
Electronic Theses and Dissertations, 2004-2019

2012

The Impact Of Timed Versus Untimed Standardized Tests On Reading Scores Of Third Grade Students In Title I Schools

Ruth Elizabeth Haniff
University of Central Florida

 Part of the [Education Commons](#)

Find similar works at: <https://stars.library.ucf.edu/etd>

University of Central Florida Libraries <http://library.ucf.edu>

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2004-2019 by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation

Haniff, Ruth Elizabeth, "The Impact Of Timed Versus Untimed Standardized Tests On Reading Scores Of Third Grade Students In Title I Schools" (2012). *Electronic Theses and Dissertations, 2004-2019*. 2202. <https://stars.library.ucf.edu/etd/2202>



THE IMPACT OF TIMED VERSUS UNTIMED STANDARDIZED TESTS
ON READING SCORES OF THIRD-GRADE STUDENTS IN TITLE I SCHOOLS

by

RUTH ELIZABETH HANIFF
B.S. Southern Illinois University, 1989
M.S. University of Central Florida, 1993
Ed.S. University of Florida, 1998

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Education
in the College of Education
at the University of Central Florida
Orlando, Florida

Summer Term
2012

Major Professor: Martha Lue Stewart

© 2012 Ruth Elizabeth Haniff

ABSTRACT

The purpose of this study was to investigate the extent to which the performance of Title I third grade students in a central Florida School District differed on tests administered under timed and untimed conditions. Further examined was the literature on reasons for the achievement gap which centered around seven different themes: (a) standardized testing, (b) achievement gap data and identified factors (c) deficit theory, (d) cultural mismatch theory, (e) extended time accommodations, (f) test anxiety and stress, and (g) timed versus untimed tests.

Six Title I schools participated in this study by assigning 194 students to take the 2006 Released FCAT Reading Test under either timed or untimed conditions. Although there were no interactions between the covariates and testing conditions, those who were in the free or reduced lunch program or were in exceptional education programs had lower FCAT scores than those who were not. However, when school was included as a moderator, there was a statistically significant interaction between testing conditions and schools on FCAT scores indicating that the relationship between testing conditions and FCAT scores varied for each individual school.

A factorial ANCOVA was conducted, and it was found that the mean differences between students who took the timed and untimed 2006 FCAT Reading Test varied from school to school after accounting for the covariates. For two schools, those students who took the untimed tests scored higher than those who took the timed tests. In contrast, those students who took the untimed tests scored lower than those students who took the

timed test for one of the schools. There was no statistically significant difference for three of the schools.

A factorial MANCOVA was used to compare reading performance on the 2006 Reading FCAT between the timed and untimed groups on domain specific tests. The relationship between testing condition and FCAT scores for each domain specific test varied depending on the individual school. Therefore, it could not be concluded from these analyses that testing conditions would consistently result in increases or decreases of student performance on standardized domain specific tests.

This dissertation is dedicated to those who are and have been the wind beneath my wings.

Thank you to those who paved the way before me, so that I might have the opportunity to fulfill this lifelong dream of earning my doctoral degree.

Thank you to my mother, Gladys Thames, who is rejoicing in heaven. I will be eternally grateful that you taught me about the love of Jesus and that I can do all things through Christ which strengthens me.

Thank you to my loving sisters, Carlene, Linnie, Roberta, Priscilla, and Jay (who we have adopted as a sister) for all of the beautiful memories that we have shared throughout the years and those to come.

I am forever grateful for your love, laughter, encouragement, and prayers.

To my husband, Shamin Haniff who always made certain that our twins were fed and well taken care of on numerous occasions when I was attending class or doing research.

Your support has played a vital role in helping me to accomplish this goal.

You are loved and appreciated more than you know.

Finally, to my intelligent and beautiful twins, Shamin Gabriel Haniff and Elizabeth Ann Haniff, thank you for your precious love which has helped to sustain me as I pursued this dream. Every time I look at you, I am reminded of God's blessings.

ACKNOWLEDGMENTS

First, I want to thank my committee chair, Dr. Martha Lue Stewart for believing in me and encouraging me to keep pursuing my dream of completing my doctoral degree. You have been such a beautiful role model and mentor to me, and I will always be grateful. Thank you for your support, guidance, and wisdom throughout this journey.

