INVESTIGATION OF DIFFERENCES IN STAR READING SCORES
FOR SECOND- AND THIRD-GRADE STUDENTS WHO RECEIVED
DIFFERENTIATED READING INSTRUCTION

by

Elizabeth Shealy Hearn

Liberty University

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

Liberty University
July, 2014
INVESTIGATION OF DIFFERENCES IN STAR READING SCORES
FOR SECOND- AND THIRD-GRADE STUDENTS WHO RECEIVED
DIFFERENTIATED READING INSTRUCTION

By Elizabeth Shealy Hearn

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

Liberty University, Lynchburg, VA
July, 2014

APPROVED BY:

Connie L. McDonald, Ph.D., Committee Chair

Gina B. Thomason, Ed.D., Committee Member

Patricia K. Adkins, Ph.D., Committee Member

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

This research study investigated the effects of teaching Response to Intervention curricula in second and third grades in a rural school district in Georgia. A causal-comparative design was used to compare scaled score gains of students who were taught using an RTI approach with curriculum available in How to Plan Differentiated Reading Instruction: Resources for Grades K-3 (Walpole & McKenna, 2009) and students who were taught with other unidentified RTI curricula. Students in both groups participated in either Tier 2 or Tier 3 reading instruction in addition to Tier 1 regular classroom instruction. The STAR Enterprise Reading Test served as the assessment instrument. An ANCOVA was employed for data analysis, statistically controlling for pretest scores. Second-grade students in the treatment group had statistically significant lower scaled scores than second-grade students in the control group. There were no statistically significant differences found between the treatment and control groups in third grade. Limitations and implications are considered with several suggestions for further research.

Keywords: Response to Intervention, STAR Enterprise Reading, differentiated reading instruction, reading, elementary school
Acknowledgments

Thanks to my husband, Greg, and children, Clay, Ellen, and William, for their encouragement. Clay’s direct words, “Mom, just keep showing up,” kept me from quitting on at least two occasions.

Thanks to my dad, Kenneth Shealy, for his faithful support and love for me through the years. Thanks also to my mother, Roberta Robinson Shealy, (1927-2000), who modeled reading and studying to her children on a daily basis.

Thanks to Dr. Connie McDonald for serving as my dissertation committee chair. Her positive emails contained both guidance and encouragement. Her knowledge of literacy was helpful. She kept the process moving so that I could one day finish.

Thanks to my committee members, Dr. Gina Thomason and Dr. Patricia Adkins. Dr. Thomason made me aware of Liberty University and guided me through the various dissertation steps. Dr. Adkins helped me narrow a topic for study and offered a location for the study to take place. Both of them provided guidance and support numerous times.

Thanks to Dr. Patrick Rosopa and Dr. Rihana Mason for their guidance with the data analysis. They answered many questions and increased my understanding of the SPSS reports. Thanks to Dr. Amanda Szapkiw for her expertise and kindness as my research consultant. Her recommendations were always given quickly and with encouragement.

Thanks to my editor, Mrs. Robyn Keeler, for her expertise in proofreading and APA.

Thanks to Dr. Beverly Oxley, an educator, psychologist, and friend, who planted the seed that one day I could obtain this degree in a note that she wrote on a test paper many years ago when I was an undergraduate student at Emmanuel College. I’ve never forgotten her vote of confidence.
Table of Contents

LIST OF TABLES ............................................................................................................. 6
LIST OF FIGURES ........................................................................................................... 7
CHAPTER ONE: INTRODUCTION .................................................................................... 8
  Background of the Study ............................................................................................. 8
  Problem Statement ...................................................................................................... 12
  Purpose Statement ...................................................................................................... 12
  Research Questions ..................................................................................................... 13
  Hypotheses and Null Hypotheses .............................................................................. 13
  Significance of the Study ............................................................................................ 14
  Overview of the Methodology ................................................................................... 16
  Definitions and Acronyms .......................................................................................... 17
  Summary ..................................................................................................................... 20
CHAPTER TWO: REVIEW OF THE LITERATURE ............................................................. 22
CHAPTER THREE: METHODOLOGY .............................................................................. 53
  Research Design .......................................................................................................... 54
  Participants .................................................................................................................. 55
  Setting .......................................................................................................................... 58
  Treatment Fidelity ....................................................................................................... 60
  Instrumentation .......................................................................................................... 61
  Procedures ................................................................................................................... 67
  Data Analysis .............................................................................................................. 68
CHAPTER FOUR: FINDINGS ............................................................................................. 71
CHAPTER FIVE: DISCUSSION ......................................................................................... 89
REFERENCES ................................................................................................................. 106
APPENDIX A: IRB APPROVAL ..................................................................................... 124
APPENDIX B: SCHOOL SUPERINTENDENT APPROVAL ............................................ 125
APPENDIX C: GEORGIA DEPARTMENT OF EDUCATION PERMISSION ................. 126
List of Tables

Table 3.1: Gender of Sample ....................................................... 56
Table 3.2: Grade Level of Sample ................................................... 57
Table 3.3: Ethnicity of Sample ...................................................... 57
Table 3.4: Reading Intervention Block Descriptions ............................. 60
Table 3.5: STAR Enterprise: Reading – Reliability ............................. 63
Table 3.6: Reliability Estimates from the STAR Reading Norming Study: Spring 2008 .... 65
Table 3.7: STAR Enterprise: Reading – Validity ............................... 66
Table 4.1: Progress Monitoring Results of Student with Outlier Score (Treatment Group) .... 73
Table 4.2: Descriptive Statistics of the Second Grade for Kurtosis and Skew ............ 74
Table 4.3: Descriptive Statistics of the Second Grade for Mean Growth .................. 74
Table 4.4: Descriptive Statistics of the Third Grade for Kurtosis and Skew ............... 75
Table 4.5: Descriptive Statistics of the Third Grade for Mean Growth ...................... 75
List of Figures

Figure 1: Multilevel RTI Model ................................................................. 42

Figure 2: Student Outlier Score in the Treatment Group ............................. 72

Figure 3: Histograms of Second-Grade 2012 and 2013 STAR Scaled Scores by Intervention Status ................................................................. 78-79

Figure 4: Scatterplots of Second-Grade 2012 and 2013 STAR Scaled Scores by Intervention Status ................................................................. 80

Figure 5: Homogeneity of Regression Slopes of Second-Grade Sample ...... 81

Figure 6: Histograms of Third-Grade 2012 and 2013 STAR Scaled Scores by Intervention Status ................................................................. 83

Figure 7: Scatterplots of Third-Grade 2012 and 2013 STAR Scaled Scores by Intervention Status ................................................................. 85

Figure 8: Homogeneity of Regression Slopes of Third-Grade Sample ........ 86
CHAPTER ONE: INTRODUCTION

Response to Intervention (RTI) has become a mandated framework in Georgia for documenting and facilitating student achievement in reading. RTI is not a specific program but a systematic process that allows flexibility in using research-based educational strategies and collection of data regarding students’ responsiveness to these instructional strategies (Linder, 2009). As a result, school administrators and teachers are looking for educational strategies and curriculum to use for at-risk students that are evidence-based for increasing student achievement. This research study used a causal-comparative design to investigate a reading curriculum used as an intervention for at-risk students in four elementary schools. The study used scaled scores from the STAR Enterprise Reading Test to determine reading growth. The study compared the scaled scores of students who received traditional reading instruction and the targeted intervention with the scaled scores of students who received traditional reading instruction and did not receive the targeted intervention. Chapter 1 provides a background of the study, identifies the problem and purpose of the research, explains the significance of the topic, and summarizes the methodology used.

Background of the Study

RTI emerged nationally in the late 1990s from a debate among educators over the most valid approach for identifying students with a reading disability. Educators wanted empirical evidence on whether a student’s reading difficulties were caused by inadequate literacy instruction or by inadequate cognitive abilities (Vellutino, Scanlon, Sipay, Small, Prat, & Chen, 1996). In earlier times, it was taught that learning difficulties were the result of a low IQ as seen by the work of educational psychologist Cyril Burt. “Capacity must obviously limit content. It is impossible for a pint jug to hold more than a pint of milk, and it is equally impossible for a
child’s educational attainment to rise higher than his educable capacity” (Burt, 1937, p. 477). Following this archaic view was the IQ-achievement discrepancy model which indicated a learning disability if there was a significant difference between a student’s scores on a general intelligence test and scores on an achievement test. More recent research suggests that the traditional IQ-achievement discrepancy can no longer be considered as a robust determinant of specific learning disabilities (Stuebing, Barth, Molfese, Weiss, & Fletcher, 2009). Multiple studies now indicate that reading achievement is determined by a students’ exposure to explicit reading interventions and that when students fail, it is more likely caused by a lack of literacy instruction.

The idea that explicit instruction is a major determiner in reading achievement aligns with Vygotsky’s Social Development Theory (Miller, 2010), which postulates that cognitive development takes place when a child has social interaction with a more skillful tutor. Vygotsky emphasized the need for students to be guided by a more capable teacher to advance them from their current level of independent capability to a higher level of potential capability when provided explicit instruction. Vygotsky’s theory is affirmed in various studies seen in the Review of the Literature in which RTI models suggest significant reading achievement improvements (see Velluntino et al., 1996; Torgesen et al., 2001; Gresham & Vellutino, 2010; Graves, Brandon, Duesbery, McIntosh, & Pyle, 2011). According to Gresham and Vellutino (2010), student responses to intervention and their ability to maintain and transfer gains from the intervention are better predictors of long-term reading achievement than IQ.

RTI gained prominence with the reauthorization of the Individuals with Disabilities Education Improvement Act (IDEA) of 2004, which affords school psychologists greater flexibility in evaluating students for special education services (Klotz & Nealis, 2005). For the
first time in almost 30 years, IDEA allowed state and local education agencies an alternative means of diagnosing learning disabilities other than the traditional IQ-achievement discrepancy model. This Act also urged school administrators and teachers to establish research-based RTI models in their school districts that would provide consistent interventions for struggling students and uniform criteria for determining student disabilities. RTI is in direct contrast to the earlier wait-to-fail model in which educators were instructed to wait and see if an at-risk student showed academic improvement or academic failure over a period of time. In the past, students sometimes showed signs of learning difficulties for years before their teachers provided intervention strategies (Stecker, Fuchs, & Fuchs, 2008).

RTI’s three foundational goals are (a) early detection measures for screening students, (b) delivery of timely interventions before struggling students have a chance to fall further behind, and (c) modifying instructional practices based on patterns of student responses (Grigorenko, 2009). By targeting students at the first sign of difficulties and providing instructional support, school systems reduce or even prevent academic failure (Gersten et al., 2009). There is also the possibility that RTI will reduce the overidentification of learning-disabled students due to the emphasis on prevention and effective teaching (Jimenez, 2010).

RTI models are currently widespread across the country as shown in annual nationwide RTI Adoption surveys. Spectrum K12 School Solutions, Inc., the National Association of State Directors of Special Education, the Council of Administrators of Special Education, and the American Association of School Administrators sponsored these surveys from 2007 to 2010 to gather data on RTI implementation across the country. The web-based surveys, sent to district administrators including RTI Directors, Special Education Directors, and Assistant Superintendents, reported increased growth from 2% full implementation in 2007 to 61% full
implementation in 2010 (Spectrum K12, 2007; Spectrum K12, 2010). The core components of any RTI model are universal screening, high-quality core instruction, progress monitoring, tiered interventions, collaborative data-based decision making, parent involvement and administrative support (Shapiro, 2012). Limitations to the RTI framework include reliability and uniformity issues due to the flexibility school administrators have in implementing the model. IDEA 2004 strongly encourages the development of research-based interventions to establish consistency and integrity in identifying learning disabilities (Klotz & Nealis, 2005). In order to establish a robust RTI model, schools need to identify and use research-based interventions that align with the intentions of IDEA.

RTI is the focus of much research currently being conducted in the field of education. There is extensive literature explaining the purpose and goals of RTI. However, there is a limited amount of evidenced-based research on specific interventions. About 75% of the respondents from the nationwide 2009 RTI Adoption survey reported that their need for evidence-based interventions was a significant obstacle to their implementation of RTI models (SpectrumK12, 2009). Again, in 2010, the RTI Adoption survey respondents reported one of the primary obstacles in fully implementing RTI was the lack of available intervention resources (SpectrumK12, 2010). Since RTI is designed for kindergarten through high school, there is also an abundance of literature describing various RTI models for elementary, middle, and high school-aged students. However, the research is limited regarding RTI outcomes for specific grades. Wanzek and Vaughn (2010) suggest that “the difficulties students face as they enter second and third grade are more complex than in kindergarten or first grade” (p. 309). This complexity creates a need for specific curriculum for second- and third-grade students. Research is needed on RTI interventions for specific grades to provide educators with evidence regarding
grade-appropriate intervention curricula (Denton et al., 2010). New interventions designed for RTI instruction need to be tested to meet the legislative requirement for research- and evidence-based instruction methods.

**Problem Statement**

School administrators and teachers in Georgia are mandated to use the RTI framework for providing appropriate and purposeful instruction for all students (Georgia Department of Education, 2008). IDEA mandates the use of research-based educational practices for collecting RTI data regarding students’ responsiveness to instruction (Linder, 2009). Since schools have the responsibility to use research- and evidence-based interventions (Wright, 2007), new interventions marketed as RTI curricula need to be researched to show empirical support on whether or not they are effective.

Although RTI has been developing over the last decade, educators still struggle to find specific grade curriculum for RTI that meets the research- and evidence-based requirements. In light of this struggle, reading teachers may attempt to develop their own curriculum. Mokhtari, Porter, and Edwards (2010) suggest that although RTI has been around for a few years, many reading teachers still have insufficient knowledge to design and implement multitiered reading instruction that is responsive to student needs. While research exists on overall RTI organization and implementation, research is needed on specific grade level interventions written for the RTI model to add curriculum evidence to existing literature (Graney, 2008; Gresham, VanDerHeyden, & Witt, 2005; Moats, 2007; Mokhtari et al., 2010; Stecker et al., 2008).

**Purpose Statement**

The purpose of this causal-comparative study was to investigate the differences in STAR Enterprise Reading (STAR) scores for second- and third-grade at-risk students who either
received or did not receive the targeted reading intervention. The Social Development Theory, which embraces the idea that students learn through social interaction from a more skillful tutor (McLeod, 2007), formed the framework of this study. The study examined whether or not a new reading curriculum specifically designed for RTI that was provided to at-risk students could significantly increase STAR scaled scores by comparing the intervention status of at-risk students and their scores on the 2013 STAR ending benchmark test, while controlling for differences in ability using their 2012 STAR beginning benchmark test. The study defined intervention status as at-risk students who received traditional reading instruction and the targeted reading intervention or at-risk students who received traditional reading instruction but did not receive the targeted reading intervention.

**Research Questions**

Research Question 1: While using 2012 STAR beginning-of-year scaled scores as a control variable for differences in achievement, do at-risk second-grade students who receive Differentiated Reading Instruction (DRI) through RTI have significantly different scaled scores on the 2013 STAR end-of-year test when compared to at-risk students who do not receive this intervention?

Research Question 2: While using 2012 STAR beginning-of-year scaled scores as a control variable for differences in achievement, do at-risk third-grade students who receive Differentiated Reading Instruction (DRI) through RTI have significantly different scaled scores on the 2013 STAR end-of-year test when compared to at-risk students who do not receive this intervention?

**Hypotheses and Null Hypotheses**

$H_1$: While using 2012 STAR beginning-of-year scaled scores as a control variable for
differences in achievement, at-risk second-grade students who receive Differentiated Reading Instruction (DRI) through RTI will have significantly different scaled scores on the 2013 STAR end-of-year test when compared to at-risk students who do not receive this intervention.

H2: While using 2012 STAR beginning-of-year scaled scores as a control variable for differences in achievement, at-risk third-grade students who receive Differentiated Reading Instruction (DRI) through RTI will have significantly different scaled scores on the 2013 STAR end-of-year test when compared to at-risk students who do not receive this intervention.

The corresponding null hypotheses that were tested were:

H01: While using 2012 STAR beginning-of-year scaled scores as a control variable for differences in achievement, at-risk second-grade students who receive Differentiated Reading Instruction (DRI) through RTI will not have significantly different scaled scores on the 2013 STAR end-of-year test when compared to at-risk students who do not receive this intervention.

H02: While using 2012 STAR beginning-of-year scaled scores as a control variable for differences in achievement, at-risk third-grade students who receive Differentiated Reading Instruction (DRI) through RTI will not have significantly different scaled scores on the 2013 STAR end-of-year test when compared to at-risk students who do not receive this intervention.

Significance of the Study

IDEA contained language in the House committee report that directed “high quality research to evaluate the effectiveness of [RTI] models” (House of Representatives Report No. 108-77, 2004). This same reauthorization contained language in the Senate committee report that encouraged local education agencies to adopt research-based models to ensure consistency and integrity in the system that diagnoses student learning disabilities (Senate Report No. 108-105, 2004). Raudenbush (2007) suggests that effectiveness studies of evidence-based
educational practices are essential to demonstrate that the innovation can produce positive effects when implemented in various settings.

Wanzek and Vaughn (2010) describe a synthesis of 18 studies they examined to determine the range of effects from small to large for students participating in interventions. They concluded that “there is . . . substantially more research for extensive interventions in the K-1 level than for students in 2nd and 3rd grade” (p. 309). Results from this study increase the knowledge base of studies specific for second- and third-grade students. More explicitly, this study contributes RTI data to the current body of knowledge regarding a reading curriculum authored by Sharon Walpole and Michael C. McKenna. These authors designed this curriculum through collaboration with many educators across the country who were trying to provide targeted instruction for specific needs. This RTI curriculum does not group children by instructional reading level, fluency level, or lexile measures. These approaches which target grade-level passages with comprehension questions and individual words and sentence lengths do not always allow individual differences in skill needs. DRI is based on a reading inventory which targets specific reading deficiencies. Targeting these deficiencies allows more precise remediation than level-based reading curricula (Walpole and McKenna, 2009).

Evidence provided from this study may be used by school administrators and school data teams in making data-based decisions as they continue their quest for implementing effective RTI reading interventions. Another benefit of the study was the opportunity for educators to engage in collaboration while working together to assist students with reading difficulties. Walpole and McKenna (2009) suggest that teachers who collaboratively use data to identify needs, form groups, and plan instruction create a problem-solving atmosphere where problems and solutions are shared, resulting in a less isolated work environment. Teachers also gained
understanding in the RTI process as they implemented the intervention and documented the data.

**Overview of the Methodology**

All second- and third-grade students in the school district used for this study were administered the STAR Test at the beginning and end of the 2012-2013 school year as part of their in-school routine experience. The U.S. Department of Education National Center on Response to Intervention (NCRTI) lists STAR as both a highly valid and reliable assessment for both universal screening and progress monitoring (NCRTI, 2012). In addition, Salvia, Ysseldyke, and Bolt (2012) reported that in 2008, STAR was renormed “on a large and representative group of students” (p. 289). Evidence for reliability and validity from this renorming was satisfactory.

