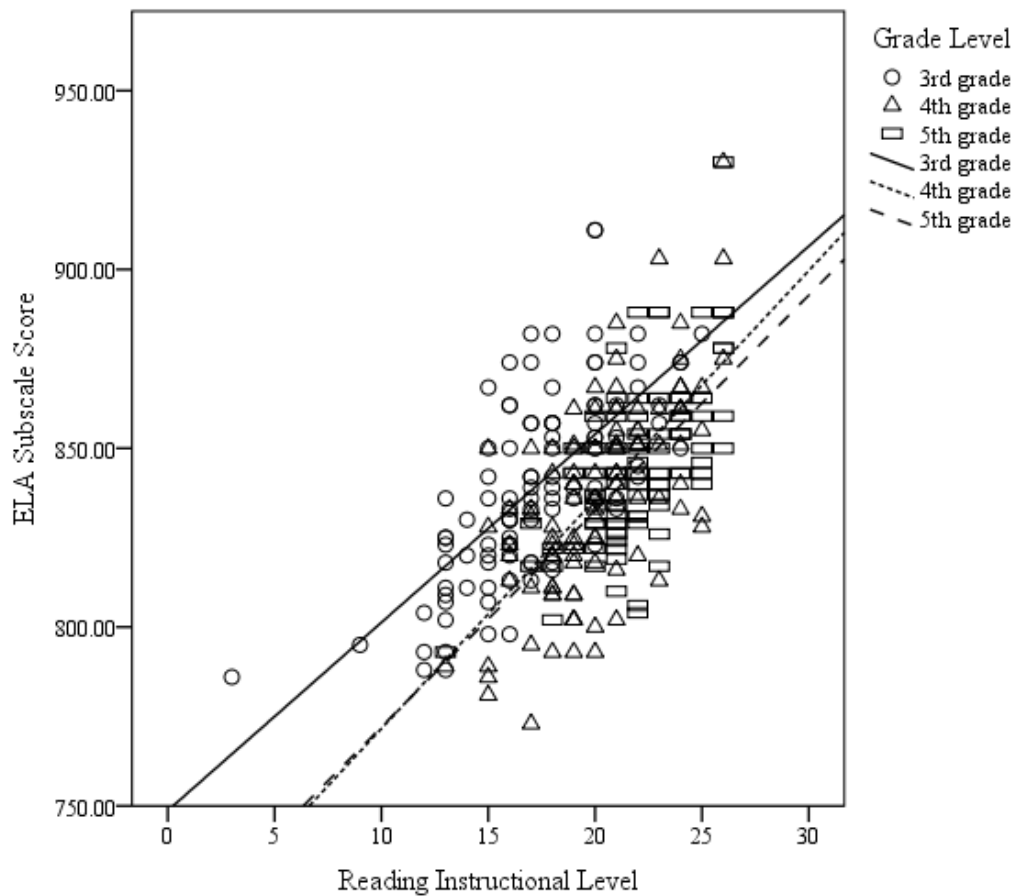Figure 1 Scatterplot of participants' reading instructional level and REA subscale scores

Figure 2 Scatterplot of participants' reading instructional level and ELA subscale scores