Next, I want to thank my committee members Dr. Jennifer Platt, Dr. Larry Holt, Dr. Kerry Purmensky, and Dr. M. H. Clark. Thank you for your insights, expertise, and commitment necessary for the successful completion of this dissertation. In addition, special thanks to Dr. Jennifer Platt for engaging me to pursue my doctorate degree at UCF. The knowledge that I have gained from this program is invaluable.

Likewise, I want to acknowledge Dr. Clara Walters for being a wonderful mentor and supporting me as an educational leader. Equally, I want to thank Dr. Mary Ann Lynn for providing wonderful guidance and support as my editor. I also want to express gratitude to Leah Mitchell, my Graduate Studies Advisor, for always being willing to go above and beyond to help me throughout this quest.

Lastly, I want to express my sincere gratitude to the principals who were able to allow their staff and students to participate in this study. I am also grateful to everyone who has been instrumental in helping me to accomplish this goal.

TABLE OF CONTENTS

LIST OF TABLES	x
LIST OF ACRONYMS/ABBREVIATIONS	xi
CHAPTER 1 THE PROBLEM AND ITS CLARIFYING COMPONENTS.....	1
Introduction.....	1
Purpose of the Study	2
Research Questions	2
Assumptions.....	3
Limitations and Delimitations of the Study	3
Background of the Study	4
Definition of Terms.....	8
Conceptual Framework.....	15
Summary.....	18
CHAPTER 2 LITERATURE REVIEW	19
Introduction.....	19
The History of Standardized Testing	19
Early Initiatives in Testing.....	19
20 th Century Standardized Testing Accomplishments.....	21
21 st Century Education Reform and High-Stakes Testing.....	25
History of Florida Comprehensive Assessment Test (FCAT).....	28
No Child Left Behind Act of 2001 and Adequate Yearly Progress (AYP).....	32
Title I Programs	36
The Achievement Gap	39
2011 National Reading Report Card.....	39
2011 State FCAT Reading Results	41
2011 Orange County District FCAT Reading Results.....	42
2011-2012 Orange County Benchmark Reading Results	43
Identified Achievement Gap Factors	45
Theories Impacting Student Performance.....	49
Deficit Theory.....	50
Cultural Mismatch Theory.....	51
Theoretical Controversy.....	52
Extended Time Accommodations and Student Performance	53
Federal Laws and Accommodations.....	54
Timed Versus Untimed Timed Studies.....	55
Anxiety Concerns of High Stakes Testing.....	59
Anxiety Studies.....	60
Summary.....	62

CHAPTER 3 METHODOLOGY	65
Introduction.....	65
Research Questions.....	66
Population and Sample	66
Instrumentation	69
Data Collection Procedures.....	70
Data Analysis Process.....	71
Ethical Considerations	71
Summary	72
CHAPTER 4 RESULTS	73
Introduction.....	73
Demographic Characteristics for Participating Schools	73
Data Analysis for Research Question 1	74
Data Analysis for Research Question 2	76
Data Analysis for Research Question 3	77
Summary	82
CHAPTER 5 SUMMARY, DISCUSSION AND RECOMMENDATIONS.....	83
Introduction.....	83
Purpose of the Study	83
Discussion of Findings.....	84
Research Question 1	84
Research Question 2	90
Research Question 3	91
Conclusion	96
Limitations	98
Implications for Policy and Practice	99
Recommendations for Future Research	103
Researcher’s Reflections.....	105
Summary	111
APPENDIX A APPROVAL OF EXEMPT HUMAN RESEARCH.....	112
APPENDIX B RECRUITMENT LETTER.....	114
APPENDIX C TEST ADMINISTRATOR SCRIPT FOR TIMED TESTS.....	117
APPENDIX D TEST ADMINISTRATOR SCRIPT FOR UNTIMED TESTS.....	119
APPENDIX E PARENT PERMISSION	121
APPENDIX F STUDENT ASSENT SCRIPT.....	126
APPENDIX G STUDENT INFORMATION DATA SHEET	128