The research sample for this study included 97 at-risk second- and third-grade students who were selected from four elementary schools in one school district in northeast Georgia. These students were identified as at-risk because their scaled scores on the 2012 beginning-of-year STAR ranked below the 40\textsuperscript{th} percentile (PR). Students were divided into two groups as follows: second- and third-grade students receiving the targeted intervention in addition to regular classroom instruction, and second- and third-grade students not receiving the targeted intervention in addition to regular classroom instruction. DRI, the intervention curriculum found in *How to Plan Differentiated Reading Instruction: Resources for Grades K-3* (Walpole & McKenna, 2009), provided three differentiation targets for this study, including phonological awareness, word recognition, and fluency. Differentiation instruction cycles embedded in these targets included basic alphabet knowledge, using letter sounds, using letter patterns, blends and digraphs, R-controlled vowels, vowel-consonant-E, and vowel teams.

One of the authors of the DRI intervention curriculum, Michael C. McKenna, a Thomas
G. Jewell Professor of Reading at the University of Virginia, reported three arguments to support the validation of using this curriculum, even though it is new to the RTI literature. “First, we built the program around instructional approaches that have already been validated individually, such as semantic feature analysis. This is strong indirect evidence of effectiveness. Second, we have conducted and reported one study contrasting DRI with guided reading at grade one. The results were very encouraging. Third, the approach has been implemented in Staunton City, [Virginia] with good, though still preliminary results” (M. C. McKenna, personal communication, July 12, 2013). In addition, the DRI curriculum was promoted by the Georgia Department of Education in 2011, which offered professional development for teachers on how to implement this curriculum through its Regional Educational Service Agencies.

The 2013 end-of-year STAR was the dependent variable for this study. The independent variable was the intervention status (receiving DRI intervention, not receiving DRI intervention), and the covariate was the 2012 beginning-of-year STAR. The Statistical Package for the Social Sciences (SPSS) was used to analyze the data. Summary statistics were computed, and the assumptions of normality, homogeneity of variances, linearity, and homogeneity of regression slopes were tested. Two analyses of covariance (ANCOVA) tests were conducted to compare the differences in the 2013 end-of-year STAR scaled scores of those students who received DRI and those students who did not receive DRI. Data were analyzed for second-grade students and third-grade students separately.

**Definitions and Acronyms**

**ANCOVA**--Analysis of covariance (ANCOVA) is an extension of analysis of variance that allows exploration of differences between groups while statistically controlling for an additional continuous variable (Pallant, 2007).
CAT--Computer-adaptive testing (CAT) is an Internet software program that provides educators with a means of evaluating the effectiveness of instruction through universal screening and progress monitoring (Shapiro, 2012).

CBM--Curriculum-based measurement (CBM) is an Internet software product that provides educators with a means of evaluating the effectiveness of instruction through universal screening and progress monitoring (Shapiro, 2012).

CBM-DD--Curriculum-based measurement-dual discrepancy model (CBM-DD) was developed by Fuchs and Fuchs (1998) and was a forerunner of the RTI model (Gresham et al., 2005).

DIBELS--Dynamic indicators of basic early literacy skills (DIBELS) is Internet-based software commonly used in the United States for universal screening (Renaissance Learning, 2010).

DRI—Differentiated reading instruction (DRI) refers to a specific reading intervention curriculum authored by Sharon Walpole and Michael C. McKenna. It is found in How to Plan Differentiated Reading Instruction: Resources for Grades K-3 (Walpole & McKenna, 2009).

DRI is the independent variable for this study.

IDEA--The Individuals With Disabilities Education Improvement Act (IDEA) is the federal statute, originally passed in 1975, that prescribes services to students aged 3-21 with disabilities (Buffum, Mattos, & Weber, 2009).

IQ-Achievement Discrepancy--IQ-achievement discrepancy is a formula used for identifying students with learning disabilities which was the fundamental marker for diagnosing learning disabled students prior to IDEA 2004 (Gresham et al., 2005).

LD--Learning disability (LD) is a disorder that affects people’s ability either to interpret what they see and hear or to link information from different parts of the brain. It may also be referred to as a learning disorder or a learning difference (www.ldonline.org).
LEA--A local education agency (LEA) is a public board of education or public authority legally constituted within a state to perform a service function for public schools in a city, county, or school district (definitions.uslegal.com).

NCRTI--The National Center on Response to Intervention (NCRTI) is a website which provides information and technical assistance to states and districts in implementing proven models of RTI (http.rti4success.org).

POI--The pyramid of interventions (POI) is a systematic program of supports that become increasingly more directive, intensive, and targeted, which was named after the intervention program created at Adlai E. Stevenson High in Lincolnshire, Illinois (Buffum et al., 2009).

PR--The percentile rank (PR) in the Renaissance Learning software is a norm-referenced score based on what students across the nation can do. A student with a percentile rank of 85 performs better than 85 percent of students nationwide of the same grade at the same time of year (Renaissance Learning, 2010).

RTI--Response to intervention (RTI) is the practice of providing high-quality instruction matched to students’ needs, monitoring progress frequently to make changes in instruction or goals, and applying child response data to important educational decisions (Buffum et al., 2009).

SPSS—The Statistical Package for the Social Sciences (SPSS) is computer software that is commonly used for calculating a wide variety of statistics in social sciences research.

SS--The scaled score (SS) on a standardized test transforms the raw score to reflect the same level of performance across different forms of a test taken at different times. This formative assessment is the most important score that STAR reports, and all other scores that are calculated are derived from the scaled scores (Tan & Michel, 2011; Renaissance Learning, 2010).

SST--A student support team (SST) is a multi-disciplinary team that engages in a collaborative
problem-solving process to address the learning or behavior problems of their students (Georgia Department of Education, 2008).

*STAR*--STAR enterprise reading test (STAR) is an Internet-based reading measurement tool recommended by the NCRTI for having the “highest scientific standards as a tool for RTI” (Renaissance Learning, 2010, p. 20; http://www.rti4success.org).

*Tier 2 Intervention Group*--Tier 2 is the second level of instruction in the pyramid of interventions designed for students who are unsuccessful in Tier 1, the core curriculum for all students (Fuchs, Fuchs, & Vaughn, 2008).

*Tier 3 Intervention Group*--Tier 3 is the third level of instruction in the pyramid of interventions designed for students who make inadequate progress in response to Tier 2 interventions (Fuchs, Fuchs, & Vaughn, 2008).

*ZPD*--The zone of proximal development (ZPD) is a developmental range with a lower limit in which one can work independently and a potential higher limit in which one can eventually succeed once provided with guidance from a more knowledgeable person. The ZPD concept, founded by Vygotsky (McLeod, 2007), is used in Renaissance Learning data reports.

**Summary**

RTI is a framework used by Georgia public schools to assure that all students are provided with instructional support to be successful. Initiated by IDEA, RTI promotes early detection of learning difficulties, effective interventions to address those learning difficulties, and use of data to show student outcomes. Educators are required to use research- and evidence-based interventions, which is a challenge to school administrators due to a limited amount of literature available on specific interventions for specific grades that has evidence of improved student achievement. This study investigated one RTI reading intervention specifically designed
for students in kindergarten through third grade. Second- and third-grade students were targeted in this study because of the need for more research for specific interventions for students in the second and third grade (Wanzek & Vaughn, 2010) and because STAR is designed to collect data for second- and third-grade students. Chapter 2 reviews literature regarding RTI, including RTI’s history and the various components that are included in RTI. Chapter 3 discusses the methodology used in the study. Chapter 4 interprets the results of the data analysis, and Chapter 5 discusses the results and highlights recommendations for practice and future research.
CHAPTER TWO: REVIEW OF THE LITERATURE

RTI gained national acclaim as a best-practice model in education when IDEA recognized the potential for the prevention of school failure in this framework and identified RTI as an alternative for identifying learning disabled (LD) students (Stecker et al., 2008). The state of Georgia adopted RTI as the framework for providing a student support team in every public school. Student support teams are established to prevent inappropriate referrals to special education and unnecessary removal of students from general education (Georgia Department of Education, 2011). In addition, RTI is the structure for providing high quality research-based instruction, interventions, and data-driven practices for all students in Georgia public schools.

In Georgia, RTI begins in general education classrooms with Tier 1 common core curriculum as the foundation for all students. Students who have learning difficulties in Tier 1 enter into Tier 2, where they receive explicit, targeted instruction in a small-group setting in addition to the core curriculum. Students who continue to have learning difficulties move into Tier 3, where the frequency and intensity of instruction increases even more. A student still facing learning difficulties after Tier 3 may qualify for a special education evaluation and, if found eligible, be placed in Tier 4, which is special education.

Educators use RTI to align appropriate assessment with expert instruction and intervention for both academic and behavioral needs. RTI envisions effective instruction to prevent problems and effective interventions in a timely manner to address problems that occur (Baker, Fien, & Baker, 2010). The effectiveness of the instruction and interventions is determined by progress monitoring and benchmark data, which are used by eligibility teams to make special education eligibility determinations. RTI provides the required evidence for diagnosing specific learning or behavioral disabilities for students who do not respond
adequately to instruction and intervention (O’Conner, 2007). However, at the core of RTI is the notion that students are placed in special education only if and when rigorous and expert instruction and intervention have not resulted in adequate growth (Mokhtari et al., 2010). The annual nationwide RTI Adoption survey for 2009, sponsored by Spectrum K12 School Solutions, Inc., the Council of Administrators of Special Education, and the American Association of School Administrators reported that 83% of the districts that responded indicated that RTI had reduced their number of special education referrals (Spectrum K12, 2009).

Literature regarding RTI has been reviewed to lay a foundation for this study and to examine how previous research influences current RTI practices. The literature review is divided into eight sections. The first section, Theoretical Framework, explains how RTI is framed in Vygotsky’s Social Development Theory. The second section outlines the history of RTI from its earliest beginnings in the late 1990s through current-day legislation and expectations. The third section discusses the benefits of RTI, and the fourth section explains various approaches and implementation norms from across the country. Section five is a description of the RTI components. The sixth section explores existing research demonstrating RTI’s effectiveness, and the seventh section describes effective RTI instruction. The eighth section highlights the specific intervention that is used in this study, DRI, found in How to Plan Differentiated Reading Instruction: Resources for Grades K-3 (Walpole & McKenna, 2009).

**Theoretical Framework**

Lev Vygotsky (1896-1934) developed the Social Development Theory based on the idea that cognitive development is promoted by social interaction. Children learn through social interaction or assistance from a more skillful tutor, such as a peer, parent, or teacher. Vygotsky further attested that everyone has a Zone of Proximal Development (ZPD), which is a range with
a lower limit in which one can work independently and a potential higher limit in which one can succeed with guidance from a more capable person (McLeod, 2007).

As applied to this study, this theory implies that teaching methods for beginning readers can impact reading growth. Vygotsky endorsed social interaction for expanding a student’s acquisition of knowledge as seen in cooperative learning groups, small groups with a teacher, peer partners, or any activity where someone knowledgeable in a skill is teaching someone else the skill. The intensity of the instruction is determined by the level of need and is diminished as the need decreases (Miller, 2010).

The structure of the Social Development Theory parallels the RTI multi-tiered framework. When students experience learning difficulties in the general education setting, they receive additional instruction in a small-group setting with an experienced teacher who provides specific instruction for a targeted weakness (Tier 2). After a period of time, the student either becomes skilled in the weak area and no longer needs the intervention, or is moved to a small-group setting with even more intense instruction (Tier 3). In RTI, the academic deficiencies of the student determine the amount and intensity of interventions they receive in an effort to give every student the best possible chance for success. Thus, RTI operates on the assumption, like the Social Development Theory, that learning difficulties can be overcome when targeted interventions are taught by a skillful teacher in an environment of a few students interacting with each other over a period of time.

Response to Intervention History

The Education for All Handicapped Children Act of 1976 began federal support for special education services in the United States. Among its many provisions, this law mandated a free and appropriate education for students with disabilities in the least restrictive environment
By 1982, Heller, Holtzman, and Messick were beginning to argue that a special education classification was only valid when the general education program and the special education program were effective and when the assessment process was accurate and well documented.

The 1976 law was renewed under the Individuals with Disabilities Act of 1990. Studies by Fuchs and Fuchs in the mid-1990s began using curriculum-based measurement (CBM) to determine students’ responsiveness to interventions in the general education classroom. They argued that educators should consider special education only if a student’s performance and learning rate are both substantially below that of classroom peers (Gresham et al., 2005).

IDEA was again reauthorized in 1997. The 1997 code required “pre-referral interventions to reduce the need for labeling children as disabled in order to address their learning needs” (IDEA Section 601[c] [5], 1997). Between 1997 and the next IDEA reauthorization of 2004, an intense debate escalated in the education field regarding the IQ-achievement discrepancy process for identifying children with learning disabilities. The heart of the debate revolved around the data that discrepant and nondiscrepant low achievers are not significantly different in achievement, behavior, and processing domains, and they do not differ in their response to instruction (Gresham et al., 2005). Speece, Case, and Molloy (2003) showed further evidence that single indicators of reading difficulty were not sensitive indicators of both low performance and below-level growth rate.

Another problematic issue regarding eligibility criteria for learning disabilities included research studies that reported substantial inconsistencies in LD eligibility criteria across the nation (Wedl, 2005). Baer (n.d.) suggested that the discrepancy model was flawed because of the assumption that IQ and achievement tests are perfectly correlated. The statistical
phenomenon of regression toward the mean explains why this assumption is not true. Students with IQs above 100 may have achievement scores below their IQ scores, and students with IQs below 100 may have achievement scores above their IQs. Also, the issue of bias for some populations of students was questioned. Minority children and children from lower socioeconomic families consistently score lower on standardized tests of ability than those from non-minority, higher socioeconomic families (Heller et al., 1982). This tendency leads to over-representation or under-representation of certain groups in special education.

Another major concern was the disconnection between assessment and instruction and the fact that the discrepancy approach does not inform instructional decisions. Gresham (2001) synthesized researcher criticisms of the IQ-achievement discrepancy model in the following statement:

The most serious flaw in the current process is the absence of a direct link between assessment procedures used for identification and subsequent interventions that might be prescribed on the basis of these assessment procedures. What appears to be needed is an approach to defining LD that is based on how students respond to instructional interventions rather than on some arbitrarily defined discrepancy between ability and achievement. (p. 3)

The literature supports the increasing irrelevance of using IQ-achievement discrepancy measurements as valid markers for diagnosing learning disabilities in children (Stuebing et al., 2009). The literature also supports the argument that deficits of experience and instruction, rather than biological cognitive deficits, cause early reading difficulties in many children (Gresham & Vellutino, 2010). These authors also suggest that the majority of children whose reading deficiencies are identified in kindergarten or sooner and who are provided adequate
interventions in foundational literacy skills perform at least in the average range of their level of reading long-term.

The need for early detection and intervention is highlighted by research found in the literature over the past two decades. Numerous studies emphasize the significance of a child’s early years of formal schooling (Shapiro, 2012). The U.S. Department of Education (2002) reports that between first and fourth grades, the odds of being labeled LD increase linearly by 450%. Specifically, between first and second grades, the LD identification rate doubles. The rate doubles again between second and third grades, and doubles again between third and fourth grades. Juel (1988) demonstrates through a longitudinal study that children who have not mastered basic reading skills by the end of first grade have an 88% probability of continued reading deficits in fourth grade. Graves (2010) suggests that about 74% of all third-grade children who are identified with reading difficulties continue to have reading problems in the sixth grade. Edmonds et al. (2009) indicate that reading problems noticeable in a child’s early years tend to develop into a permanent, lifelong struggle with reading.

As the significance of early childhood education and investigative discrepancies became apparent, the RTI model surfaced as a variation of the scientific method, which is commonly used to study natural phenomena (Wedl, 2005). Beginning with the late 1990s, researchers began focusing on instructional models or programs which used tiered frameworks to promote learning for all students, including those identified with learning difficulties. According to Allington (2009), Marie Clay’s research-based Reading Recovery program, which was introduced in the United States in the 1980s, was one of the first programs to lay out a framework for RTI. The Reading Recovery program emphasized intensive, one-on-one instruction, extensive teacher training, and a comprehensive collection of data on individual
students’ progress (Lyons, 1998). The Reading Recovery program and the RTI model are similar in that both promote early identification of reading difficulties, progress monitoring for gathering data, placement based on academic deficiencies, and intensive, expert, one-on-one or small-group instruction with high teacher fidelity. Reading Recovery primarily emphasizes the early grades (Mokhtari et al., 2010) whereas RTI is designed for all students in need of targeted instruction, regardless of the age or grade (Buffum et al., 2009).

In the early stages, RTI was called the Curriculum-Based Measurement Dual Discrepancy model (CBM-DD) and consisted of three phases: (a) documentation of adequate classroom instruction and dual discrepancies, (b) implementation of a pre-referral intervention focusing on a student’s dual discrepancy, and (c) the design and implementation of an extended intervention plan (Gresham et al., 2005). These early phases can be seen in the current RTI components as (a) collecting data through universal screenings and progress monitoring, (b) implementing research-based interventions with specific targets, and (c) using data-based decision-making in developing individualized instructional plans (NCRTI, 2012). The reauthorization of IDEA 2004 provided that “in determining whether a child has a specific learning disability, a local education agency (LEA) may use a process which determines if a child responds to a scientific, research-based intervention” (IDEA, 2004, p. unknown). This law allowed the RTI model to be used as an alternative to the IQ-achievement discrepancy model.

As noted, IDEA did not require or eliminate the IQ-achievement discrepancy model for LD children, but offered the alternative of using the RTI model to identify LD children. Since that time, some states have completely discontinued the use of the IQ-achievement discrepancy model. NCRTI reports that as of 2010, seven states use RTI only for LD identification, and five states use RTI and Other (but no IQ discrepancy) for LD determination. The remaining 38 states
continue to use RTI and IQ discrepancy or RTI and IQ discrepancy and other for determining learning disabilities (NCRTI, 2010). Cassidy and Cassidy (2008, 2009) report that RTI is one of the top five topics in their annual What’s Hot and What’s Not Literacy Survey along with adolescent literacy, English-language learners, high-stakes assessment, and literacy coaching. The implementation and data collection of RTI continues to challenge educators and school psychologists as they modify and refine the model so that the actual application of it in school settings will resemble the potential it has for all children (Gresham et al., 2005).

**Response to Intervention Benefits**

Gresham (2001) and Vaughn and Fuchs (2003) record four benefits of implementing the RTI model: (a) providing immediate help to struggling children, (b) perceiving the learning problem as a risk rather than as a deficit, (c) decreasing the opportunities for race or gender biases by eligibility teams, and (d) providing a high focus on student outcomes. Early identification and remediation of problems prevent students from waiting to fail. The IQ-achievement discrepancy approach was discredited, in part, because students had to fail severely enough for long periods of time before they showed large enough deficits in academic achievement to satisfy the severe discrepancy requirement and be referred for remediation. The IQ-achievement discrepancy approach was described as the “wait-to-fail” approach (Gresham et al., 2005). Ciolfi and Ryan (2011) also suggest growing evidence that the discrepancy approach is particularly problematic for students living in poverty, students from other cultures, and students whose native language is not English.

MacMillan and Siperstein (2002) suggest that referrals without explicit data often result in false positives (inaccurate identification of students as LD) and/or false negatives (failure to identify students who are LD). Garcia and Guerra (2004) suggest that the false positive/false
negative phenomenon results from teachers identifying problems within students, families, and communities, while failing to consider school practices and student outcomes. The RTI approach assists in diminishing the false positive/false negative phenomenon while closing the gap between the identification of students with learning difficulties and the intervention for students with learning difficulties (Vaughn & Fuchs, 2003). The RTI approach also provides screening and early interventions to remediate academic difficulties (Jenkins & O’Connor, 2002).