APPENDIX H FLORIDA DEPARTMENT OF EDUCATION COPYRIGHT
PERMISSION..... 130

LIST OF REFERENCES 132

LIST OF TABLES

Table 1	Adequate Yearly Progress (AYP): Subgroup-Based Components.....	33
Table 2	Adequate Yearly Progress Benchmarks in Florida.....	35
Table 3	Annual Income Eligibility Guides (Effective July 1, 2010 to June 30, 2011)...	37
Table 4	Median Combined SAT Scores by Family Income (400-1600 Scale)	38
Table 5	Demographic Characteristics of Participating Schools	74
Table 6	Interactions Between Covariates and Test Conditions	75
Table 7	Main Effects for Covariates.....	75
Table 8	Means, Standard Deviations, and <i>p</i> Values: Timed and Untimed FCAT Reading Tests.....	77
Table 9	Interactions Between Testing Conditions and School for Each Domain Test...	78
Table 10	Means and Standard Deviations: <i>F</i> Ratios Between Timed and Untimed Conditions for FCAT Reading Section on Words and Phrases	79
Table 11	Means and Standard Deviations: <i>F</i> Ratios Between Timed and Untimed Conditions for FCAT Reading Section on Main Idea	79
Table 12	Means and Standard Deviations: <i>F</i> Ratios Between Timed and Untimed Conditions for the FCAT Reading Section on Plot and Purpose.....	80
Table 13	Means and Standard Deviations: <i>F</i> Ratios Between Timed and Untimed Conditions on FCAT Reading Section for Comparison and Cause and Effect	81
Table 14	Means and Standard Deviations: <i>F</i> Ratios Between Timed and Untimed Conditions for FCAT Reading Section on Reference and Research	81

LIST OF ACRONYMS/ABBREVIATIONS

AYP	Adequate Yearly Progress
CCS	Common Core Standards
ELL	English language learners
ESE	Exceptional Student Education or Learners
FCAT	Florida Comprehensive Assessment Test
FCAT 2.0	Florida Comprehensive Assessment Test 2.0
NCLB	No Child Left Behind Act of 2001
NGSS	Next Generation Sunshine State Standards
OCPS	Orange County Public Schools
SES	Socioeconomic Status
SSS	Sunshine State Standards
VAM	Value Added Model

CHAPTER 1

THE PROBLEM AND ITS CLARIFYING COMPONENTS

Introduction

Achievement gaps occur when one group of students outperforms another group and the difference in averages scores for the two groups is statistically significant (National Center for Education Statistics, 2011a). The National Assessment of Educational Progress (NAEP) reported on the Hispanic-White achievement gap and the Black-White achievement gap using the NAEP scores in mathematics and reading for these groups to illuminate patterns and changes in these gaps over time. Historically, some groups of students have achieved less academically (Artiles, Kozleski, Trent, Osher, & Oriz, 2010; Nieto & Bode, 2008). Artiles et al. (2010) used the acronym RCELD to describe these groups of underserved students who are racially, culturally, ethnically, and linguistically diverse. The achievement gap for this study also included students who were considered to be Title I students.

Closing the achievement gap remains a topic of focus for educators and the nation's leaders. Educators face the significant challenge of meeting the mandate of No Child Left Behind (NCLB) legislation which is to have 100% of the students reading on grade level by 2014. President Barack Obama granted several states waivers which have allowed them to escape the harsh sanctions of the NCLB goal by 2014, but those states have been required to provide accountability plans which include improving academic achievement for 15% of their most troubled schools (McNeil, 2011). Though considerable attention has been devoted to the reasons for this achievement gap, the gap

remains. In 2011, the National Report Card showed a significant gap between minority and economically disadvantaged students and their more affluent white peers.

The present study was conducted to further examine student achievement variables centered on seven different themes: (a) standardized testing and school reform initiatives, (b) achievement gap data and identified factors, (c) deficit theory, (d) cultural mismatch theory, (e) extended time accommodations, (f) test anxiety and stress, and (g) timed versus untimed tests. All of these themes were addressed in this study.

Purpose of the Study

The purpose of this study was to investigate the extent to which the performance of Title I third-grade students in a central Florida School District differed on standardized tests administered under timed and untimed conditions. In addition, this study contributed to the research and literature which focused on closing the achievement gap.

Research Questions

This study was guided by the following questions:

1. To what extent do gender differences and students who are Black, who receive free lunch, who are English language learners, and who are in exceptional student education moderate the relationship between timed and untimed testing conditions and standardized tests scores?
2. To what extent does reading achievement differ on standardized tests administered under timed and untimed conditions when controlling for

students who are Black, who receive free lunch, who are English language learners, who are in exceptional student education and have gender differences?

3. To what extent does reading achievement differ on standardized tests administered under timed and untimed conditions for domain specific reading tests for Words and Phrases in Context, Main Idea, Plot and Purpose, Comparisons and Cause and Effect, and Reference and Research when controlling for the same covariates in Question 2?

Assumptions

The following assumptions were made in this study.