In the past, the deficit model was used for LD remediation in which a processing deficit was identified and an instructional strategy was matched to it. After 20 years of disappointing research, Cronbach (1975) replaced the deficit model with a process similar to a problem-solving risk model (Tilly, Reschly, & Grimes, 1999). The problem-solving risk model is now used in the RTI framework. RTI uses a risk model in which all students are screened for learning difficulties and specific learning weaknesses. Students identified as at-risk are then given evidence-based remediation targeted at the specific weakness. Evidence-based interventions are interventions that are accompanied by data showing that the intervention results in positive student outcomes (Georgia Department of Education, 2009).

RTI is no longer just a means for identifying LD students. Rather, the risk model expands RTI into a general education practice (Kavale & Spalding, 2008) for all students through the core curriculum in conjunction with providing an alternative means of identifying students with learning disabilities (Buffum, Mattos, & Weber, 2012). Moore and Whitfield (2009) propose that RTI activities are fundamentally a part of the general education classroom, and Graner, Faggella-Luby, and Fritschmann (2005) assert that all educators in schools are using RTI methods. Baker et al. (2010) further state that schools cannot validly employ RTI for
identifying LD students unless they have first established an RTI program that demonstrates effective prevention and intervention instruction through the Tier 1 level of core curriculum.

Donovan and Cross (2002) allege that the teacher referral process used in the past for selecting students for special education evaluations may have been biased. Before RTI surfaced as an alternative for identifying students with learning disabilities, special education records indicated an overidentification of boys and an underidentification of girls as LD. The teacher referral process was based on relativity: a student’s performance relative to the performance of the class. MacMillan and Siperstein (2002) report that in the past a teacher’s decision to refer a student could have been influenced by factors such as gender, socioeconomic status, and race, in addition to academic deficiencies. The RTI approach for referring students for special education evaluation eliminates the teacher referral process based on relativity. RTI requires evidence of targeting specific weaknesses, providing interventions to address those weaknesses, use of problem-solving efforts by support teams, and evidence-based data to make referral decisions. RTI has reduced the disproportionate representations in special education of certain minority and socioeconomic groups as well as males versus females (Gresham et al., 2005).

RTI concentrates on three factors that affect student academic outcomes: (a) useful and appropriate instruction, (b) direct and frequent measurement of achievement, and (c) instructional environment. The National Reading Panel (2000) maintains that many children identified as LD in reading were either exposed to ineffective reading curricula or were exposed to effective reading curricula that were implemented with poor integrity. Vaughn, Linan-Thompson, and Hickman (2003) label these children as “instructional casualties.” Clay (1987) argues that these children “learn to be learning disabled” from lack of exposure to fundamental skills and from poor teacher fidelity. RTI is an educational reform targeted for preventing
“instructional casualties” and is designed to provide every child with the additional time and support needed for high levels of learning (Burns, Appleton, & Stehouwer, 2005). “Applied to both academics and behavior, RTI is increasingly viewed as an important and major effort to reform the delivery of instruction, especially at the elementary level” (Shapiro, 2012, p. 6).

The instructional environment of RTI includes the assessment of evidence-based strategies, allocated learning time, pacing and sequencing of instruction, and teacher fidelity in a general education classroom to promote positive child outcomes for all students (Gresham et al., 2005). The intensity of the intervention is designed to match the severity of the problem (Gresham, 2001) which is seen in the multi-tiered design of the model. Teacher fidelity means that the delivery of the instruction or intervention is implemented in the manner in which it was designed (Georgia Department of Education, 2009). RTI requires regular documentation of teacher implementation measures and accurate data of universal screenings and progress monitoring to ensure that the instruction and interventions are delivered consistently and with a high degree of accuracy (Togut, 2012).

**Response to Intervention Approaches**

Fuchs, Mock, Morgan, and Young (2003) describe two basic approaches used to provide RTI instruction: (a) the standard protocol approach and (b) the problem-solving approach. Fuchs et al. (2003) claim that researchers generally support the standard protocol approach for both empirical and logistical reasons. Since the issue of treatment fidelity is so important, this approach is more easily validated and provides a higher assurance that programs are researched-based (Stecker et al., 2008). Standard protocol, also called standard treatments, uses a prescriptive set of procedures that can be implemented in a given domain, which is supported by research as being effective for low-performing students. Most standard protocol instruments are
scripted and offer better quality control and integrity of instruction (Gresham et al., 2005).

Renaissance Learning (2010) suggests that standard protocol methods allow for decisions about placement to be made within fewer meetings, and fewer resources are required to meet student needs. Stecker et al. (2008) also claim that professional development for instructional delivery is usually easier to manage with a standard protocol approach. Professional development is a challenging issue for RTI according to Mokhtari et al. (2010), who suggest that RTI is still misunderstood by many classrooms and reading clinics in the real world.

To our knowledge, we have little or no evidence with respect to whether (a) reading teachers have sufficient knowledge of RTI and its intent, structures, and challenges, (b) they are adequately prepared to design and implement expert, multitiered reading instruction that is responsive to student needs, and (c) they have the expertise and resources to systematically document the potential effectiveness of RTI relative to its dual purposes of preventing serious reading problems and placing students in special education services. (p. 692)

Stecker et al. (2008) also suggest that schools in rural areas with limited financial means may be more successful with the standard protocol approach due to limited resources or lack of an available pool of candidates to hire as intervention specialists.

In a problem-solving approach, the instruction is designed by teachers or a student support team based on targeted needs from universal screenings or progress monitoring. Renaissance Learning (2010) states that problem-solving approaches require teachers and specialists to do additional diagnostic testing and hold multi-staff meetings to analyze a student’s deficits and plan individualized intervention strategies. The quality of this instruction is dependent both on the explicit design of the intervention and the structured delivery of the
intervention. Noell, Duhon, Gatti, and Connell (2002) suggest academic and behavior interventions developed within a problem-solving approach can be reliably implemented if follow-up and support procedures such as daily or weekly performance feedback are firmly established. This approach is challenging, however, since the successful implementation of RTI requires extensive knowledge and skill on the part of classroom teachers (National Joint Committee on Learning Disabilities, 2005).

Graney (2008) asserts that teachers may be unaware of effective reading instruction, and Moats (2007) indicates that teachers may engage in practices with non-existent research. Fuchs et al. (2003) claim that practitioners infrequently evaluate the outcomes of the problem-solving approach and generally fail to produce evidence of fidelity and improved achievement. Intervention instruction may be based on convenience or popularity rather than based on researched evidence of its effectiveness. More specifically, Fuchs et al. (2003) report consistent findings that teachers conducting interventions without follow-up support typically deliver poor intervention instruction. The problem-solving approach allows multiple types of interventions to be employed, and schools must be able to verify that these teacher-designed interventions are empirically validated. RTI delivered with a problem-solving approach is possible but places more responsibility on school administrators to ensure the precision and intensity of intervention efforts and the effectiveness of the ongoing process of interpreting time-series intervention data (Barnett, Daly, Jones, & Lentz, 2004). Some schools employ various combinations of both approaches.

**Response to Intervention Components**

NCRTI lists four components of the RTI model, including (a) a multilevel prevention system, (b) universal screening, (c) progress monitoring, and (d) data-based decision making
Throughout the literature, authors agree on these core elements, although they may be organized differently. Some RTI models may also contain additional sublevel components such as high-quality, standards-aligned instruction, parental involvement, administrative support, and professional development.

**Multilevel prevention system.** The RTI model is designed in a pyramid shape with sections to reflect the tiered levels of intervention. Students’ data determine where they are placed in a Tier. Brown-Chidsey and Steege (2010) suggest that about 80% of students are able to be successful with Tier 1 instruction alone, with the remaining 20% of students needing additional tiered interventions.

**Tier 1.** Whether the model is three-, four-, or five-tiered, instruction always begins in the general education classroom with a high-quality, standards-based core program, which is known as Tier 1. Since RTI data can potentially be used for LD identification, schools must be able to verify that they are implementing empirically-tested core programs and that their instructional practices are not contributing to students’ lack of achievement. The core instruction should be research-based, have an accountability structure for high teacher fidelity, and should incorporate differentiation to meet the needs of a broad range of students (Stecker et al., 2008). Research-based curriculum refers to curriculum in which the methods, content, materials, delivery, etc., were developed using guidance from the collective research and scientific community (Georgia Department of Education, 2009).

Universal screeners are given at the beginning of the school year to identify a student’s current level of achievement in relation to a performance standard (Baker et al., 2010). However, Compton, Fuchs, Fuchs, and Bryant (2006) advise that students who perform low on initial screeners may not need the additional support of Tier 2, but rather may need to be given
time to respond adequately to the core program. Their research suggests that general education classroom teachers who monitor closely those students who are low-performing on initial screening tests may see rates of improvement over a period of several weeks and ultimately determine that those students are responding well to the core program without the need for Tier 2 interventions. Clements and Kratochwill (2008) advise that if adequate growth is not seen, the first step is to rule out poor instruction or implementation of the core program as a probable cause. If the instruction is high-quality and a student has not made adequate progress, RTI calls for the instruction to be changed in some manner, such as increasing instructional intensity or targeting a specific weakness, so that the student has another opportunity to respond positively to the Tier 1 core instruction.

The National Association of School Psychologists (2009) employs data from the mental health field to estimate that about 80% of the students in a given school should be successful with a well-implemented, evidence-based core curriculum in Tier 1. In addition, Renaissance Learning (2010) and the Georgia Department of Education (2008) suggest that when fewer than 80% of a school’s students are performing successfully at Tier 1, their poor performance may indicate a problem with the core curriculum.

Stecker et al. (2008) urges school administrators to address the professional development on effective use of materials and practices beginning with the Tier 1 implementation. Without a pervasive knowledge and implementation of Tier 1 instruction, additional interventions (Tiers 2 and 3) will not reach their maximum potential (Georgia Department of Education, 2009). Vaughn and Chard (2006) state that initial professional development alone may be insufficient and should be followed with ongoing support and coaching. Schools should both assure the fidelity of selection of core programs as well as substantiate the delivery and monitoring of the
instruction, usually documented by classroom observations.

The Georgia Department of Education provides five non-negotiables for Tier 1 instruction. Standards-based classrooms for all students in general education shall provide: (a) universal screenings or benchmarks conducted at the school level, (b) evidence-based curricula and strategies in place for all students, (c) differentiation documented by general education teachers through the general education environment, (d) at-risk students identified in an area of instructional delay (language, academics, behavior), and (e) data analyzed by classroom general education teachers for decision-making (Togut, 2012).

**Tier 2.** Students who do not respond adequately to the core instruction are given needs-based interventions at the secondary level of the RTI model, Tier 2. The primary differences between Tier 1 and Tier 2 are group size, instruction interventions of moderate to heavy intensity, and increased feedback from teachers (Baker et al., 2010). The RTI framework allows for the fluid movement in and out of the different levels of instruction as needed (Stecker et al., 2008). On average, about 15% of students will need Tier 1 plus Tier 2 instruction to be successful (Brown-Chidsey & Steege, 2010). “Tier 2 . . . instruction is supposed to directly supplement (not replace) Tier 1 instruction” (Vaughn, Wanzek, Woodruff, & Linan-Thompson, 2007, p. 19). Tier 2 interventions are usually accomplished through small groups in which teachers model the skill, provide multiple practice opportunities, correct and address student errors, and instruct students with fast-paced lessons to increase student engagement (Baker et al., 2010).

Fuchs et al. (2008) state that Tier 2 services are provided on a frequent basis, either every day or several days per week. In 2009, Gersten et al. published research that suggested that Tier 2 interventions are highly effective if the instruction is approximately 30 minutes of small-group
instruction three to five days per week. The literature reflects flexibility in the duration of Tier 2 intervention groups, which is usually determined by the RTI model for a school district. “Intervention length can vary due to publisher recommendations, research data, student scheduling factors, availability of resources and student attendance” (Wedl, 2005, p. 15). Interventions that are too short will have minimal impact, and interventions that are too long, without gains, will delay instructional changes. By continually evaluating progress-monitoring data, educators can make informed decisions regarding how long interventions are needed. Fuchs and Fuchs (2008) suggest six to eight weeks as an appropriate duration for Tier 2 intervention, while Vaughn and Denton (2008) report most schools schedule 10 to 30 weeks to allow for sufficient demonstration of progress.

The RTI framework allows for schools to vary in the number of tiers or rounds of supportive instruction that are provided prior to a special education referral. Stecker et al. (2008) recommend that students receive at least one round of preventive instruction in Tier 1, and one or two rounds of preventive instruction in Tier 2 or Tier 3 before being referred for a special education evaluation. Due to the flexibility allowed in RTI models, some may include four or five tiers with increasingly intensive instruction, although most literature reflects a three-tiered model. In this study, the state of Georgia four-tiered model is used.

The Georgia Department of Education provides five non-negotiables for Tier 2 instruction. Needs-based learning groups shall provide learning that is different by including: (a) parent notification that additional small-group instruction is needed for the student, (b) parent notification of strategies to be attempted, (c) small-group instruction provided in addition to the core curriculum, (d) progress monitoring administered frequently to determine whether a change in strategies is needed, and (e) data analyzed by classroom general education teachers for
decision-making (Togut, 2012).

**Tier 3.** If a student’s progress is insufficient in Tier 2, the student can be moved to Tier 3 for additional instruction with even higher intensity and frequency. In Tier 3, a student support team (SST) is initiated to develop an educational plan with continuous monitoring and evaluation (Georgia Department of Education, 2008). The SST’s Tier 3 role is to investigate why the student has continued to struggle regardless of Tier 2 interventions. This team uses a collaborative problem-solving approach to correct or bypass learning problems, which promotes individualized interventions which are more stringent than the lower tiers. If students are satisfactorily helped by the Tier 3 SST analysis and intervention, they can return to Tier 2 or Tier 1 once their continued progress allows them to maintain academic success (Georgia Department of Education, 2009).

The Georgia Department of Education provides four non-negotiables for Tier 3 instruction. SST-driven learning experiences shall provide instruction that is different by including: (a) baseline and progress monitoring data from Tier 2 analyzed to create specific goals for student improvement, (b) SST determinations for the need of additional information on a student, including the use or administration of informal and formal measures to gather individual data in the area of concern, (c) continued SST interventions if the student is making progress; however, if progress is minimal, SST members will revise the intervention, and (d) SST determination on whether referral to Tier 4 for a special education evaluation is needed (Togut, 2012).

**Tier 4.** Students who demonstrate minimal gains after receiving high-quality instruction during Tiers 1, 2, and 3 are provided Tier 4 interventions. At this level, students are served who exhibit the most severe reading difficulties and are sometimes referred to as nonresponders due
to their inadequate response to interventions (Wanzek & Vaughn, 2010). According to Reschly (2005), instruction and measurement at this level indicate intervention groups that are either individualized or very small groups with intensified instruction, increased instructional time, and ongoing progress monitoring.

Tier 4 intervention, normally for about 5% of the students, is typically the result of a comprehensive evaluation for special education (Brown-Chidsey & Steege, 2010). The special education referral should be made only after sufficient time has passed with high-quality, evidence-based instruction as well as documented data indicating that the student continues to have severe difficulties (Bursuck & Blanks, 2010). Stecker et al. (2008) caution, however, that since the last tier provides unique services on an individualized basis, Tier 4, in this RTI model, is special education. They suggest that poor response in Tier 3 triggers the need for a special education evaluation.

Stecker et al. (2008) also maintain that it is problematic for general education teachers to deliver increasingly more individualized instruction. Individualization is the essence of special education services delivered by trained professionals, and continued preventive instruction without successful academic growth may deny students with true disabilities their rights provided through the IDEA legislation. School administrators are obligated under the legal parameters of IDEA to ensure that evaluations of children suspected of having a learning disability are not delayed or denied due to any unfinished RTI strategies or due to RTI not being fully implemented in a school (Togut, 2012).

Wanzek and Vaughn (2010) maintain that Tier 4 is distinct from the other levels in that: “(a) participating students have more severe difficulties and for some students life-long difficulties, (b) more intensive interventions are provided, and (c) it is expected that [Tier 4]
instructors are professionally trained and demonstrate very high levels of expertise and knowledge” (p. 306).

The Georgia Department of Education provides two non-negotiables for Tier 4 instruction. A specially designed learning environment shall provide learning that is different by including: (a) specialized programs, methodologies, or instructional deliveries and (b) an even greater frequency of progress monitoring to assess a student’s response to intervention(s) (Togut, 2012).

In addition, to validate that a child suspected of having a learning disability has been adequately exposed to a rich learning environment and has been offered multiple opportunities to respond to instruction, the special education evaluation must include: (a) data that demonstrates that prior to, or as part of, the referral process, the child was provided appropriate instruction in regular education settings, delivered by qualified personnel, and (b) data-based documentation of repeated assessments of student progress during instruction. This documentation must also have been provided to the child’s parents. A special education evaluation must include multiple sources of evidence concluding that a student exhibits a pattern of strengths and weaknesses relative to age, grade level, and intellectual development (Togut, 2012).

A multilevel RTI model is displayed in Figure 1.
Figure 1. Georgia Student Achievement Pyramid of Interventions. From “Response to Intervention: Georgia’s Student Achievement Pyramid of Interventions.” Copyright 2011 by Georgia Department of Education. Reprinted with permission (see Appendix C).

Universal screening. In addition to the tiered framework, universal screening is essential to RTI as it attempts to find students in danger of academic failure so that prevention services can be implemented without delay. RTI requires the systematic evaluation of all students to identify those who are: “(a) making adequate progress, (b) at some risk of failure if not provided extra assistance, or (c) at high risk of failure if not provided specialized supports” (Togut, 2012, Response to Intervention (RTI) section, para. 1). Shapiro (2012) defines universal screening as
“periodic windows into student performance by comparing against the performance of peers” (p. 9). Shapiro describes these benchmark assessments as brief, inexpensive, easily administered, and easily scored. The Georgia Department of Education (2009) suggests that universal screenings are not designed to identify why students are underperforming, but rather to target students who are not exhibiting the expected performance criteria for a given grade level.

Universal screening cannot be used alone to identify at-risk students due to the understanding that more students will be identified as at-risk than actually are and that some students who are at-risk may not be identified. Fuchs and Fuchs (2008) recommend flagging students below the 50th percentile as suspected at-risk “so that no student who might develop severe problems is missed” (p. 31) although most RTI models have a lower percentile cutoff score for Tier 2. The Georgia Department of Education (2009) recommends that performance expectations be set in advance by POI teams.

The POI team for the school district in this study was formed at the onset of the district RTI implementation for several reasons. The lead psychologist, who had received the most POI/RTI training, needed an avenue to share information and communicate RTI goals, as well as convey new pyramid of intervention special education eligibility requirements. At the time, there was also an emphasis in the district on Professional Learning Communities for collaboration and sharing of data. Over time, administrators realized that a district-wide POI team could provide systematic, consistent RTI initiatives in developing interventions and collecting data. Therefore the district POI Professional Learning Community was formed, facilitated by the lead school psychologist and consisting of school administrators who were responsible for RTI, SSTs, and Special Education services.