1. The 2006 Released FCAT Reading Test was selected because it was determined that this standardized test was recognized as a statewide test that could be used throughout Florida.
2. A quantitative research methodology was an appropriate and useful approach for this study.
3. The study would be able to be replicated in other grade levels and schools to provide an extensive comparison of its results and conclusions.

Limitations and Delimitations of the Study

There were some limitations and delimitations imposed on this research.

Delimitations were created by the researcher in identifying the population which was the

focus of the study. Existing conditions related to the students and the extent to which they were representative of the Title I population were beyond the researcher's control and were, therefore, limitations of the study. The limitations and delimitations follow:

1. This study focused only on the reading achievement of third-grade, Title I students as opposed to any other academic areas.
2. Some of the students who were categorized as Title I may have previously been students of privilege who had recently become economically disadvantaged due, in part, to the recession in the U.S. economy.
3. Not all of the Title I students who took the test were economically disadvantaged even though they attended a Title I school. This may be attributed, in part, to the way the attendance zones were drawn by the school district. Though zoned to attend a Title I school, some of the students may have lived in more affluent neighborhoods.

Background of the Study

On January 8, 2002, President George W. Bush signed into law the No Child Left Behind Act (NCLB) of 2001. The NCLB Act was a reauthorization of the Elementary and Secondary Education Act (ESEA) which was first passed in 1965 and last reauthorized in 1994. ESEA encompassed the Title I Program to aid disadvantaged children and help to close the achievement gap between minority and nonminority students, and between disadvantaged children and their more affluent peers. NCLB contained four basic education reform principles: stronger accountability for results,

increased flexibility and local control, expanded options for parents, and an emphasis on teaching methods that had been proven to work (U.S. Department of Education, 2010c).

The National Report Card was also introduced to provide assessment results. The 2011 Report Card indicated that the average reading scores for fourth grade students have remained unchanged since 2009, and the average reading scores for eighth grade students showed a one-point increase. A closer look at these results revealed that significant academic achievement gaps have continued between economically disadvantaged students who receive free or reduced meals and those students who are not economically disadvantaged because their families make too much money to qualify for the free or reduced meals.

In an effort to meet the requirements of NCLB, the state of Florida has required that all school districts administer the Florida Comprehensive Assessment Test (FCAT) which is a high-stakes, timed test given to students in Grades 3-11 each year. The FCAT measured overall student achievement in reading, mathematics, writing, and science as well as for the various subgroups mandated by NCLB for Adequate Yearly Progress. The 2011 FCAT achievement gap results were consistent with those of the 2011 National Report Card, confirming that there was still much work to do to close the gap.

In 2009, President Barack Obama kicked off his education reform with the Race to the Top initiative (U.S. Department of Education, 2012). This initiative focused on increasing accountability and expectations for all students with an emphasis on providing more funding for education. Those goals included higher standards for early childhood, elementary, and high school programs as well as increased teacher pay and recruitment,

and increased college aid. The reform called for longer school days and a longer calendar year, less difference between the 50 states in standardized testing, and innovative and effective approaches to turn around struggling schools. Awards for the Race to the Top Grants were designed for States that were leading the way with ambitious yet reachable plans for implementing coherent, compelling education reform (U.S. Department of Education, 2012).

President Obama also launched the “Educate to Innovate” campaign which aimed to increase science, technology, engineering and mathematics (STEM) literacy. Federal grants in the amount of \$4.35 billion were to be offered to schools that could innovate in STEM education. Additionally, the private sector has contributed an additional \$260 million in related funding and programs. The goal has been to increase STEM literacy for students and improve the national standing from below average and average to the top rated standing (U.S. Department of Education, 2012).

The Federal Government has placed a great deal of emphasis on providing funding and sharing best practices to help level the academic playing field for economically disadvantaged students. As a result, a vast amount of research and literature has been published on the reasons for the achievement gap between these groups of students. These have been debated in deficit theory and cultural mismatch theory (Gay, 2000; Ladson-Billings, 1994; Nieto & Bode, 2008; Weiner, 2003).

Another area of focus related to this achievement gap was that anxiety and stress may be caused by the increase of high-stakes testing. Researchers have suggested that state testing programs have resulted in increased student anxiety, increased stress, and

lowered motivation (Abrams, Pedulla & Madaus, 2003; Barksdale-Ladd & Thomas, 2000; Segool, 2009). This stress was shown to affect the performance of some students on standardized tests and cause them to perform at levels below their full potential.

Some of the research in this study was a review of the literature which focused on whether the timing of tests places increased anxiety on some students which ultimately affected how they performed (Case, 2003; Lee, 1995; Marquart, 2000; Schlemer, 2007; Segool, 2009; Thurlow & Bolt, 2001; Willner, River & Acosta, 2009). According to Cizek and Berg (2006), students' grade point averages were inversely related to test anxiety levels, and significantly more test-anxious students dropped out of school than non-test anxious students. The results of high-stakes testing have played a vital role in the achievement gap, and it is important that educators work together to close this gap for all students. Continued research is needed into educational practices that might support or hinder the No Child Left Behind mandate with the goal of closing the achievement gap by 2014.

Data were collected for this research by administering the 2006 FCAT Reading Test to Title I students with the variable of timed versus untimed tests. The purpose of this study was to determine to what extent the FCAT performance of the third-grade Title I students differed between students taking the test under timed and untimed conditions.

Definition of Terms

In this section, terms that are relevant in this study are explained and defined:

Accommodation: An accommodation is generally a change in assessment materials or procedures that addresses aspects of students' disabilities that may interfere with the valid assessment of their knowledge and skills (Bolt & Thurlow, 2004).

Achievement gap: A difference in scores between two groups of students, e.g., male and female, Black and White, or Hispanic and White, can only be considered an achievement gap if the difference is statistically significant. There is more than one way that you can note changes in achievement gaps. Whether a gap narrows depends on the amount of change in the average scores for the two student groups. A gap can narrow if the average scores of both groups improve, but one group's scores improve more. If scores for both groups increase at the same rate, the score gap may not change (National Center for Education Statistics, 2011a, para 1).

Adequate Yearly Progress (AYP): The No Child Left Behind Act of 2001 (NCLB) requires all states to report student achievement based on results of reading and mathematics statewide assessments and several other academic indicators for all schools, districts, and the state. The AYP Report provides a disaggregated achievement test results for major racial groups, economically disadvantaged students, students with disabilities, and English language learners. All groups must reach the annual proficiency target for their schools to make Adequate Yearly Progress (Florida Department of Education, 2010c).

Affluent students or affluent peers: For the purposes of this study, affluent students or affluent peers are students whose families earn too much money to qualify for free or reduced meals according to the Federal Guidelines.

Common Core Standards (CCS): On July 27, 2010, the Florida State Board of Education adopted Common Core Standards which are scheduled to eventually replace the Next Generation Sunshine State Standards. These standards define the knowledge and skills students should have within their K-12 education careers so that they will graduate from high school able to succeed in entry-level, credit-bearing academic college courses, and in workforce training programs. States are asked to adopt the Common Core State Standards in their entirety and include the core in at least 85% of the state's standards in English language arts and mathematics. This is the first step in a process timeline that was intended to complete implementation in all content areas in all grades, during the 2013/14 school year, and new CCSS aligned assessments, to replace FCAT 2.0, in the 2014/15 school year (Florida Department of Education, 2010b).

Economically disadvantaged: In common usage, “disadvantaged” is a generic term for those from lower-income backgrounds or the disadvantaged poor. Economically disadvantaged is a term used by government institutions in, for example, allocating free school meals to “a student who is a member of a household that meets the income eligibility guidelines for free or reduced-price meals (less than or equal to 185% of federal poverty guidelines).” (Dallas Indicators, 2010).

Exceptional Student Education (ESE): Educational programs and services for students with special learning needs. This is sometimes referred to as special education.

For purposes of this study, gifted students are not included in the ESE group (Florida Senate, 2011).

Extended time: This is the most frequently used and allowed accommodation on a standardized test that is often given in combination with other accommodations such as read aloud, Braille, or separate testing location. Extended time can range from time and a half, to double time, and to unlimited time on the assessment (Cawthon, Ho, Patel, Potvin & Trundit, 2009). For the purpose of this study, extended time was also used interchangeably with untimed testing conditions.

504 Plan: This is a plan developed to ensure that children who have a disability identified under the law and are attending an elementary or secondary educational institution receive accommodations that will ensure their academic success and access to the learning environment (Florida Senate, 2011).

Florida Comprehensive Assessment Test (FCAT): This is the older version of the test first administered to public school students in 1998 in select grades. There were approximately 200,000 students in public schools in Florida in 2009 who took this version of the FCAT for the last time. This version of the FCAT measured the Sunshine State Standard in reading and mathematics (Grades 3-10), science (Grades, 5, 8, & 11) and writing (Grades 4, 8, & 10) (Florida Department of Education, 2010c).