The instrument for universal screening in this study was STAR by Renaissance Learning.
This screener, adopted by the school district POI team, was selected based on published research data. NCRTI identifies AIMSWeb and STAR as the two universal screening and progress monitoring tools with the strongest convincing evidence of mastery measurements (NCRTI, 2012). As of the 2011 updated review, both of these instruments scored equally on reliability and validity. In 2008, Renaissance Learning restandardized STAR using a population of 69,738 students. According to Salvia et al. (2012), reliability was tested using scaled scores and instructional reading levels. “Test-retest reliabilities ranged from .85 to .95 for scaled scores and averaged .91 for instructional reading level” (p. 290). Salvia et al. also confirmed that the 2008 standardization showed high validity because performance on STAR was closely related with performance on a number of different standardized measures. STAR, the dependent variable used in this study, recommends below the 40th percentile as the cutoff score for Tier 2, although school systems are allowed to choose a different cutoff score. In this study, the 40th percentile was the cutoff score for receiving the reading intervention.

Typically, universal screening is done in consistent intervals across the school year (beginning of fall, winter, and end of spring) (Renaissance Learning, 2010). Baker et al. (2010) assert that since universal screeners are administered at the beginning, middle, and end of the year, this data can be used to analyze how well the multiple tiers of instruction have

(a) decreased the percentage of students at the highest level of reading risk within and across school years, (b) increased the percentage of students who reach grade level reading goals, and (c) increased strong rates of student reading growth and improvement over a specified time period. (p. 12)

**Progress monitoring.** RTI requires continuous progress monitoring on a regular and frequent basis for the identification of student growth trends that reveal increasing instructional support is
needed (Togut, 2012). Stecker et al. (2008) report that progress monitoring differs from universal screenings and benchmark assessments in terms of “duration of assessments, frequency of administration, consistency and equivalence of content assessed, and usefulness of information for determining both level and rate of student academic growth” (p. 11). They also argue that progress monitoring is crucial to the RTI model because (a) it aids teachers in making instructional decisions, and (b) it provides data for LD identification. Buffum et al. (2009) agree that progress monitoring is vital for assessing students’ academic performance and for monitoring and evaluating the effectiveness of the instruction. Shapiro (2012) claims that to be applicable and useful in planning instruction, progress monitoring must be ongoing, frequent, flexible, and consistent with the measuring instrument.

Buffum et al. (2012) recommend simplifying progress monitoring so that it is not too broad or does not create overly demanding documentation. They suggest that intervention groups need to focus on specific targets rather than diverse needs so that teachers know explicitly what to teach and monitor. In addition, curriculum-based measures and/or computer-adaptive tests are especially useful in simplifying the measuring required for progress monitoring. Both of these methods provide a similar process using Internet software for progress monitoring and allow immediate feedback regarding the student’s responsiveness to the intervention. Whether from computer programs or more traditional methods, progress monitoring data is normally plotted on a graph to compare the student’s rate of growth with the grade level goal over time. A line-of-best-fit is superimposed on the data to show the student’s actual rate of improvement and compared to the long-term goal to predict whether or not the student is likely to meet long-term expectations. If a student’s growth rate is not sufficient, the teacher’s decision should be to change the instruction to try to improve achievement a different way (Stecker et al., 2008).
In addition to monitoring low performing students, progress monitoring also serves another function in the RTI framework. Stecker et al. (2008) indicate that when progress monitoring is administered to all students, although not as frequently as with at-risk students, it serves as a useful tool for evaluating the quality of the instructional program and the effectiveness of individual teachers. When data indicate that most students have either made adequate progress or have made little change over a period of time, school administrators are provided information to make accommodations and modifications in curriculum and/or teacher professional development to help improve the overall effectiveness of the instructional program. Also, teachers are provided data to determine whether individual students and/or all students are progressing as they should and have information to better accommodate academic diversity.

Regardless of the number of tiers and duration of the intervention, “progress monitoring data [throughout the tiered process] are critical for determining overall student unresponsiveness to instruction. Progress monitoring aids in eliminating the lack of effective instruction as a contributing factor to the student’s learning problems” (Stecker et al., 2008, p. 13).

**Data-based decision making.** Research confirms the significant effects on student achievement when teachers use progress-monitoring data to formatively devise instructional programs best suited to the individual needs of students (Stecker, Fuchs, & Fuchs, 2005). Togut (2012) maintains that instructional decision making is at the heart of the RTI process, which modifies instruction when student performance on curricular outcomes signals insufficient growth. The Georgia Department of Education (2008) promotes the use of school data teams for instructional decision making. Data teams are trained to use RTI data for monitoring growth and for targeting areas of needed improvement throughout the year. Data teams work to address specific issues related to underperformance areas and are the driving force for implementing interventions that
Impact achievement.

Curriculum-based measurement (CBM) and computer-adaptive testing (CAT) are two assessment systems which have received strong research support as tools for gathering student data and providing graphs for RTI documentation requirements. For many years, the CBM model of assessment “has been most closely associated with RTI models” (Shapiro, 2012, p. 6). However, Shapiro suggests that the CAT model is currently emerging as the foundation for the next generation of statewide assessment measures. CATs focus on skills (number correct) within various domains of academic areas. CATs have a vast bank of test questions, and students are asked questions based on their accuracy of answering prior items. Renaissance Learning uses the CAT model for their STAR tests. “Because the measure adjusts the difficulty of the items presented to students depending on the accuracy of their responses, the items answered correctly reflect a broad range of skills acquired by students” (Shapiro, 2012, p. 15). Reports produced in the CAT model suggest targets for instruction based on a very large database of student growth patterns, and also indicate where a student’s skills fall across a longitudinal scale spanning from kindergarten through grade 12. The CAT assessment system accomplishes RTI’s goals of sound decision making and targeted instruction based on good data while also providing the specifics of instructional planning as part of the routine assessment process (Shapiro, 2012). “Although sound instruction is paramount to successful implementation of RTI, assessment data should drive decision making” (Stecker et al., 2008, p. 10).

Existing Research

Multiple studies in existing literature indicate that instructional interventions in a timely manner produce significant reading improvements. Vellutino et al. (1996) report findings from a longitudinal study in which two thirds of the sample caught up to their peers in reading by
explicit intervention instruction, suggesting that these students were instructional casualties and did not have processing deficits.

Another study by Torgesen et al. (2001) reports that about 40% of LD students in their sample were returned to general education, no longer in need of special education services, after completing intensive remedial instruction. This study provides evidence that specific interventions can be used to remediate reading difficulties in many (but not all) young students.

Gresham and Vellutino (2010) completed a five-year longitudinal study of struggling readers in 2008 in which RTI was tracked from kindergarten through fourth grade. Students were assessed on measures of reading achievement and reading-related cognitive abilities. This study suggests that a student’s ability to consolidate and maintain reading achievement gains through intervention may be a more valid criterion in predicting long-term reading growth than the IQ-achievement discrepancy.

Graves et al. (2011) suggest through a Tier 2 literacy instruction study that RTI reading interventions have positive effects on struggling readers. Sixty sixth-grade students designated as far below basic and below basic based on the California English Language Development Test were participants in a study. The control group received typical language arts instruction, and the experimental group received small-group instruction three days per week using evidence-based materials designed for Tier 2 of the RTI model. Groups of 3 were instructed for 3 hours per week over 10 weeks in decoding, phonemic awareness and phonics, fluency, comprehension, and vocabulary development. Results of the study indicated that although both control and experimental groups improved during the intervention period, the experimental group improved 30% more than the control group in fluency, with a gained average of 10 words per minute in 10 weeks compared to 0 words per minute in 10 weeks for the control group. This growth is even
more significant given consistent findings of success in fluency as a predictor of reading comprehension (Kim, Petscher, Schatschneider, & Foorman, 2010; Yovanoff, Duesbery, Alonzo, & Tindal, 2005). Another study suggested that average sixth graders might be expected to grow at a rate of less than one word per minute per week (Silberglitt & Hintze, 2007). Clearly, students in the Tier 2 study experimental group far exceeded this expectation.

**Effective Instruction**

The underlying foundation to all of the RTI components is evidence-based instruction taught with high-quality fidelity. The National Reading Panel (2000) suggests that the absence of evidence-based practices has serious consequences as seen in studies by Lyon (1998) and Juel (1988). Their research indicates that without evidence-based instruction, 30-60% of students may fall behind in reading, and once behind, are unlikely ever to catch up. Stichter, Stormont, and Lewis (2009) report that students who are most at risk for reading failure are the least likely to receive effective instruction.

Evidence-based reading instruction includes phonemic awareness, phonics, fluency, vocabulary, and reading comprehension (National Reading Panel, 2000; National Early Literacy Panel, 2008). Phonemic awareness, the ability to hear and manipulate sounds, is a predictor of decoding ability. Students at risk need explicit instruction in segmenting and blending sounds. Moats (2007) suggests that some general education reading programs treat phonemic awareness instruction incidentally rather than as an essential component.

Phonics, sound-symbol relationships in a sequence, facilitates automatic word decoding and is applied and practiced with decodable reading materials. Bursuck and Damer (2007) indicate that phonics helps children learn the alphabetic principle, which is the understanding that there are predictable relationships between letters and sounds. Students at risk for reading
may have a phonics deficiency and need to be provided explicit phonics instruction (Moats, 2007).

Fluency, the ability to read text accurately, rapidly, and with expression, is taught by providing guided repeated oral reading activities (Therrien, Gormley, & Kubin, 2006). Fluency enables readers to devote their attention to comprehension rather than decoding. Bursuck and Damer (2007) suggest that many general education reading programs provide insufficient practice in reading fluency to address the needs of struggling students.

Vocabulary knowledge accelerates comprehension (August, Carlo, Dressler, & Snow, 2005). Due to lack of literacy exposure, at-risk students are often deficient in vocabulary and need direct teaching of new words and strategies for interpreting word meanings (Stahl & Nagy, 2006). Coyne, McCoach, and Kapp (2007) suggest that teaching specific vocabulary words to at-risk students is appropriate and highly effective although many general education reading programs merely mention and assign words without direct vocabulary instruction (Beck & McKeown, 2007).

Comprehension, understanding the meaning of the text, includes (a) activating background knowledge, (b) questioning, (c) drawing conclusions, (d) predicting, (e) summarizing, (f) awareness of understanding or not understanding, and (g) using text structures to grasp the meaning of the text (National Reading Panel, 2000; Baumann, Seifert-Kessell, & Jones, 1992; Armbruster, Lehr, & Osborn, 2001; Boulware-Gooden, Carreker, Thornhill, & Joshi, 2007; & Dymock, 2007). Moats (2007) indicates that the general education program must explicitly teach, model, and practice the structures of narrative and expository text in a planned progression. Bursuck and Blanks (2010) suggest that it is not enough for general education reading programs to mention comprehension strategies, but rather, empirically based
instructional designs incorporating systematic and explicit teaching are necessary for students to acquire these skills.

The five recommended components for evidence-based instruction (National Reading Panel, 2000; National Early Literacy Panel, 2008) are found in the reading intervention curriculum examined in this study. The instruction materials from How to Plan Differentiated Reading Instruction: Resources for Grades K-3 (Walpole & McKenna, 2009) are specifically designed for reading interventions to be taught in addition to the core curriculum. In this study, the areas of phonemic awareness, phonics and word study, and fluency were learning goals in the treatment intervention groups.

**Reading Curriculum Intervention**

This study investigated a new curriculum in the RTI literature specifically designed for reading interventions, which was authored by Sharon Walpole and Michael C. McKenna. In 2007, Walpole and McKenna authored a book entitled Differentiated Reading Instruction: Strategies for the Primary Grades, which provided “simple instructional strategies with strong research pedigrees” (Walpole & McKenna, 2009, p. 1) and which targeted specific skills for small group instruction. After much feedback from principals, literacy coaches, and teachers who used Differentiated Reading Instruction: Strategies for the Primary Grades for intervention curriculum, the authors revised the curriculum to provide even more explicit lessons for reading instruction. In addition, the revised curriculum provided strategies for instructional planning so that teachers could employ strategies in creating their own differentiated reading instruction based on student needs. The revised curriculum is provided in the authors’ 2009 book, How to Plan Differentiated Reading Instruction: Resources for Grades K-3, which is the RTI curriculum examined in this study. One of the authors, Michael C. McKenna, refers to this curriculum
simply as Differentiated Reading Instruction (DRI), and so for the purposes of this study, the treatment curriculum is referred to as DRI. The STAR Diagnostic Report, available with each administration of a test, provides a Domain and Skill Area score with key ideas and details, which can be directly matched to a DRI target. This data, provided by STAR, enables teachers to choose the appropriate DRI cycle for students that will provide them explicit instruction for a specific area of weakness.

Reading instruction for this study included phonemic awareness, phonics and word study, and fluency in seven differentiated instruction cycles as follows: (a) basic alphabet knowledge, (b) using letter sounds, (c) using letter patterns, (d) blends and digraphs, (e) R-controlled vowels, (f) vowel-consonant-E, and (g) vowel teams. Fluency passages were incorporated into the last four of these cycles. Students were placed in a cycle based on weaknesses indicated by a phonics inventory. Reading cycles were 15 days each with the exception of vowel teams, which was 30 days. The scripted instruction was designed for 15 to 20 minutes. Students in Tier 2 received this instruction three or four times per week, and students in Tier 3 received this instruction four or five times per week, depending on the model adopted by individual schools.

DRI was chosen because it was promoted by the Georgia Department of Education. Walpole and McKenna assisted the state of Georgia in implementing a Reading First Grant from 2004 - 2007 by serving as professional development architects. Since that time, Walpole and McKenna have periodically provided professional development to Georgia educators. The school district in this study was focused on providing a research-based reading curriculum for students that was consistent county-wide for RTI reading interventions. Results from this study provide data regarding the effectiveness and possible future use of this curriculum.
CHAPTER THREE: METHODOLOGY

The Georgia Department of Education requires a fully implemented RTI framework in all public schools to assist students with learning difficulties (Georgia Department of Education, 2008). IDEA legislation requires that interventions taught in this model must be research-based and show evidence of improving student outcomes. One RTI curriculum for reading in the early grades is found in the book *How to Plan Differentiated Reading Instruction: Resources for Grades K-3* (Walpole & McKenna, 2009). This curriculum, referred to as DRI in this study, is recommended by the authors and by the Georgia Department of Education as an appropriate intervention for the RTI model. Since this material is new to the RTI literature, research is needed to support the use of this curriculum. This study investigated the impact of the DRI reading curriculum on the STAR scaled scores for second- and third-grade students in a school district in northeast Georgia. Scaled scores were used because they equally compare levels of performance across different versions of a test, and Renaissance Learning states that scaled scores show absolute growth over time (Tan & Michel, 2011; Renaissance Learning, 2010). The study investigated the following research questions:

Research Question 1: While using 2012 STAR beginning-of-year scaled scores as a control variable for differences in achievement, do at-risk second-grade students who receive Differentiated Reading Instruction (DRI) through RTI have significantly different scaled scores on the 2013 STAR end-of-year test when compared to at-risk students who do not receive this intervention?

Research Question 2: While using 2012 STAR beginning-of-year scaled scores as a control variable for differences in achievement, do at-risk third-grade students who receive Differentiated Reading Instruction (DRI) through RTI have significantly different scaled scores
on the 2013 STAR end-of-year test when compared to at-risk students who do not receive this intervention?

The corresponding null hypotheses were:

\( H_{o1} \): While using 2012 STAR beginning-of-year scaled scores as a control variable for differences in achievement, at-risk second-grade students who receive Differentiated Reading Instruction (DRI) through RTI will not have significantly different scaled scores on the 2013 STAR end-of-year test when compared to at-risk students who do not receive this intervention.

\( H_{o2} \): While using 2012 STAR beginning-of-year scaled scores as a control variable for differences in achievement, at-risk third-grade students who receive Differentiated Reading Instruction (DRI) through RTI will not have significantly different scaled scores on the 2013 STAR end-of-year test when compared to at-risk students who do not receive this intervention.

Chapter 3 explains the research design, the participants, the setting, and treatment fidelity of the study. The testing instrument is described, showing validity and reliability data of the dependent variable. The chapter concludes with an explanation of the procedures and data analysis that were used for the study.

**Research Design**

The causal-comparative research design was used in this study. According to Gall, Gall, and Borg (2007), although this research design “[does] not permit strong conclusions about cause and effect, [it is] useful for initial exploratory investigations or in situations where it is impossible to manipulate the independent variable” (p. 306). Exploratory research on this curriculum is defensible because DRI was new both to RTI literature and to the school district that implemented it. The school district in this study already had an RTI framework in place. In Spring 2012, the district POI team selected the RTI reading curriculum presented in Walpole and
McKenna’s 2009 book, *How to Plan Differentiated Reading Instruction: Resources for Grades K-3*, noted as DRI in this study, as an option for educators to use in the 2012-2013 school year. School RTI coordinators chose whether or not their school would use DRI for teaching Tier 2 and Tier 3 at-risk students. Participants for Tier 2 and Tier 3 interventions were chosen according to their percentile rank (PR) on the Fall 2012 administration of the STAR Test. Random assignment of students to receive DRI was not possible by the researcher, and school administrators determined the curriculum and structure for teaching at-risk students during the school year prior to the beginning of the study, thus justifying the use of the causal-comparative design.

In analyzing the data from the 2012-2013 STAR scores, a treatment group and control group were used. The control group served to lower the potential threat to validity posed by the statistical regression threat of using students with low test scores and the selection threat to validity posed by using nonequivalent groups.

**Participants**

This study used a convenience sample. The participants were second- and third-grade students identified as being at-risk for reading and enrolled in any one of four elementary schools in one school district in northeast Georgia. At the beginning of the 2012-2013 school year, all second- and third-grade students in the district were universally screened using the STAR Computer-Adaptive Test (CAT) by Renaissance Learning. Students who scored below the 40th percentile on this screener were identified as at-risk students. Teachers placed at-risk students in an intervention group based on their STAR percentile rank, as well as available space in the scheduled allotment of intervention groups. Students in an intervention group were taught with either the treatment curriculum, DRI by Walpole and McKenna, or some other reading
intervention curriculum determined by the school they attended and according to results from a phonics inventory showing evidence of need.

At-risk students who received DRI in addition to traditional reading instruction were the participants of the treatment group. At-risk students who did not receive DRI in addition to traditional reading instruction were the participants of the control group. Any student who was part of an intervention group but transferred to another school during the school year was excluded from the sample population.

The sample size for this study began with 97 participants. The treatment group consisted of 53 students (54.6%), and the control group consisted of 44 students (45.4%). The gender of the sample consisted of 45 male students (46.4%) and 52 (53.6%) female students as seen in Table 3.1. The grade level of the sample consisted of 62 (63.9%) second-grade students and 35 (36.1%) third-grade students as seen in Table 3.2.