Florida Comprehensive Assessment Test 2.0 (FCAT 2.0): This the latest revision of the FCAT which measures student success with the Next Generation Sunshine State Standards and includes assessments in reading (Grades 3-10) and mathematics (Grades 3-8) and was implemented in the 2010-2011 school year. The FCAT 2.0 was scheduled to

begin assessing science (Grades 5 and 8) in spring of 2012 (Florida Department of Education, 2010a).

Florida EOC Assessments: The Florida End-of-Course Assessments are computer-based tests designed to measure student achievement of the Next Generation Sunshine State Standards for specific middle- and high-school level courses as outlined in their course descriptions. The Algebra 1 EOC Assessment was administered for the first time in May 2011; the Biology 1 and Geometry EOC Assessments were administered for the first time in May 2012; the U.S. History EOC Assessment was scheduled to be administered for the first time in 2013 with the Civics EOC Assessment scheduled to be administered to seventh graders for the first time in 2014 (Florida Department of Education, 2010a).

Gender Differences: For purposes of this study, gender differences are defined as the differences in the scores on the 2006 FCAT between third grade boys and third grade girls.

Multicultural Education and Training Advocacy (META) Consent Decree: In August 1990, a consent decree was signed giving the courts the power to enforce an agreement between the Florida State Board of Education and a coalition of eight groups regarding the identification and provision of services to students whose native language is other than English. These students are referred to as English Language Learners (ELL) (Florida Department of Education, 1990).

The Next Generation Sunshine State Standards (NGSSS): The Next Generation Sunshine State Standards were adopted in 2007 to help students in Florida effectively

engage, communicate, and compete globally with students around the world. These standards incorporate important skills such as critical thinking, problem-solving, creativity, innovation, collaboration and communication. The NGSSS have significantly changed the standards and benchmarks students are expected to know and be able to do at each grade level. The NGSSS require concepts and content to be taught in far greater depth than the previous Sunshine State Standards and promote a deeper understand of the subject matter with increased rigor (Florida Department of Education, 2010a).

No Child Left Behind Act of 2001(NCLB): This Act was signed into law on January 8, 2002 by President George W. Bush with a priority on closing the achievement gap with accountability, flexibility and choice so that no child is left behind. The foundation is built on four common-sense principles: holding schools accountable for results, giving states and districts flexibility in spending federal funds, using scientific research to guide classroom practices, and involving parents by giving them information and choices about their children's education. This law also mandated that schools must have all students reading on grade level by 2014 or there would be tough sanctions (No Child Left Behind, 2002).

Race to the Top (RTTP): In 2009, President Barack Obama announced his comprehensive education reform with the Race to the Top Grant Program. This initiative focused on four core areas: (a) adopting rigorous standards and assessments that prepare students for success in college and the workplace; (b) recruiting, developing, retaining, and rewarding effective teachers and principals; (c) building data systems that measure student success and inform teachers and principals how they can improve their practices;

and (d) turning around the lowest performing schools. Awards for the Race to the Top grants were designed for states that were leading the way with ambitious yet reachable plans for implementing coherent, compelling education reform. States that received this award were granted funding and a waiver from the No Child Left Behind mandate (U.S. Department of Education, 2012).

Standardized assessment tests: Standardized assessment tests are normed to a specific group, written at a specific grade level, traditionally use a multiple-choice format, and are considered cost-effective and easily scored (Hirsch, 1996).

Sunshine State Standards (SSS): The Sunshine State Standards were approved by the State Board of Education in 1996 to provide expectations for student achievement in Florida. The Standards approved in 1996 were written in seven subject areas, each divided into four separate grade clusters (Pre K-2, 3-5, 6-8, and 9-12). This format was chosen to provide flexibility to school districts in designing curriculum based on local needs. However, as Florida moved toward greater accountability for student achievement at each grade level, the 1996 SSS were further defined. In the subject areas of language arts, mathematics, science, and social studies, the SSS expanded to include grade level expectations (Florida Department of Education, 1996).

Untimed test: For the purpose of this study, an untimed test was defined as a test that allows any additional time a student needs or is given beyond the predetermined scheduled set time to finish a standardized test. It may also be used interchangeably with the term extended time or extended time accommodation (Cawthon et al., 2009).