Table 3.1

Gender of Sample

<table>
<thead>
<tr>
<th></th>
<th>Entire Sample</th>
<th>Control Group</th>
<th>Treatment Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>45 (46.4%)</td>
<td>16 (16.5%)</td>
<td>29 (29.9%)</td>
</tr>
<tr>
<td>Female</td>
<td>52 (53.6%)</td>
<td>28 (28.9%)</td>
<td>24 (24.7%)</td>
</tr>
<tr>
<td>Total</td>
<td>97 (100.0%)</td>
<td>44 (45.4%)</td>
<td>53 (54.6%)</td>
</tr>
</tbody>
</table>

*Table 3.1.* (Infinite Campus, 2012).
Table 3.2

*Grade Level of Sample*

<table>
<thead>
<tr>
<th></th>
<th>Entire Sample</th>
<th>Control Group</th>
<th>Treatment Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Second Grade</td>
<td>62 (63.9%)</td>
<td>22 (22.7%)</td>
<td>40 (41.2%)</td>
</tr>
<tr>
<td>Third Grade</td>
<td>35 (36.1%)</td>
<td>22 (22.7%)</td>
<td>13 (13.4%)</td>
</tr>
<tr>
<td>Total</td>
<td>97 (100.0%)</td>
<td>44 (45.4%)</td>
<td>53 (54.6%)</td>
</tr>
</tbody>
</table>

*Table 3.2.* (Infinite Campus, 2012).

The sample group mirrored the ethnicity of the total school district elementary school population with the majority of students being Caucasian, as seen in the next section. The ethnicity of the sample consisted of 73.2% Caucasian, 19.6% African American, 5.2% Hispanic, and 2.0% Asian. The breakdown of the entire sample disaggregated by ethnicity is seen in Table 3.3.

Table 3.3

*Ethnicity of Sample*

<table>
<thead>
<tr>
<th></th>
<th>Entire Sample</th>
<th>Control Group</th>
<th>Treatment Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caucasian</td>
<td>71 (73.2%)</td>
<td>31</td>
<td>40</td>
</tr>
<tr>
<td>African American</td>
<td>19 (19.6%)</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Hispanic</td>
<td>5 (5.2%)</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Asian</td>
<td>2 (2.0%)</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>97 (100.0%)</td>
<td>44 (45.4%)</td>
<td>53 (54.6%)</td>
</tr>
</tbody>
</table>

*Table 3.3.* (Infinite Campus, 2012).
The sample also included 73 students (75.2%) who were considered economically disadvantaged.

**Setting**

A causal-comparative research design examined the scaled scores from the 2012 beginning-of-year STAR and the 2013 end-of-year STAR of at-risk students in second- and third-grades in a public school district in northeast Georgia. The school district had four elementary schools with the following characteristics: (a) student enrollment between 382 and 459; (b) each school received Title 1 funding (The Governor’s Office of Student Achievement, 2011); (c) each school employed a full-time academic coach; and (d) each school shared the same two school psychologists. The principal served as the RTI coordinator at two of the schools, and the assistant principal served as the RTI coordinator at two of the schools. Each school had a small student population identified as special education (from 9% to 11.6%). The schools served culturally similar populations consisting of a majority of white Caucasian (from 69% to 91%), and a minority of African American (from 3% to 15%), Hispanic (from 3% to 11%), and other (from 2% to 5%). The identified gifted students at the four schools ranged from 5.4% to 8%, and the identified Early Intervention students at the four schools ranged from 8.8% to 22% (Infinite Campus, September, 2012).

Because state legislation was passed in 2010 requiring Georgia public schools to adopt national Common Core standards by the fall of 2012, all certified teachers except two participated in collaborative training on the Common Core curriculum. One certified teacher was funded by a grant and served part-time, and another certified teacher served as a paraprofessional. These two teachers did not receive Common Core curriculum training. It is understood that Common Core training was ongoing throughout the year through grade level
meetings with academic coaches as the new standards were implemented in the school system. In this study, the Common Core curriculum provided the reading standards taught in Tier 1.

Twenty-nine credentialed teachers taught Tier 2 and Tier 3 reading groups. Twenty-eight teachers were female, and one teacher was male. All teachers were Caucasian. Twenty-four teachers had graduate degrees, and five teachers had bachelor’s degrees. Twenty-seven teachers had taught from four to 31 years, and two teachers had taught for one year.

Twenty-five second- and third-grade classes in these schools were used in the study. Three schools provided students for the treatment group, and all four schools provided students for the control group. All four schools provided traditional reading instruction using the Common Core standards. A variety of teaching resources were used, including a traditional Basal series, Reading First materials from a previous federal grant, framework lessons provided by the state of Georgia, and various other resources that teachers found appropriate for instruction. While the Common Core standards were required for lesson foundations, teachers were allowed to teach the standards with any resources they chose. The reading block at each school consisted of two sections – a block of time for traditional reading instruction and a block of time for reading intervention. The intervention instruction for this study took place during the reading intervention block, and descriptions for this instruction are seen in Table 3.4.

The treatment group intervention curriculum came from Walpole and McKenna’s 2009 book, *How to Plan Differentiated Reading Instruction: Resources for Grades K-3*. In this study, Differentiated Reading Instruction (DRI), a broad term in the education field, refers to this specific curriculum compiled and marketed by these authors to provide educators with a research-based framework for implementing differentiated instruction. A thorough description of DRI is found at the end of Chapter 2 under Reading Curriculum Intervention (see p. 51).
Table 3.4

*Reading Intervention Block Descriptions*

<table>
<thead>
<tr>
<th>Reading Groups</th>
<th># in Groups</th>
<th>Intervention</th>
<th># of Twenty Minute Sessions Per Week</th>
<th>Progress Monitor</th>
<th>Data Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tier 2 (Treatment)</td>
<td>6 or less</td>
<td>DRI</td>
<td>3 or 4</td>
<td>Biweekly</td>
<td>STAR</td>
</tr>
<tr>
<td>Tier 2 (Control)</td>
<td>6 or less</td>
<td>Variety of Other</td>
<td>3 or 4</td>
<td>Biweekly</td>
<td>STAR</td>
</tr>
<tr>
<td>Tier 3 (Treatment)</td>
<td>6 or less</td>
<td>DRI</td>
<td>4 or 5</td>
<td>Weekly</td>
<td>STAR</td>
</tr>
<tr>
<td>Tier 3 (Control)</td>
<td>6 or less</td>
<td>Variety of Other</td>
<td>4 or 5</td>
<td>Weekly</td>
<td>STAR</td>
</tr>
</tbody>
</table>

*Treatment Fidelity*

To ensure the equality of treatment instruction, the POI team administered training to RTI coordinators and academic coaches at the beginning of the 2012 school year. The RTI coordinators and academic coaches relayed this information to teachers administering the treatment intervention. The assistant principal at one school and the academic coaches at two schools monitored the treatment implementation. Teachers were provided with identical sets of DRI curriculum, which included scripted lesson plans, student worksheets, and the book *How to Plan Differentiated Reading Instruction: Resources for Grades K-3* (Walpole & McKenna, 2009).

Certified elementary school teachers administered all treatment interventions. Tier 2 students received intervention instruction three or four days a week, and Tier 3 students received intervention instruction four or five days a week depending on school schedules. Groups ranged in size from five to six students and instruction time involved 15 to 20 minutes per lesson. The curriculum provided seven sections of instruction involving phonemic awareness, phonics and
word study, and fluency with an assessment at the end of each section. Students were placed in a section based on results from a phonics inventory administered by either a teacher or an academic coach. When a 15-lesson section of the treatment was completed, teachers, academic coaches, and/or RTI coordinators determined through progress monitoring data whether students needed to remain in a treatment group for continued interventions or could return to a lower Tier because they either needed less intensive interventions or no longer needed interventions. Since placement in a group was fluid based on a student’s progress monitoring results, some students tested out of the intervention treatment during the middle of the year while other students entered the intervention treatment during the middle of the year. Some students remained in the treatment intervention group for the entire year. Intervention instruction was delivered in addition to Tier 1 instruction.

This treatment fidelity information was confirmed in three face-to-face interviews with one assistant principal and two academic coaches regarding the implementation of the treatment intervention at their school. Based on the interviews and the above information, it was concluded that the implementation of the treatment across the three schools was similar.

Instrumentation

STAR served as the control variable and as the dependent variable. The Fall 2012 beginning-of-year scaled scores were the covariate to adjust for pre-existing differences in the participants at the onset of the study. The Spring 2013 end-of-year scaled scores served as the dependent variable.

STAR is a computer-adapted test (CAT) that compares students’ reading achievement to that of students across the nation (Renaissance Learning, 2010). NCRTI reports that STAR has the highest rating of all screening assessments and is among the highest rated progress
monitoring assessments (NCRTI, 2012). Available data collected from STAR includes a scaled score (SS), a grade-equivalent score, a percentile rank (PR), a normal curve equivalent, an instructional reading level, a zone of proximal development (ZPD), and an estimated oral reading fluency score (Renaissance Learning, 2010).

This study used STAR percentile rank and scaled scores. The percentile rank is the proportion of scores in a distribution to which a specific score is greater than or equal. This means that a student with a percentile rank of 30 performs equal to or better than 30 percent of students of the same grade at the same time of year nationwide (Renaissance Learning, 2010). All students who scored below the 40th percentile on the Fall 2012 beginning-of-year STAR were identified as needing reading interventions. The scaled score provides equal comparison of performance level on different forms of a test and shows absolute reading growth over time. In STAR, scaled scores range from 200 to 900 and are used to divide students into the following categories:

- At/Above Benchmark = At/above 40th percentile;
- On Watch = Below 40th percentile;
- Intervention = Below 25th percentile;
- Urgent Intervention = Below 10th percentile (Renaissance Learning, 2010).

Computer-adaptive technology allows STAR to provide accurate data in a short amount of time. Comprehension is measured by test questions that require background knowledge, vocabulary, and semantic and syntactical skills. The psychometric test design for the assessment is the Item Response Theory based on the Rasch model. The test is individualized for each student and continually adjusts questions to the student’s reading ability. Questions in the assessment have been administered to large samples of students and calibrated to determine the
difficulty of each one. The statistical analysis relates the probability of a student’s correctly answering a question to the student’s ability and the difficulty of the question. The assessment uses a student’s pattern of correct or incorrect answers to provide a statistically reliable and valid estimate of the student’s reading ability (Renaissance Learning, 2010).

Reliability, consistent results over multiple measurements, is a requirement for a good assessment. Reliability is high for STAR. Since this assessment individualizes each test through the software, high levels of reliability are possible with fewer questions. The STAR Reading Instruction Manual also provides teachers with standardized administration directions to be used with each STAR administration. NCRTI affirms that a reliability level of .60 is good and of .80 is very good (Renaissance Learning, 2010). Using a collection of generic, split-half, and test-retest reliability data, the reliability level of STAR exceeds .80 as reported by NCRTI in Table 3.5 below (NCRTI, 2012).

Table 3.5

**STAR Enterprise: Reading – Reliability**

<table>
<thead>
<tr>
<th>Type of Reliability</th>
<th>Age or Grade</th>
<th>n (range)</th>
<th>Coefficient Range</th>
<th>Coefficient Median</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic</td>
<td>1 – 5</td>
<td>7,523 – 10,476</td>
<td>.89 – .91</td>
<td>.90</td>
<td>36 – 62 Median: 48</td>
</tr>
<tr>
<td>Split Half</td>
<td>1 – 5</td>
<td>7,523 – 10,476</td>
<td>.88 – .89</td>
<td>.89</td>
<td></td>
</tr>
<tr>
<td>Retest</td>
<td>1 – 5</td>
<td>296 – 300</td>
<td>.82 – .89</td>
<td>.83</td>
<td></td>
</tr>
<tr>
<td>Generic</td>
<td>2 – 12</td>
<td>1,153 – 6,462</td>
<td>.90 – .93</td>
<td>.92</td>
<td>71 – 83 Median: 81</td>
</tr>
<tr>
<td>Split Half</td>
<td>6 – 12</td>
<td>1,153 – 6,462</td>
<td>.89 – .91</td>
<td>.90</td>
<td></td>
</tr>
<tr>
<td>Retest</td>
<td>6 – 12</td>
<td>209 – 295</td>
<td>.80 – .90</td>
<td>.81</td>
<td></td>
</tr>
</tbody>
</table>

The STAR Reading Technical Manual states that “traditional internal consistency
reliability coefficients such as Cronbach’s alpha and Kuder-Richardson Formula 20 (KR-20) cannot be calculated for adaptive tests” (Renaissance Learning, 2012, p. 45). Therefore, Renaissance Learning calculates internal consistency using the split-half method. The split-half reliability estimates from 2008 are displayed in Table 3.6, showing similar results displayed in Table 3.5 above. “These reliability estimates are quite consistent across grades 1 – 12, and quite high . . . a result of the measurement efficiency inherent in the adaptive nature of the STAR Reading test” (Renaissance Learning, 2012, p. 45).
<table>
<thead>
<tr>
<th>Grade</th>
<th>N</th>
<th>ρxx</th>
<th>ρxx</th>
<th>N</th>
<th>ρxx</th>
<th>Average Days between Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7,523</td>
<td>0.91</td>
<td>0.88</td>
<td>298</td>
<td>0.89</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>10,132</td>
<td>0.90</td>
<td>0.89</td>
<td>296</td>
<td>0.85</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>10,476</td>
<td>0.89</td>
<td>0.89</td>
<td>297</td>
<td>0.82</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>9,984</td>
<td>0.89</td>
<td>0.89</td>
<td>297</td>
<td>0.83</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>8,352</td>
<td>0.90</td>
<td>0.89</td>
<td>300</td>
<td>0.83</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>6,462</td>
<td>0.90</td>
<td>0.89</td>
<td>294</td>
<td>0.81</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>4,767</td>
<td>0.91</td>
<td>0.90</td>
<td>288</td>
<td>0.83</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>4,364</td>
<td>0.91</td>
<td>0.90</td>
<td>284</td>
<td>0.80</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>2,921</td>
<td>0.92</td>
<td>0.90</td>
<td>241</td>
<td>0.86</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>2,079</td>
<td>0.93</td>
<td>0.90</td>
<td>214</td>
<td>0.80</td>
<td>7</td>
</tr>
<tr>
<td>11</td>
<td>1,795</td>
<td>0.93</td>
<td>0.90</td>
<td>209</td>
<td>0.80</td>
<td>5</td>
</tr>
<tr>
<td>12</td>
<td>1,153</td>
<td>0.93</td>
<td>0.91</td>
<td>245</td>
<td>0.90</td>
<td>8</td>
</tr>
<tr>
<td>Overall</td>
<td>69,738</td>
<td>0.95</td>
<td>0.92</td>
<td>3,263</td>
<td>0.91</td>
<td>7</td>
</tr>
</tbody>
</table>

Validity, testing what is meant to be tested, is another requirement for a good assessment. STAR claims high content validity by using authentic children’s literature or nonfiction texts to develop relevant test items. Validity for this measurement is also established in a comparison of
students’ scores on STAR to their scores on other assessments, “including the California Achievement Test, DIBELS, FCAT, Iowa Test of Basic Skills, and Stanford Achievement Test” (Renaissance Learning, 2010, p. 9). The high correlation between these multiple assessments exceed the guidelines provided by NCRTI and are shown in Table 3.7 (NCRTI, 2012).

Table 3.7

*STAR Enterprise: Reading — Validity*

<table>
<thead>
<tr>
<th>Type of Validity</th>
<th>Age or Grade</th>
<th>Test or Criterion</th>
<th>n (range)</th>
<th>Coefficient Range</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concurrent</td>
<td>1 – 12</td>
<td>Various</td>
<td>12,000+</td>
<td>.71 – .73</td>
<td>.72</td>
</tr>
<tr>
<td>Predictive</td>
<td>3 – 6</td>
<td>SAT9 and CST</td>
<td>1,000+</td>
<td>.81 – .83</td>
<td>.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.78 – .81</td>
<td>.80</td>
</tr>
<tr>
<td>Predictive</td>
<td>2 – 6</td>
<td>SAT9</td>
<td>44 – 389</td>
<td>.66 – .73</td>
<td>.68</td>
</tr>
<tr>
<td>Concurrent</td>
<td>1 – 8</td>
<td>Suffolk Reading Scale</td>
<td>2,694</td>
<td>.78 – .86</td>
<td>.82</td>
</tr>
<tr>
<td>Construct</td>
<td>3, 5, 7, 10</td>
<td>DRP</td>
<td>273 – 424</td>
<td>.76 – .86</td>
<td>.82</td>
</tr>
<tr>
<td>Concurrent</td>
<td>1 – 4</td>
<td>DIBELS ORF</td>
<td>12,220</td>
<td>.71 – .87</td>
<td>.81</td>
</tr>
<tr>
<td>Predictive</td>
<td>1 – 6</td>
<td>Various</td>
<td>74,877 – 200,929</td>
<td>.68 – .82</td>
<td>.79</td>
</tr>
<tr>
<td>Predictive</td>
<td>7 – 12</td>
<td>Various</td>
<td>3,107 – 64,978</td>
<td>.81 – .86</td>
<td>.82</td>
</tr>
<tr>
<td>Concurrent</td>
<td>3 – 8</td>
<td>Various</td>
<td>1,200 – 2,329</td>
<td>.71 – .74</td>
<td>.73</td>
</tr>
<tr>
<td>Predictive</td>
<td>3 – 8</td>
<td>Various</td>
<td>2,974 – 4,493</td>
<td>.66 – .70</td>
<td>.68</td>
</tr>
</tbody>
</table>

The condition (receiving DRI, not receiving DRI) served as the independent variable.

DRI, published in 2009, is new to the RTI literature. At the time of this investigation, there was one study available that showed preliminary results regarding the validity of using DRI.

According to Michael C. McKenna, coauthor of DRI, (personal communication, July 12, 2013) a
study was conducted contrasting DRI with guided reading at grade one in which the results were “very encouraging.” Another study was in progress in Staunton City, Virginia, which showed “good, though still preliminary results.” A third argument validating the use of this curriculum was the authors’ clarification that DRI was built around instructional approaches that have already been validated individually, such as semantic feature analysis. A fourth argument for the validation of this curriculum is found in the credibility of the coauthors, who are both considered literacy experts. Sharon Walpole, a professor in the School of Education at the University of Delaware, “has extensive school-based experience designing and implementing tiered instructional programs” (http://www.education.udel.edu/sharonwalpole/faculty). Michael C. McKenna, Thomas G. Jewell Professor of Reading at the University of Virginia, “has focused on areas ranging from children’s attitudes toward reading to differentiated reading instruction to educational technology” (http://curry.virginia.edu/articles/michael-c.-mckenna). Walpole and McKenna have coauthored four books on literacy topics. In addition, the DRI curriculum was deemed appropriate by the Georgia Department of Education, which promoted this intervention throughout the state and offered professional development for teachers on how to implement DRI through their Regional Educational Service Agencies.

**Procedures**

Students in reading intervention were taught using either DRI by Walpole and McKenna or non-Walpole and McKenna curriculum during the 2012-2013 school year. Students who scored below the 40th percentile on the beginning-of-year STAR were placed in a reading intervention group. Students were given a phonics inventory to determine reading weaknesses. Teachers, academic coaches, and/or RTI coordinators formed homogeneous reading groups based on need. Students were not able to choose whether they were in a group or not.
Assignment to the treatment or control group was determined by the school location and administration. Placement in a group was fluid. Some students were dismissed from a reading intervention group at different times during the year due to increased achievement reflected by progress monitoring data. Some students were added to a reading intervention group mid-year, and some students remained in a reading intervention group for the entire year. Participants in the study included all students in the treatment or control groups who were provided reading interventions during the 2012-2013 school year.

After gaining approval from IRB (see Appendix A) to conduct the research, the Renaissance Learning software provided STAR data. The sample population was categorized into either the treatment group or the control group. Students in the treatment group were coded as a 1 because they received DRI in addition to traditional reading instruction, and students in the control group were coded as a 0 because they did not receive DRI in addition to traditional reading instruction. Once this coding was completed, all student names were removed and no student identification was available on the data reports. The beginning-of-year and the end-of-year scaled scores, gender, and grade level were extracted from the data reports for all participants coded as a 1 or a 0 and used to answer the research questions.

Data Analysis

Both research questions were analyzed the same way using specific data for the second grade and the third grade. Data analyses were conducted in SPSS. Specifically, a one-way analysis of covariance (ANCOVA) was deemed the appropriate procedure for both research questions because randomization was not possible and because prior education differences of the participants needed to be controlled. According to Gall et al. (2007), “the effect of ANCOVA is to make . . . two groups equal with respect to one or more control variables” (p. 320). The 2012
beginning-of-year STAR served as the control variable for differences in previous achievement. The 2013 end-of-year STAR was the dependent variable. The condition (receiving DRI intervention, not receiving DRI intervention) served as the independent variable. Note that gender was not used as an additional covariate because a chi-square test of independence between gender and condition was not significant (\( \chi^2 = 66.126, p = .438 \)). Thus, it was not necessary to add gender as a control variable in the analysis to control for the selection threat to validity.

Assumptions that applied to both research questions included normality, homogeneity of variances, linearity, and homogeneity of regression slopes. The normality assumption was examined by intervention status using histograms and skew and kurtosis statistics. The homogeneity of variances assumption was tested using Levene’s Test. The linearity assumption was investigated by intervention status using scatterplots between the covariate and the dependent variable. A Pearson correlation between the treatment and control groups and the dependent variable also analyzed the linear relationship. The homogeneity of regression slopes assumption was tested statistically by using an F test that included an interaction term between the covariate and the condition (treatment or control). This test of between-subjects effects generated a significance level for the interaction between the condition and the covariate. Regression slopes were also examined to check for homogeneity. Results of the ANCOVAs and the tests of the various assumptions and their interpretations are discussed in detail in Chapter 4.

The statistical level of significance for this study was alpha (\( \alpha \)) = .05, which was used to determine if the hypotheses should be supported or rejected. The effect size and the observed power for the research questions were calculated by SPSS as part of the ANCOVAs. The partial eta squared value was compared to Cohen’s guidelines (1988, pp. 284-287) to determine if the
effect size was small (.01), medium (.06), or large (.14). The necessary sample size for power was also compared to Cohen’s guidelines, which suggest that power should be fixed at .80 to prevent too great a risk of a Type II error and to keep the number in the study manageable for the researcher. Using a table by Cohen (1992) for the necessary sample size for power of .80, with (\(\alpha\)) = .05 and a medium effect size, the sample size should be at least 64. Since a convenience sample was used, the researcher had no control over how many students would be in the sample at the onset of the study. The sample size for Research Question 1 was 62 students and the sample size for Research Question 2 was 35 students. Limitations caused by the small sample size for Research Question 2 are discussed further in Chapter 5.
CHAPTER FOUR: FINDINGS

The purpose of this study was to investigate the differences in STAR scores for second- and third-grade at-risk students who either received or did not receive reading intervention using the RTI curriculum entitled “Differentiated Reading Instruction” (DRI) by Sharon Walpole and Michael C. McKenna during the 2012-2013 school year. The study examined the differences in STAR scaled scores from the Fall 2012 beginning-of-year test to the Spring 2013 end-of-year test, while controlling for differences in ability using the Fall 2012 beginning-of-year test.

This chapter contains four sections. The first section discusses the demographic data of the participants. The second section presents the descriptive statistics of the sample population. The results of the ANCOVAs, the assumptions, and their interpretations are presented in the third section. The fourth section is a summary of the findings.

Demographic Data

The participants for this study were 97 at-risk students from either second or third grade in four elementary schools in a school district in northeast Georgia. The students were selected because they scored below the 40th percentile on the 2012 beginning-of-year STAR. All 97 students received identical Tier 1 reading instruction found in the Common Core standards. For Research Question 1, 62 second-grade students were divided into two groups. A treatment group of 40 students received additional reading instruction in Tier 2 or Tier 3 using the DRI curriculum. A control group of 22 students did not receive additional Tier 2 or Tier 3 reading instruction using the DRI curriculum. For Research Question 2, 35 third-grade students were divided into two groups. A treatment group of 13 students received additional reading instruction in Tier 2 or Tier 3 using the DRI curriculum. A control group of 22 students did not receive additional Tier 2 or Tier 3 reading instruction using the DRI curriculum. This
demographic data is provided in Table 3.2.

**Descriptive Statistics**

Before computing any statistics, it was observed that the treatment group contained a third-grade student with an outlier score as seen in Figure 2. According to Gall et al. (2007), “if the outlier’s score is not attributable to a calculation or recording error, you need to search elsewhere for an explanation” (p. 154). It was determined that this student’s participation in the study was not due to clerical error. The study included all students who scored below the 40th percentile, and this student’s pretest score was in the 37th percentile. However, further review suggested that the initial 2012 STAR score was likely an inaccurate chance occurrence. Beginning with the first progress monitoring test and all progress monitoring tests thereafter, the student’s percentile rank was in the range of 78 to 90, as seen in Table 4.1. According to RTI guidelines, this student should have been transferred back to Tier 1 since progress monitoring results indicated that this student no longer needed Tier 2 intervention.

![Figure 2. Student Outlier Score in the Treatment Group.](image-url)
Table 4.1

Progress Monitoring Results of Student with Outlier Score (Treatment Group)

<table>
<thead>
<tr>
<th>Test Date</th>
<th>Scaled Score</th>
<th>Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>August 28, 2012 (covariate)</td>
<td>299</td>
<td>37</td>
</tr>
<tr>
<td>October 10, 2012</td>
<td>523</td>
<td>86</td>
</tr>
<tr>
<td>November 2, 2012</td>
<td>495</td>
<td>80</td>
</tr>
<tr>
<td>December 5, 2012</td>
<td>573</td>
<td>88</td>
</tr>
<tr>
<td>March 8, 2013</td>
<td>540</td>
<td>78</td>
</tr>
<tr>
<td>May 1, 2013 (dependent variable)</td>
<td>678</td>
<td>90</td>
</tr>
</tbody>
</table>

According to O’Halloran (2005), if the conclusion changes when a case is deleted, and if there is reason to believe that the case belongs to a population other than the one under investigation, then the case should be omitted before proceeding with the analysis. In addition, Cook’s Distance Measure was used to determine the impact this outlier score would have on results of the study. The Cook’s Distance test revealed that the outlier score was extremely influential. All other third-grade scores had Cook’s distance values less than .30 while the outlier score had a Cook’s distance value of 1.77. Therefore, it was determined that this score should be eliminated before conducting the ANCOVA so the results would not be distorted. At this point in the study, the deletion of this participant changed the sample size to 96.

To answer Research Question 1, descriptive statistics for the STAR scaled scores for the 2012 beginning-of-year test and the 2013 end-of-year test of the second-grade sample are shown in Table 4.2. Descriptive statistics for the 2012 beginning-of-year test and the 2013 end-of-year test by intervention status of the second-grade sample are shown in Table 4.3. To answer Research Question 2, descriptive statistics for the STAR scaled scores for the 2012 beginning-of-
year test and the 2013 end-of-year test of the third-grade sample are shown in Table 4.4. Descriptive statistics for the 2012 beginning-of-year test and the 2013 end-of-year test by intervention status of the third-grade sample are shown in Table 4.5.

Table 4.2

*Descriptive Statistics of the Second Grade for Kurtosis and Skew*

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>Kurtosis</th>
<th>Skew</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013 Treatment Group</td>
<td>40</td>
<td>198.78</td>
<td>74.855</td>
<td>-1.047</td>
<td>0.063</td>
</tr>
<tr>
<td>2013 Control Group</td>
<td>22</td>
<td>259.95</td>
<td>76.878</td>
<td>0.236</td>
<td>-0.915</td>
</tr>
<tr>
<td>2013 Combined</td>
<td>62</td>
<td>220.48</td>
<td>80.548</td>
<td>-1.065</td>
<td>-0.179</td>
</tr>
<tr>
<td>2012 Treatment Group</td>
<td>40</td>
<td>98.85</td>
<td>30.555</td>
<td>0.128</td>
<td>1.182</td>
</tr>
<tr>
<td>2012 Control Group</td>
<td>22</td>
<td>97.36</td>
<td>23.615</td>
<td>2.686</td>
<td>1.405</td>
</tr>
<tr>
<td>2012 Combined</td>
<td>62</td>
<td>98.32</td>
<td>28.096</td>
<td>0.619</td>
<td>1.241</td>
</tr>
<tr>
<td>Combined Scaled Score Mean Growth</td>
<td>122.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3

*Descriptive Statistics of the Second Grade for Mean Growth*

<table>
<thead>
<tr>
<th>Intervention Status</th>
<th>n</th>
<th>2012 M</th>
<th>2012 SD</th>
<th>2013 M</th>
<th>2013 SD</th>
<th>Mean Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>40</td>
<td>98.85</td>
<td>30.555</td>
<td>198.78</td>
<td>74.855</td>
<td>99.93</td>
</tr>
<tr>
<td>ANCOVA Adjusted</td>
<td>40</td>
<td>98.85</td>
<td>30.555</td>
<td>198.35</td>
<td>11.488</td>
<td>99.50</td>
</tr>
<tr>
<td>Control</td>
<td>22</td>
<td>97.36</td>
<td>23.615</td>
<td>259.95</td>
<td>76.878</td>
<td>162.59</td>
</tr>
<tr>
<td>ANCOVA Adjusted</td>
<td>22</td>
<td>97.36</td>
<td>23.615</td>
<td>260.73</td>
<td>15.491</td>
<td>163.37</td>
</tr>
</tbody>
</table>
Table 4.4

*Descriptive Statistics of the Third Grade for Kurtosis and Skew*

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>Kurtosis</th>
<th>Skew</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013 Treatment Group</td>
<td>12</td>
<td>285.17</td>
<td>68.669</td>
<td>-1.120</td>
<td>-0.032</td>
</tr>
<tr>
<td>2013 Control Group</td>
<td>22</td>
<td>352.91</td>
<td>78.134</td>
<td>0.379</td>
<td>-0.773</td>
</tr>
<tr>
<td>2013 Combined</td>
<td>34</td>
<td>329.00</td>
<td>80.849</td>
<td>-0.645</td>
<td>-0.333</td>
</tr>
<tr>
<td>2012 Treatment Group</td>
<td>12</td>
<td>142.50</td>
<td>46.915</td>
<td>-0.801</td>
<td>0.726</td>
</tr>
<tr>
<td>2012 Control Group</td>
<td>22</td>
<td>192.41</td>
<td>60.265</td>
<td>-1.047</td>
<td>-0.344</td>
</tr>
<tr>
<td>2012 Combined</td>
<td>34</td>
<td>174.79</td>
<td>60.257</td>
<td>-1.296</td>
<td>0.096</td>
</tr>
<tr>
<td>Combined Scaled Score Mean Growth</td>
<td></td>
<td>154.21</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.5

*Descriptive Statistics of the Third Grade for Mean Growth*

<table>
<thead>
<tr>
<th>Intervention Status</th>
<th>n</th>
<th>2012 M</th>
<th>2012 SD</th>
<th>2013 M</th>
<th>2013 SD</th>
<th>Mean Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>12</td>
<td>142.50</td>
<td>46.915</td>
<td>285.17</td>
<td>68.669</td>
<td>142.67</td>
</tr>
<tr>
<td>ANCOVA Adjusted</td>
<td>12</td>
<td>142.50</td>
<td>46.915</td>
<td>300.99</td>
<td>21.711</td>
<td>158.49</td>
</tr>
<tr>
<td>Control</td>
<td>22</td>
<td>192.41</td>
<td>60.265</td>
<td>352.91</td>
<td>78.134</td>
<td>160.50</td>
</tr>
<tr>
<td>ANCOVA Adjusted</td>
<td>22</td>
<td>192.41</td>
<td>60.265</td>
<td>344.27</td>
<td>15.626</td>
<td>151.86</td>
</tr>
</tbody>
</table>

As shown by this data, the overall growth for the second-grade treatment group, before the ANCOVA adjusted for the covariate, was a mean increase in scaled scores of 99.93. The overall growth for the second-grade control group, before the ANCOVA adjusted for the covariate, was a mean increase in scaled scores of 162.59. The overall growth for the third-grade
treatment group, before the ANCOVA adjusted for the covariate, was a mean increase in scaled scores of 142.67. The overall growth for the third-grade control group, before the ANCOVA adjusted for the covariate, was a mean increase in scaled scores of 160.50.

**Results**

\( H_{o1} \): While using 2012 STAR beginning-of-year scaled scores as a control variable for differences in achievement, at-risk second-grade students who receive Differentiated Reading Instruction (DRI) through RTI will not have significantly different scaled scores on the 2013 STAR end-of-year test when compared to at-risk students who do not receive this intervention.

To test Null Hypothesis 1, a one-way ANCOVA was performed to determine if there were significantly different scaled scores on the 2013 STAR end-of-year test for second-grade students who received DRI compared to second-grade students who did not receive DRI. The intervention status (receiving DRI, not receiving DRI) was the independent variable. The 2013 STAR end-of-year scaled scores was the dependent variable. Each participant’s 2012 beginning-of-year scaled score was the covariate. Prior to the ANCOVA, assumption testing determined normality, homogeneity of variances, linearity, and homogeneity of regression slopes.

The reliability of the covariate was affirmed by data provided by NCRTI (NCRTI, 2012) as seen in Table 3.5. Since Cronbach’s alpha cannot be calculated for computer adaptive tests, internal consistency was affirmed by using the split-half method (Renaissance Learning, 2010) as seen in Table 3.6. Based on reliability data that exceeds .80 and high reliability estimates that were consistent across grades 1 – 12, it was assumed that STAR was reliable.

Normality for the second-grade sample was analyzed using descriptive statistics for the STAR scaled score data listed in Table 4.2 before adjusting for the ANCOVA. Four histograms were created to determine if normality could be assumed, as seen in Figure 3. The SPSS
histogram for the treatment group 2012 STAR pretest computed skewness as 1.182 and kurtosis as 0.128. The SPSS histogram for the treatment group 2013 STAR posttest computed skewness as 0.063 and kurtosis as -1.047. The SPSS histogram for the control group 2012 STAR pretest computed skewness as 1.405 and kurtosis as 2.686. The SPSS histogram for the control group 2013 STAR posttest computed skewness as -0.915 and kurtosis as 0.236.

Several arguments regarding what constitutes normality and nonnormality are found in the research literature as follows:

The multivariate normal distribution is … characterized by skewness equal to 0 and kurtosis equal to 3. However, it is common practice to subtract the constant value of 3 from the kurtosis estimate so that the normal distribution is characterized by zero skewness and zero kurtosis. (Curran, West, and Finch, 1996, p.17)

However, Micceri (1989) suggested that much of behavioral research data will never be entirely normally distributed, and Pearson (1895) questioned the occurrence of normality among real-world distributions. Nunnally (1978) doubted normality in education studies because:

Strictly speaking, test scores are seldom normally distributed. The items of a test must correlate positively with one another for the measurement method to make sense. Average correlations as high as .40 would tend to produce a distribution that was markedly flatter than the normal. (p. 160)

Curran, West, and Finch (1996) noted that nonnormality values are not precisely set:

On the basis of previous results, we [recommend] for the practicing researcher [that] we have not identified at what point the data appreciably deviate from multivariate normality. Similar to previous researchers (e.g., Muthen and Kaplan,
1985, 1992), we found significant problems arising with univariate skewness of 2.0 and kurtoses of 7.0. Further research is needed to better understand more precisely when nonnormality becomes problematic, but it seems clear that obtained univariate values approaching at least 2.0 and 7.0 for skewness and kurtoses are suspect. (p. 26)

Garson (2012) suggested identical skewness range values – “skew should be within the +2 to -2 range when the data are normally distributed” (p. 18), but recommended more conservative kurtoses range values – “kurtosis should be within the +2 to -2 range when the data are normally distributed (a few authors use the more lenient +3 to -3, while other authors use +1 to -1 as a more stringent criterion when normality is critical)” (p. 19).

All skewness values were in the range of +2 and -2, and all kurtoses values were in the range of +3 and -3 (see Table 4.2), suggesting that normality for the second-grade treatment group and second-grade control group could be assumed.
Figure 3. Histograms of Second-Grade 2012 and 2013 STAR Scaled Scores by Intervention Status.

The assumption that the sample was taken from populations with homogeneous variances was examined using Levene’s Test. The test statistic, \( F(1,60) = .02, p = .899 \), was not significant, indicating that the homogeneity of variance assumption was not violated.

The assumption of the linear relationship between the dependent variable and the covariate was tested by generating scatterplots using SPSS. A scatterplot was created for the treatment group and the control group as seen in Figure 4. The dependent variable (2013 STAR scores) was on the Y axis, and the covariate (2012 STAR scores) was on the X axis. A straight-line pattern was apparent for the treatment group, indicating that the assumption of linearity was satisfied. In addition, the Pearson correlation between the treatment pretest and posttest was \( r = .35, p = .029 \) with the correlation being significant at the 0.05 level (two-tailed), suggesting support for a linear relationship. Due to the small sample size \( (n=22) \), a curvilinear pattern was apparent for the control group, indicating a violation of linearity. The Pearson correlation between the pretest and posttest was \( r = .21, p = .350 \), also suggesting nonlinearity for the control group. Pallant (2007) documented that “scatterplots can be used to test for linearity, but these need to be checked separately for each of your groups. If you discover any curvilinear
relationships, these may be corrected by transforming your variable or alternatively [dropping] the offending covariate from the analysis. Disposing of covariates that misbehave is often easier, given the difficulty in interpreting transformed variables” (p. 293). The Northwestern University Statistics Department (1997) reported “the impact of an assumption violation on the linear regression result depends on the extent of the violation such as … how skewed the Y population distribution is. Some small violations may have little practical effect on the analysis” (retrieved from http://www.basic.northwestern.edu/statguidefiles/linreg_ass_vio.html). Due to the low skew value of the second-grade control group’s 2013 STAR data of -0.915 (see Table 4.2), it was determined that transforming or disposing of the offending variable was unnecessary and the linearity violation could be tolerated. However, caution should be used in interpreting the results of the control group due to the possibility of a Type error caused by a small sample size (n=22).

![Figure 4](image)

*Figure 4.* Scatterplots of Second-Grade 2012 and 2013 STAR Scaled Scores by Intervention Status.

The assumption for the homogeneity of regression slopes, that there is no interaction between the covariate and the dependent variable, was evaluated by creating two scatterplots as seen in Figure 5. The two lines corresponding to the pretest and posttest were similar in their
slopes, which satisfied the assumption that for a traditional ANCOVA the regression slopes are equal across all groups. Homogeneity of regression slopes was also evaluated by a test of between-subjects effects, which generated a significance level for the interaction between the treatment and the covariate. This test revealed that the interaction was not statistically significant, $F(1,58) = .05, p = .833$, and suggested that the assumption was not violated.