Test anxiety: Test anxiety consists of psychological, physiological, and behavioral responses associated with concern about experiencing negative outcomes as a result of failure or poor performance on an assessment or in an evaluative situation (Cizek & Berg, 2006).

Timed test: For purposes of this study, a timed test is a test that is administered under certain time constraints and must be completed within the allotted time.

Title I program: This program has provided financial assistance to local education agencies and schools with high percentages of economically disadvantaged children to help ensure that all children meet challenging state academic standards. Federal funds have been allocated through four statutory formulas based primarily on census poverty estimates and the cost of education in each state (U.S. Department of Education, 2010).

Title I school: For purposes of this study, Title I schools had at least 75% of their student population receiving free or reduced meals.

Title I student: Title I students attend a Title I school. They may or may not be eligible to receive free or reduced meals but are considered to be Title I students because they attend a Title I school.

Value Added Models (VAMs): They are specific types of growth models in the sense that they are based on changes in test scores over time. VAMs are complex statistical models that generally attempt to take into account student or school background characteristics in order to isolate the amount of learning attributable to a specific teacher or school. Teachers or schools that produce more than typical or expected growth are said to “add value” (U.S. Department of Education 2012, p. 19).

Conceptual Framework

This researcher has been a principal for more than 12 years and has had many conversations with colleagues about ways in which achievement can be increased for every student, thus closing the achievement gap between economically disadvantaged students and their more affluent peers. One topic of continuing debate has been whether providing Title I students with the accommodation of extended time or unlimited time on the FCAT would increase their overall performance. This question has remained at the forefront of conversations because of the research on the importance of prior knowledge, background experience and vocabulary as it related to reading comprehension, standardized testing, and poverty (Delpit, 2006; Gay, 2004; Kincheloe & Steinberg, 2004; Kozol, 2005; Ladson-Billings, 1994; Nieto & Bode, 2008; Payne, 2008). The researcher's colleagues, typically fellow principals, have had firsthand experience with the differences in children who come from a stable home environment and arrive at school ready to learn versus those children whose home environments are in a constant state of turmoil. Regardless, school leaders have been faced with the challenge of meeting the mandate of NCLB which is to close the achievement gap and have all students reading on grade level by 2014.

This study, therefore, was focused on the achievement gap as it related to economically disadvantaged students and their challenges that could affect their performance on standardized tests and the AYP results for their schools. A review of the history of standardized testing was examined as well as the history of FCAT (Black, 1998; Fletcher, 2009; Gallagher, 2003; Gould, 1981; Minton, 1988; Reber, 2005;

Sherman, 2008; Walsh & Betz, 1995). An emphasis was also placed on the deficit theory and the cultural mismatch theory which explored possible insights on ways educators could help economically disadvantaged students improve their academic performance (Cicchelli & Cho, 2007; Davis et al., 2008; Gay, 2000, 2004; Kozol, 2007; Ladson-Billings, 1994, 2001; Nieto & Bode, 2008). The effects of test anxiety on student performance were also considered along with the research on extended time as an accommodation (Kaflowitz, 1983; Lee, 1995; Segool, 2009; Tobias, 1985). Many Title I students have a multitude of challenges facing them on a daily basis. A closer look at timed versus untimed testing would allow educators to evaluate whether or not untimed tests would help to level the academic playing field for these students. As shared by Paulson and Marchant (2009), this was vital since standardized tests are high-stakes tests which could result in a student's retention, a teacher's termination, a principal's demotion or the replacement of a superintendent.

In third grade, students take their first actual FCAT "high-stakes" test. The FCAT is especially high-stakes for students in the third grade, because a score of Level 1 on the reading section could mean retention for them. Nieto and Bode (2008) wrote that retention decisions are usually based on high-stakes which exacerbate inequality. According to National Center for Education Statistics (2011b), the 2007 data revealed about 21% of Black students in kindergarten through Grade 12 had been retained. That was a higher percentage of retentions than for White students (9%) and Asian students (3%). The percentage of Hispanic students retained was 12%; students who were two or more races averaged a 14% retention rate. The retention rate for boys (14%) was higher

than that of girls (9%) throughout the United States. These findings partially reaffirmed Black's (2004) earlier research in which the highest retention rates were found among boys, minorities, special education students, and those who come from low-income families and live in the inner city (p. 42).