![Figure 5](image)

**Figure 5.** Homogeneity of Regression Slopes of Second-Grade Sample.

After determining that the linearity assumption could be tolerated and that all other assumptions were not violated, an ANCOVA was performed to test Null Hypothesis 1. Results from the ANCOVA, after adjusting for the covariate, indicated that the adjusted mean for the second-grade treatment group was 198.35 ($SD = 11.488$) and the adjusted mean for the second-grade control group was 260.73 ($SD = 15.491$). The overall growth for the second-grade treatment group, after adjusting for the covariate, was a mean increase in scaled scores of 99.5 (adjusted ending mean of 198.35 less pretest mean of 98.85). The overall growth for the second-grade control group, after adjusting for the covariate, was a mean increase in scaled scores of 163.37 (adjusted ending mean of 260.73 less pretest mean of 97.36). The adjusted means did indicate a statistically significant difference between groups, $F(1,59) = 10.46, p = .002$, partial $\eta^2 = .15$, indicating that the treatment group did have a significantly different mean score,
although it was significantly lower. According to Cohen (1988), the effect size of .15 is large, indicating that 15% of the variance in the 2013 STAR scores can be accounted for by DRI instruction. The covariate accounted for 9% of the variance in the dependent variable. SPSS indicated an observed power of .89, which suggested adequate statistical power and the probability of a Type error to be low. Based on these results, there was a statistically significant difference on the dependent variable due to the independent variable after adjusting for the covariate. Thus, Null Hypothesis 1 is rejected.

$H_{02}$: While using 2012 STAR beginning-of-year scaled scores as a control variable for differences in achievement, at-risk third-grade students who receive Differentiated Reading Instruction (DRI) through RTI will not have significantly different scaled scores on the 2013 STAR end-of-year test when compared to at-risk students who do not receive this intervention.

To support or fail to support Null Hypothesis 2, a one-way ANCOVA was performed to determine if there were significantly different scaled scores on the 2013 STAR end-of-year test for third-grade students who received DRI compared to third-grade students who did not receive DRI. The intervention status (receiving DRI, not receiving DRI) was the independent variable. The 2013 STAR end-of-year scaled scores was the dependent variable. Each participant’s 2012 beginning-of-year scaled score was the covariate. Prior to the ANCOVA, assumption testing was conducted to determine normality, homogeneity of variances, linearity, and homogeneity of regression slopes.

Since the covariate did not change, the reliability of the covariate for Research Question 2 was identical to Research Question 1 (see affirmation on p. 76). STAR was assumed reliable.

Normality for the third-grade population was analyzed using descriptive statistics for the STAR scaled score data listed in Table 4.4 before adjusting for the ANCOVA. Four histograms
were created to determine if normality could be assumed as seen in Figure 6. The SPSS histogram for the treatment group 2012 STAR pretest computed skewness as 0.726 and kurtosis as -0.801. The SPSS histogram for the treatment group 2013 STAR posttest computed skewness as -0.032 and kurtosis as -1.120. The SPSS histogram for the control group 2012 STAR pretest computed skewness as -0.344 and kurtosis as -1.047. The SPSS histogram for the control group 2013 STAR posttest computed skewness as -0.773 and kurtosis as 0.379.

The normality and nonnormality arguments discussed for Research Question 1 applied to Research Question 2 (see page 77). Since all skewness and kurtoses values were in the range of +2 and -2 (see Table 4.4), normality was assumed for both third-grade groups.

*Figure 6.* Histograms of Third-Grade 2012 and 2013 STAR Scaled Scores by Intervention
Status.

The assumption that the sample was taken from populations with homogeneous variances was examined using Levene’s Test. The test statistic, $F(1,31) = .312, p = .580$, was not significant, indicating that the homogeneity of variance assumption was not violated.

The assumption of the linear relationship between the dependent variable and the covariate was tested by generating scatterplots using SPSS. A scatterplot was created for the treatment group and the control group as seen in Figure 7. The dependent variable (2013 STAR scores) was on the Y axis, and the covariate (2012 STAR scores) was on the X axis. Due to the small sample size ($n=12$), a curvilinear pattern was apparent for the treatment group, indicating a violation of linearity. In addition, the Pearson correlation between the treatment pretest and posttest was $r = .44, p = .156$ with the correlation being significant at the 0.05 level (two-tailed), suggesting a nonlinear relationship. Due to the small sample size ($n=22$), a curvilinear pattern was apparent for the control group, indicating a violation of linearity. The Pearson correlation between the pretest and posttest was $r = .34, p = .120$, also suggesting nonlinearity for the control group. The linearity and nonlinearity arguments discussed for Research Question 1 applied to Research Question 2 (see page 79). Due to the low skew value of the third-grade treatment group’s 2013 STAR data of -0.032 and the low skew value of the third-grade control group’s 2013 STAR data of -0.773 (see Table 4.4), it was determined that transforming or disposing of the offending variables was unnecessary and the linearity violations could be tolerated. However, caution should be used in interpreting the third-grade results due to the possibility of a Type error caused by small sample sizes for both the treatment group ($n=12$) and the control group ($n=22$).
The assumption for the homogeneity of regression slopes, that there is no interaction between the covariate and the treatment, was evaluated by creating two scatterplots as seen in Figure 8. The two lines corresponding to the pretest and posttest were similar in their slopes, which satisfied the assumption that for a traditional ANCOVA the regression slopes are equal across all groups. Homogeneity of regression slopes was also evaluated by a test of between-subjects effects, which generated a significance level for the interaction between the treatment and the covariate. This test revealed that the interaction was not statistically significant, $F(1,29) = .136, p = .715$, and suggested that the assumption was not violated.

*Figure 7. Scatterplots of Third-Grade 2012 and 2013 STAR Scaled Scores by Intervention Status.*
Figure 8. Homogeneity of Regression Slopes of Third-Grade Sample.

After determining that the linearity violations could be tolerated and all other assumptions were not violated, an ANCOVA was performed to test Null Hypothesis 2. Results from the ANCOVA, after adjusting for the covariate, indicated that the adjusted mean for the treatment group was 300.99 ($SD = 21.711$) and the adjusted mean for the control group was 344.27 ($SD = 15.626$). The overall growth for the third-grade treatment group, after the ANCOVA adjusted for the covariate, was an increase in mean scaled scores of 158.49 (adjusted ending mean of 300.99 less pretest mean of 142.50). The overall growth for the third-grade control group, after the ANCOVA adjusted for the covariate, was an increase in mean scaled scores of 151.86 (adjusted ending mean of 344.27 less pretest mean of 192.41). The adjusted means did not indicate a statistically significant difference between groups, $F (1,31) = 2.424, p = .130$, partial $\eta^2 = .07$, indicating that the group receiving DRI did not have a significantly different mean score. According to Cohen (1988), the effect size of .07 is medium, indicating that 7% of the variance in the 2013 STAR scores can be accounted for by DRI instruction. The covariate accounted for 13% of the variance in the dependent variable. SPSS indicated an observed power of .33, which suggested the possibility of a Type II error due to the small sample size ($n=34$). Based on these results, there was no statistically significant difference on the
dependent variable due to the independent variable after adjusting for the covariate. Thus, it is necessary to support Null Hypothesis 2. However, it is noted that caution should be observed due to the likelihood of a Type II error.

**Additional Analysis**

The focus of this study was to investigate the reading achievement of students who were taught with the treatment curriculum through RTI Tier 2 and Tier 3. As noted in the analyses above, students in the control group had higher growth results on the end-of-year STAR than students in the treatment group, although both groups of students showed gains. A paired sample t test was conducted for the treatment group (n=52) to determine if the mean growth after adjusting for the ANCOVA was significant. Results from this analysis showed a significant increase from pretest to posttest for the treatment group, t(51) = -11.19, p = .00, indicating that students taught with the treatment curriculum made significant gains in reading achievement, although they were lower than those of students taught with the control group curricula.

**Results Summary**

The purpose of this study was to investigate the impact of a new RTI curriculum, Differentiated Reading Instruction (DRI), on at-risk second- and third-grade students. The 2013 end-of-year STAR scaled scores for Tier 2 and Tier 3 students receiving and not receiving DRI instruction were examined to determine any statistically significant differences between the two groups. Results indicated that, while using the 2012 STAR beginning-of-year scaled scores as a control variable, second-grade at-risk students who received DRI instruction had significantly lower scaled scores on the 2013 end-of-year STAR compared to second-grade at-risk students who did not receive DRI instruction. Results also indicated that while using the 2012 STAR beginning-of-year scaled scores as a control variable, third-grade at-risk students who received
DRI instruction had no significantly different scaled scores on the 2013 end-of-year STAR compared to third-grade at-risk students who did not receive DRI instruction. Possible explanations of these results are discussed further in Chapter 5.
CHAPTER FIVE: DISCUSSION

Chapter 5 restates the research problem of this causal comparative study and discusses the findings from the investigation. The sections of this chapter include: the research problem, summary of the findings, discussion of the results, implications, limitations, and recommendations for further research.

Research Problem

RTI is a framework mandated for use in Georgia public schools (Georgia Department of Education, 2008) for providing services for at-risk students. Based on a model of prevention as opposed to a model of failure, RTI, in the context of reading, is defined as a “comprehensive, systematic approach to teaching and learning designed to address language and literacy problems for all students through increasingly differentiated and intensified language and literacy assessment and instruction” (Reading Today, 2009, p. 1). With the implementation of the model nationwide over the past decade, a potential benefit of using RTI has been the delivery of high quality instruction for all students, while giving students who struggle to learn the opportunity to receive intervention before experiencing academic failure (Fletcher & Vaughn, 2009).

Since the reauthorization of IDEA in 2004, which allowed RTI as an alternative for diagnosing learning disabilities, there has been much research and discussion regarding how to make the framework reliable and uniform across the country. The specific language of IDEA included establishing research-based RTI models with consistent and evidence-based interventions (Stecker, Fuchs, & Fuchs, 2008) but also allowed flexibility across states and local school districts in implementing the framework. With this flexibility, procedures for obtaining reliable and valid knowledge relevant to observing students, data analysis, evaluations, etc. were expected (Zucker, 2004).
Although the language for RTI is found in federal and state policies, the local level is where educators teach, intervene, diagnose, refer, and apply RTI practices. In implementing the multi-tiered approach effectively, considerable time, planning, and commitment from educators in school systems have been required (Reed, Wexler, and Vaughn, 2012). Fuchs and Fuchs (2009) agree that RTI has been a multifaceted framework to put into practice:

RTI is complex in its intent and scope. It is also complex in terms of structure (multiple levels) and [due to the fact that] various kinds of assessments (screenings and progress monitoring) must be integrated meaningfully with different forms of instruction (core, small-group, and individualized). It is challenging for another reason: It requires close coordination of services delivered by different personnel at different prevention levels (e.g., teachers at primary prevention, paraprofessionals at secondary prevention, reading specialists or special educators at tertiary prevention). (p. 251)

A challenge in implementing RTI has been identifying research-based interventions that align with the objectives of IDEA. IDEA encouraged the development of researched-based interventions, but new curricula developed for RTI must be tested, analyzed, revised, and refined to be able to empirically support the outcomes expected in the classroom. According to Fuchs and Fuchs (2006), most public schools lack the personnel and the expertise to authenticate and confirm reliability of interventions and curricula used for RTI. Therefore, schools have become dependent on textbook companies and software producers to validate the curriculum they are teaching as research-based.

As documented in Chapter 1, many respondents from the nationwide RTI Adoption Surveys of both 2009 and 2010 reported that a major obstacle for RTI in their school districts
was finding research-based interventions appropriate for specific grade levels (SpectrumK12, 2009) (SpectrumK12, 2010). Furthermore, Denton et al. (2010) suggested that administrators need evidence regarding grade-appropriate interventions they are providing for teachers to use to fulfill RTI requirements. According to Reutzel, Petscher, and Spichtig (2012), today’s environment of high accountability supports the need for “carefully constructed evaluations of commercially available supplementary invention programs” (p. 406). This study tested a new RTI curriculum written for elementary students (kindergarten through third grade) with the purpose of providing evidence concerning an RTI curriculum in a public school setting.

**Summary of the Findings**

**Research Question 1.** While using 2012 STAR beginning-of-year scaled scores as a control variable for differences in achievement, do at-risk second-grade students who receive Differentiated Reading Instruction (DRI) through RTI have significantly different scaled scores on the 2013 STAR end-of-year test when compared to at-risk students who do not receive this intervention?

Research Question 1 examined the differences in STAR pretest and posttest scaled scores for a group of second-grade at-risk students (n=62) who either received or did not receive a specific reading intervention through RTI during the 2012-2013 school year. Students in the treatment group (n=40) received DRI intervention, and students in the control group (n=22) did not receive DRI intervention. The students were selected from four elementary schools in one rural school district in northeast Georgia. The use of the causal comparative design was justified because random assignment to a group was not possible. All students scored below the 40th percentile on the 2012 STAR beginning-of-year test, making them eligible to receive either Tier 2 or Tier 3 reading interventions. The independent variable was the intervention status.
(receiving DRI intervention, not receiving DRI intervention), the dependent variable was the 2013 STAR end-of-year test, and the covariate was the 2012 STAR beginning-of-year test.

An ANCOVA was conducted to make the treatment and control groups equal with respect to differences in previous achievement. On the 2012 STAR beginning-of-year test, the treatment group had a mean score of 98.85, and the control group had a mean score of 97.36. After adjusting for the covariate, the 2013 STAR end-of-year test mean score of the treatment group was 198.35, and the 2013 STAR end-of-year mean score of the control group was 260.73. This data indicated that the overall mean score gain for the second-grade treatment group was 99.50, and the overall mean score gain for the second-grade control group was 163.37 (see Table 4.3). The 2013 STAR scaled scores of the students who received intervention using the DRI curriculum were significantly lower than the 2013 scaled scores of the students who did not receive intervention using the DRI curriculum.

**Research Question 2.** While using 2012 STAR beginning-of-year scaled scores as a control variable for differences in achievement, do at-risk third-grade students who receive Differentiated Reading Instruction (DRI) through RTI have significantly different scaled scores on the 2013 STAR end-of-year test when compared to at-risk students who do not receive this intervention?

Research Question 2 examined the differences in STAR pretest and posttest scaled scores for a group of third-grade at-risk students ($n=34$) who either received or did not receive a specific reading intervention through RTI during the 2012-2013 school year. Students in the treatment group ($n=12$) received the DRI intervention, and students in the control group ($n=22$) did not receive the DRI intervention. The selection and random assignment limitations were identical to those of Research Question 1. The independent variable, dependent variable, and covariate were
also the same.

An ANCOVA was conducted to make the treatment and control groups equal with respect to differences in previous achievement. On the 2012 STAR beginning-of-year test, the treatment group had a mean score of 142.50, and the control group had a mean score of 192.41. After adjusting for the covariate, the 2013 STAR end-of-year test mean score of the treatment group was 300.99, and the 2013 STAR end-of-year test mean score of the control group was 344.27. This data indicated that the overall mean score gain for the third-grade treatment group was 158.49, and the overall mean score gain for the third-grade control group was 151.86 (see Table 4.5). The 2013 STAR scaled scores of the students who received intervention using the DRI curriculum were not significantly different from the 2013 STAR scaled scores of the students who did not receive intervention using the DRI curriculum.

**Discussion of the Results**

Theoretically, this study aligned with Vygotsky’s Social Development Theory that students learn best through social interaction with a more skillful tutor (Miller, 2010). All RTI groups, both treatment and control, were led by qualified teachers in small groups of 6 or fewer students (see Table 3.4).

Empirically, a vast amount of research has verified or disproved instructional practices intended for teaching children to read. Allington (2012) suggested that “we now have an essential research base demonstrating that virtually every child could be reading grade level by the end of first grade” (p. 520). Research literature exists regarding core reading programs, phonics proficiency, teacher qualifications, independent reading, inventive writing, and more, all of which should influence what happens in RTI groups. The remainder of this section discusses how this study aligned or did not align with existing research on how children become
accomplished readers.

Many core reading programs and RTI curricula are marketed as evidenced-based, reliable, replicable, etc. Although curriculum does have a purpose in delivering instruction, evidence does not support the use of a commercial core reading program in its entirety. In 2006, McGill-Franzen, Zmich, Solic, and Zeig reported that in Florida, various core reading programs were used when a quarter of the third-grade population failed the state reading test, indicating that the use of core reading programs failed to develop reading growth. In 2007, What Works Clearinghouse reported that the only core reading program (out of 153 programs) with strong evidence of improving reading achievement was Reading Recovery. Dewitz, Jones, and Leahy (2009) examined five core reading programs and found that none of the curriculum fostered reading comprehension development.

Evidence does indicate that core reading programs may have value when used as instructional tools, rather than the total reading program. Dewitz and Sullivan (2010) suggested that core reading programs promote themed instruction, foundational lessons (that need to be expanded on), texts for modeling and guided practice (but not for independent reading), and structure for new teachers (but not for experienced teachers). Aligned with this research, the core reading program in this study, the national Common Core curriculum, provided the reading standards, but teachers were allowed to teach the standards with practices-of-best-fit for their students. A variety of teaching resources such as a traditional Basal series, Reading First materials from a previous federal grant, framework lessons provided by the state of Georgia, and various other resources were used to provide quality Tier 1 instruction.

This evidence regarding the use of a single program both aligns and does not align with the Tier 2 and Tier 3 instruction provided in this study. The treatment group used a new RTI
curriculum (DRI), which had been introduced to the RTI literature in 2009. The DRI curriculum was used the entire year for the treatment group while the control group used various curricula depending on the needs of the students. All students were assessed with a phonics inventory to target specific weaknesses before entering an RTI group. The treatment group curriculum provided instruction in the areas of phonemic awareness, phonics and word study, and fluency, whereas the control group curricula could provide instruction with a variety of resources in any needed area.

Although end-of-year results showed that the second-grade treatment group scored significantly lower than the second-grade control group, and the third-grade treatment group and third-grade control group comparison was not significant, both groups made academic gains (see Table 4.3 and Table 4.5). The unknown curricula used for students in the control group (n=44) was effective. RTI coordinators and administrators of the participating schools approved any curricula used in the control group. The unknown curricula were not documented due to this study’s focus on determining the effectiveness of the new DRI curriculum in improving at-risk students’ reading ability. The end-of-year scores of the students in the control group suggested that reading achievement is possible with a variety of curricula choices. This aligns with an investigation by Blair, Rupley, and Nichols (2007) suggesting that effective teachers combined methods and used a variety of texts to make the instruction more individualized for specific needs. This concept is discussed further in the implications section.

The DRI curriculum used for students in the treatment group (n=52) was also effective, with eight students showing an increase of over 200% in their raw scaled scores, 20 students showing an increase of over 100% in their raw scaled scores, and an additional 10 students showing an increase of between 50% and 99% in their raw scaled scores (see also the Additional
Analysis section at the end of Chapter 4). Renaissance Learning provides STAR Common Core State Standards functional grade-level cut scores based on an average of 70% mastery across standards (Renaissance Learning, 2013). The functional grade-level classifications are criterion-referenced and indicate whether or not students have learned the grade-level standards. According to the 2012 STAR beginning-of-year scaled scores, the second-grade treatment group had zero students who were on grade level. All scaled scores were below the cut score of 179. The 2013 STAR end-of-year scaled scores indicated that 21 students (53%) in the second-grade treatment group had improved to On Grade Level status with scaled scores of 206 or above. According to the 2012 STAR beginning-of-year scaled scores, the third-grade treatment group had one student who was on grade level with a scaled score of 232. All other scaled scores were below the cut score of 219. The 2013 STAR end-of-year scaled scores indicated that seven students (58%) in the third-grade treatment group had improved to On Grade Level status with scaled scores of 273 or above. Overall, 28 students (54%) in the treatment group improved to On Grade Level status. The end-of-year scores of the students in the treatment group suggest that the DRI curriculum was beneficial for some students in promoting reading achievement. To align with existing research, the DRI curriculum should be noted as one effective method for developing phonics proficiencies in young children because “no single approach works for every child and effective teachers adapt their teaching until they locate the best method for developing [reading] proficiencies for each child” (Allington, 2012, p. 522).

Additional support for not using a one-size-fits-all approach is found in existing phonics’ research on decoding. Decoding of text was emphasized by The National Reading Panel Report (2000), which initiated purchases of decodable texts in schools across the country, a practice that is still current. However, Jenkins, Peyton, Sanders, and Vadasay (2004) found that first-grade
students produced the same reading outcomes whether they had decodable texts or predictive texts when they had consistent decoding lessons. Teaching children to read decodable nonsense words, also a current practice in early childhood classrooms, was assumed to show a child’s true decoding ability. Pressley (2002) observed that too much emphasis on nonsense words interfered with children using cross-checking and self-regulating strategies when trying to read actual texts. Allington (2012) stated that developing early decoding skills is a necessary prerequisite for reading, but effective teachers should know several decoding methods because no single approach works for every child. Cunningham (2011) agreed that research does not indicate a single best method for teaching decoding:

The key conclusion of this research is that children do need systematic phonics instruction, but there is no one best way to teach phonics. This conclusion is disturbing to those who would like for there to be a specified best way so that everyone would be mandated to do it that way. (p. 221)

Another alignment of this study with existing research is found in teacher qualifications. There is evidence that supports the view that quality reading instruction is less about the curriculum and more about the effectiveness of the teacher. Nye, Konstantopoulos, and Hedges (2004) and Stuhlman and Pianta (2009) both reported research which indicated the expertise of the teacher was the critical factor in the quality of reading instruction. Qualifications of the teachers in this study were documented in the setting section in Chapter 3. Out of the 29 teachers in the study, 24 teachers had graduate degrees, and 27 teachers had a range of experience from four to 31 years. However, documenting teacher fidelity in this study did not align with current research and is discussed in detail in the limitations section of Chapter 5 (see p. 103).
Since this study addressed struggling readers, existing research on the value of independent reading practice and inventive writing for children should be considered. The intervention groups in this study consisted of 20-minute sessions 3 to 5 times per week, depending on the Tier and on the school schedule. The DRI curriculum for the treatment group was scripted, which is recommended by some researchers to offer better quality control and integrity of instruction (Gresham et al., 2005). Walpole and McKenna (2009), authors of the DRI curriculum, stated their philosophy early on in their book that “reading practice will happen when children are not with the teacher” (p. 8). They designed the DRI lessons heavily with modeling, repetition, and an emphasis on every child attending every minute they were in the differentiated instruction group. The details of the control group curricula are unknown since the focus of this study was on the treatment curriculum, but the design of both RTI groups was skills-focused instruction.

Numerous studies indicate that struggling readers should spend more time reading and writing daily. Stahl and Nagy (2006) found that vocabulary growth is developed mostly by engaged independent reading. Allington (2009) avowed that “there is also evidence that almost everything, from phonemic awareness, to phonics, to comprehension, is developed through independent reading and writing” (p. 526). Knapp (1995) suggested that struggling readers learned more through meaning-focused lessons than skill-focused lessons. Torgesen (2004) warned against “enormous reading practice deficits” (p. 365), and Adams (1990) reported that inventive spelling and inventive writing foster letter-sound relationships which develop phonemic awareness and promote understanding of the alphabetic principle. Adams further stated that “the most important activity for developing literacy is that of inducing students to read independently” (1990, p. 295). Evidence is also available on the importance of self-selected
texts to improve reading (Krashen, 2004) and the significance of exposure to print in the elementary years (Cunningham and Stanovich, 1997).

The RTI Tier 2 and Tier 3 groups in this study mostly did not align with existing research regarding independent reading and writing. It is possible that these students received this instruction through the Tier 1 lessons, but Allington (2012) suggested that “struggling readers spend two-thirds of every lesson engaged in the actual reading of texts” (p. 528). Since the RTI instruction for both groups was designed to focus on skills, (which is suggested in most RTI literature), there was little time for independent reading or writing during the reading intervention block. The fluency cycle of the DRI curriculum did include reading passages, but they were not from self-selected texts. Ehri, Dreyer, Flugman, and Gross (2007) noted that struggling readers in their study showed gains primarily due to reading texts at a high level of accuracy (texts matched to reading levels rather than matched to grade levels). According to Allington (2012), “struggling readers just participate in too little high-success reading activity every day. This is one reason so few struggling readers ever become achieving readers” (p. 525).

This study was conducted in actual classroom environments rather than the structured environment of a laboratory. Therefore, the end results could have been affected by other uncontrollable variables such as differences in parental involvement, other family factors between members of the treatment group and members of the control group, student motivation, peer group dynamics, the presence or absence of disruptive students in the groups, and truancy. This study aligned with research literature in some areas (core reading program, qualifications of teachers, the control group’s use of a variety of resources) and did not align with research literature in other areas (the treatment group’s use of a single curriculum, independent reading practice, inventive writing practice). Results of the study showed reading achievement success
for students in both groups and the new RTI treatment curriculum did produce evidence of helping struggling readers. Existing research about struggling readers, along with evidence from this study, may help school administrators and RTI coordinators in the future implement Tier 2 and Tier 3 instruction that will yield even higher growth on reading test scores, but more importantly give every child the highest opportunity to become a successful reader.

**Implications**

Results from this study suggest that reading intervention in Tier 2 and Tier 3 can improve academic achievement in at-risk students. Tier 2 and Tier 3 interventions were taught in addition to the Tier 1 core curriculum for all students. For the treatment group and the control group combined, the mean raw scaled score data increased 135.60 on the 2013 STAR end-of-year test. Only two students out of the entire population ($N=96$) did not show progress at the end of the year. One student from the treatment group scored 10 points less on the 2013 STAR end-of-year test, and one student from the control group scored 24 points less on the 2013 STAR end-of-year test. Other than these two, all students in Tier 2 or Tier 3 made progress during the school year in reading achievement based on the 2013 STAR end-of-year scaled scores.

This study investigated the efficacy of an RTI curriculum at the Tier 2 and Tier 3 level. Hill, King, Lemons, and Partanen (2012) suggest that before a Tier 2 intervention is evaluated, the contribution of the Tier 1 instruction to student achievement should be considered. The importance of the high-quality Tier 1 core instruction has been established by Fuchs, Fuchs, and Stecker (2010), who suggest that low-quality Tier 1 instruction may produce false positives. Students who are moved to Tier 2 could be true non-responders or could be casualties of poor instruction. “Fidelity at Tier 1 is directly related to responsiveness at Tier 2” (Hill et al, 2012, p. 117). In addition, alignment between the instruction provided in all the Tiers should be in place.
to permit students to move in and out of Tiers as needed (Fuchs, Fuchs, & Compton, 2012). A major implication for researchers of RTI curriculum should be to include the quality of Tier 1 instruction. Gersten and Dimino (2006) suggest that nonresponsive students who received low-quality Tier 1 instruction would be easier to remediate at Tier 2 than nonresponsive students who received high-quality Tier 1 instruction.

Results from this study also suggest that there may not be one particular curriculum that is the best for every student. Evidence from this study indicates that students taught with the treatment curriculum, DRI, did achieve successful reading gains during the 2012-2013 school year. Evidence from this study also indicates that students taught with a different intervention curriculum (a variety of unknown resources) made successful reading gains. This evidence suggests the importance of diagnostic evaluations for students at the onset of intervention instruction to ensure that specific reading weaknesses are being targeted in the intervention lessons. In this study, RTI coordinators were given the option to use the treatment curriculum or some other type of curriculum to address the needs of the students in their schools. An effort was made to diagnose specific reading weaknesses by administering a phonics inventory to each student in the sample population prior to beginning intervention instruction. Use of additional diagnostic tools may be needed to further assist in the detection of root causes of reading problems so that these issues can be addressed appropriately for successful outcomes.

It is suggested by Baker et al. (2010) that data should be collected and examined in an ongoing manner for both Tiers 1 and 2. O’Donnell (2008) proposes that collecting adequate amounts of data in school-wide studies can be very costly. A widely accepted practice for data collection for Tier 1 is a benchmark test three times a year (beginning, middle, and end). Benchmark tests for Tier 2 and Tier 3 students are administered more frequently. An implication
from these practices might be that if Tier 1 testing could be conducted more often, a larger proportion of students at moderate risk for reading difficulty could be detected earlier and interventions to address their reading weaknesses could be implemented sooner.

**Limitations**

The causal-comparative design was a limitation of this study because it cannot establish a cause and effect relationship. According to Gall et al. (2007), “a disadvantage of causal-comparative research designs is that inferences about causality on the basis of the collected data are necessarily tentative” (p.310). Therefore, caution must be used when making conclusions from the results because alternative interpretations of the findings are possible. Although the ANCOVA equalized the initial achievement level of students, other variables, as mentioned in the discussion of the results, could have had an impact on the study.

Another limitation of this study was the small sample size that was collected from one rural school district in Georgia. A larger sample size would lower the risk of any Type error and also increase the observed power of the study. The results may be localized to the rural area where the study was conducted and not be applicable to a more urban setting with different demographics.

A third limitation of this study was the inability to assign students randomly to the treatment group and the control group, which caused unequal sample sizes. All students in the population were chosen by school administrators of the participating schools. Howell (2010) suggests that if it is not possible for the sample sizes to be equal, “the smaller group should be as large as possible relative to the larger group” (p. 236). For Research Question 1, the treatment group consisted of 40 students, and the control group consisted of 22 students. For Research Question 2, the treatment group consisted of 12 students, and the control group consisted of 22
students. Although a covariate (2012 STAR beginning-of-year test) was utilized and the demographics across the population were similar, a selection threat due to non-equivalent groups existed.

Fidelity was another limitation. Prior to the commencement of the study, treatment instruction training was administered to RTI coordinators and academic coaches at each site by the POI team. The RTI coordinators and academic coaches relayed this training to the teachers at their individual schools. Teachers were provided with all materials and guidelines needed for the treatment curriculum, including scripted lesson plans, student worksheets, and the book *How to Plan Differentiated Reading Instruction: Resources for Grades K-3* (Walpole & McKenna, 2009). The implementation of the treatment across all classes in the study was confirmed in three face-to-face interviews with one assistant principal and two academic coaches at the conclusion of the study. The reading intervention block descriptions, personnel, amount of time devoted to teaching the curriculum, and the fluid placement within the groups were discussed. Based on the interviews, it was determined that the implementation of the treatment across each site in the study was similar. Details regarding treatment fidelity are described in Chapter 3.

Literature regarding teacher fidelity suggests that more involvement may be needed to ensure compliance with following teaching methods and procedures. Pence, Justice, and Wiggins (2008) suggest that teacher fidelity should be checked at least three different times over an academic year using a curriculum fidelity checklist. In a fidelity study by Abry, Rimm-Kaufman, Larsen, and Brewer (2013), it is recommended that teachers’ fidelity be rated at least five times throughout the year using some type of classroom observation measure. They also suggest that teachers complete a survey regarding their classroom practices at the end of each year. These recommendations for measuring teacher fidelity appear attainable when certain
conditions are in place. However, Stecker et al. (2008) acknowledge that schools in rural areas can have limited resources for providing personnel for intervention practices. The rural school district in this study had a shortage of personnel to conduct teacher fidelity observations and limited financial means to provide additional personnel for this task. Therefore, classroom observations were not formally conducted, and it was not possible to relate student outcomes to measures of fidelity although RTI coordinators tried to ensure the treatment was being delivered as designed by the authors. Thus, whether or not teacher or treatment fidelity had an impact on the academic gains of the students cannot be determined.

**Recommendations for Further Research**

This educational research study tested a new curriculum designed to increase reading achievement in primary-aged children. Evidence from this study suggests that some students who were taught with this curriculum made tremendous progress. This tentative evidence lends support for investigations of future additional studies on this curriculum. Since evidence from this type of study can have an influence on school policies regarding appropriate curriculum, future studies on this curriculum would be more beneficial if they had an experimental design. A carefully designed experiment with random assignment would suggest stronger evidence that any observed findings resulted from the treatment and not from extraneous variables (Gall et al, 2007).

Future research should also include a more rigorous fidelity measurement. Although most educators agree that documentation of fidelity is needed, the large amount of resources that may be required to implement fidelity procedures prohibits many researchers from adequately reporting fidelity. Without some documentation on the fidelity of instruction, it remains uncertain if students who are referred to a higher Tier are true nonresponders or casualties of
poor instruction (Hill et al., 2012).

Since the treatment curriculum was written for kindergarten through grade three, further research needs to be conducted involving all of those grades. This study only examined second-and third-grade students. Additional research could indicate that the treatment curriculum is more effective for a particular grade. In addition, a longitudinal study examining the effects of the treatment curriculum over a period of years might suggest that as teachers become more comfortable with teaching the material, the treatment fidelity would increase, possibly resulting in higher academic achievement for the students.

Conclusion

IDEA (2004) and the Georgia Department of Education (2008) require educators to use RTI as the framework for addressing the needs of struggling students in public schools in Georgia. Interventions used in RTI must be research-based with evidence of successful student outcomes. The RTI curriculum entitled How to Plan Differentiated Reading Instruction: Resources for Grades K-3 (Walpole & McKenna, 2009) is one intervention curriculum to consider. The evidence provided in this study suggested that reading interventions in Tier 2 and Tier 3 in an RTI model can increase STAR scaled scores. More specifically, students taught with the treatment intervention (DRI) did make impressive reading progress, although the gains were lower than students in the control groups. Additional research on this RTI curriculum is needed to confirm these results, establish possible strengths and weaknesses of the curriculum, and increase understanding of best practices in implementing the Response to Intervention education initiative.
REFERENCES


Coyne, M., McCoach, D., & Kapp, S. (2007). Vocabulary intervention for kindergarten students:
Comparing extended instruction to embedded instruction and incidental exposure.  

*Learning Disability Quarterly, 30, 74-88.*


Georgia Department of Education. (2009, June). Response to intervention: The Georgia student achievement pyramid of interventions. PowerPoint slides presented at the Title Programs Conference, Atlanta, GA.


Summit: Building a Foundation for the Future, Washington DC.


Infinite Campus. (2012). Franklin County School District, GA.


Renaissance Learning. (2010). Getting the most out of STAR reading. Wisconsin Rapids, WI.

Renaissance Learning. (2013). Functional grade levels (FGL) for STAR reading enterprise. Wisconsin Rapids, WI.


Educational Research, 105, 404-415.


Associates.


Grimes (Eds.), *Special education in transition: Functional and noncategorical programming* (pp. 285-301). Longmont, CO: Sopris West.


Torgesen, J. K. (2004). Lessons learned from research on interventions for students who have difficulty learning to read. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 355-382). Baltimore: Paul H. Brookes.


APPENDIX A: IRB CONSENT FORM

IRB Exemption 1631.071613: Investigation of Differences in STAR Reading Scores for Second- and Third-Grade Students Who Received Reading Intervention Instruction

IRB, IRB [IRB@liberty.edu]
Sent: Tuesday, July 16, 2013 8:56 AM
To: Heam, Elizabeth Shealy
Cc: IRB, IRB; McDonald, Connie L (School of Education); Garzon, Fernando (Center for Counseling and Family Studies)
Attachments: HeamExemption_07_13.pdf (203 KB); ChangeinProtocolUpdated.doc (68 KB)

Dear Elizabeth,

The Liberty University Institutional Review Board has reviewed your application in accordance with the Office for Human Research Protections (OHRP) and Food and Drug Administration (FDA) regulations and finds your study to be exempt from further IRB review. This means you may begin your research with the data safeguarding methods mentioned in your approved application, and that no further IRB oversight is required.

Your study falls under exemption category 46.101 (b)(4), which identifies specific situations in which human participants research is exempt from the policy set forth in 45 CFR 46:

(4) Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects.

Please note that this exemption only applies to your current research application, and that any changes to your protocol must be reported to the Liberty IRB for verification of continued exemption status. You may report these changes by submitting a change in protocol form or a new application to the IRB and referencing the above IRB Exemption number.

If you have any questions about this exemption, or need assistance in determining whether possible changes to your protocol would change your exemption status, please email us at irb@liberty.edu.

Please retain this letter for your records. Also, if you are conducting research as part of the requirements for a master’s thesis or doctoral dissertation, this approval letter should be included as an appendix to your completed thesis or dissertation.

Sincerely,

Fernando Garzon, Psy.D.
Professor, IRB Chair
Counseling

(434) 592-4054

LIBERTY UNIVERSITY

Liberty University | Training Champions for Christ since 1971
APPENDIX B: SCHOOL SUPERINTENDENT CONSENT FORM

School System

"Preparing Students for the Challenges of Tomorrow"

Superintendent

August 7, 2012

Institutional Research Board
Liberty University
1971 University Boulevard
Lynchburg, VA 24502

Liberty University IRB Office:

As Superintendent of [Redacted] County Schools, I have given Ms. Beth Hearn permission to conduct her research in our school system. I have spoken with Ms. Hearn and understand the scope of her research and how she will collect and present her data. All information to be gathered will be done in a confidential and appropriate manner.

Should you have any questions, please feel free to contact me.

Sincerely,

[Redacted] Superintendent

[Redacted]
APPENDIX C: GEORGIA DOE PERMISSION FOR RTI PYRAMID GRAPHIC

Response from the Georgia Department of Education

askdoe@doe.k12.ga.us

Response from AskDOE:

Good afternoon,
Thank you for contacting the Georgia Department of Education. We appreciate the opportunity to assist you.
You may want to contact Debbie, Director of Special Education.
She can be contacted [redacted].
Thank you again for contacting us. Please let me know if I can provide further assistance.

Debbie
Education Administration Specialist
AskDOE

Thank you again for contacting the Georgia Department of Education. This response has been provided by AskDOE, the Superintendent’s Helpdesk. If you have any questions or concerns regarding this response, please don’t hesitate to contact AskDOE at 1-866-656-5280 or 302-311-3527 and refer to Help Ticket number [redacted].

Email Received:

Can you tell me who I need to email at GADOE to ask permission to use a GADOE RTI graphic as a figure in my dissertation?

Thanks,
Beth Hearn

Permission to use RTI Graphic

Beth Hearn <bethylhearn@gmail.com>

Hello.
I am a doctoral student at Liberty University working on a dissertation on RTI. With your permission, I would like to insert the attached Georgia Department of Education RTI graphic into my dissertation as a figure. This graphic was retrieved online from “Response to Intervention: Georgia’s Student Achievement Pyramid of Interventions” - 2011 and will be cited appropriately.

Thank you for your consideration.

[File: RTI Graphic.docx]

Deborah

Please just note the source

From: Beth Hearn
Sent: Friday, May 31, 2013 2:18:04 PM
To: Deborah
Subject: Permission to use RTI Graphic