Some school based administrators have observed and discussed the following concerns that may also lead to a third-grade student failing the FCAT and being retained. There are unidentified third-grade students who are taking the FCAT for the first time who may need extended time to pass the FCAT with a Level 2 and above, but that need is not discovered until they have already taken the FCAT test and failed with a score of a Level 1. Similarly, there are ELL students who make a passing score on the FCAT with the extended time accommodation and use of a dictionary. However, once they exit the ESOL Program and they are no longer eligible for the extended time accommodation and use of a dictionary, they fail the FCAT. In addition, there are third-grade students who speak English as a first language but still struggle with reading, writing, and speaking of the English language on the same level as the ELL students. Nonetheless, these third grade students are not given the opportunity to qualify for the extended time accommodation as are ELL students because English is their first language.

Hammond, Linton, and Drew (2007) conducted a comprehensive study on the dropout crisis in the United States for the National Dropout Prevention Center and concluded:

Four factors were found in at least two data sources to significantly impact dropout at all three schools levels. Three of these four factors are individual ones and include low achievement, retention/over-age for grade and poor attendance.

The fourth factor found to be significant across all school levels was the family factor of low socioeconomic status. (p. 5)

Hammond et al.'s (2007) previous study validated Holmes' (2006) belief that student retention was linked to the dropout rate. The dropout rates in 2007 were highest for Hispanic students (21%) followed by Black students (8%), Asian/Pacific Islanders (6%) and White students (5%) (National Center for Education Statistics, 2011c). Thus, performance on standardized tests has become increasingly important with multiple implications for third-grade students, their teachers, and their schools if they do not perform well on standardized tests (Segool, 2009).

Summary

This chapter has provided an overview of the study. The purpose of the study, assumptions, limitations and delimitations of the study were presented, and relevant terms were defined. The research questions which guided the study were stated, and additional information was shared in regard to the background of the study and the conceptual framework. Chapter 2 contains a review of the literature and related research. Chapters 3 and 4 present the methodology used in conducting the study and the analysis of data. Chapter 5 concludes the report of the research and consists of a summary and discussion of findings along with implications for policy and practice and recommendations for future research.

CHAPTER 2 LITERATURE REVIEW

Introduction

This literature review contained in this chapter addresses a number of themes that are pertinent to the study. These themes include: standardized testing and reform initiatives, achievement gap data and identified factors, deficit theory, cultural mismatch theory, extended time accommodations, test anxiety and stress, and timed versus untimed tests as they relate to Title I student performance.

The History of Standardized Testing

Early Initiatives in Testing

It has been reported that the use of some form of standardized testing was documented as early as 2200BC when candidates for Civil Service employment were tested on their knowledge of Confucian Philosophy and poetry through rote memorization on written exams (Fletcher, 2009). By 1803, standardized testing was being used on a large scale by many European universities as part of their entrance requirements. In concert, the United States began wide use of standardized tests for government funded schools in 1845, and Harvard was the first school to use it as an entrance exam in 1851 (Ward & Murray-Ward, 1999). Before the Industrial Revolution, most people worked as farmers and did not need much education or any type of certification. However, the Industrial Revolution promoted a time where literacy became

vital, and employers needed educated people to work for them (Black, 1998). This era brought school-age children out of farms and factories into schools and classrooms and made standardized tests a more conducive way to evaluate large numbers of students (Fletcher, 2009).

In 1845, Horace Mann convinced the Boston Public School Committee (BPSC) to give written examinations to their students. Mann's vision was to gather information about the quality of teaching and learning in urban schools. Mann persuaded the BPSC to administer additional testing to decide when students were ready to go to the next grade, because the results of the testing revealed major achievements for Boston's students. His model of testing was considered to be very successful, resulting in the New York Regents Examinations in 1865, and adopted by most of the schools in the United States (Gallagher, 2003).

Another important milestone for standardized testing occurred in 1911 when a prominent American psychologist Henry Goddard brought Binet's model of testing to the United States. Binet was a French psychologist who created a test of intelligence used to identify when children had learning deficiencies and would not benefit from a formal education (Walsh & Betz, 1995). This examination was revised and expanded in 1916 by Lewis Terman, a professor at Stanford, and renamed the Stanford-Binet Test of Intelligence. Educators, captivated by this movement, began to believe that scientific evidence had been discovered which explained an inherent lack of ability in individuals and propelled the massive use of ranking, sorting, and standardized testing (Gould, 1981).

20th Century Standardized Testing Accomplishments

As the Army prepared for World War I, there was a pressing urgency to efficiently identify the numbers of officer candidates and recruits needed, so they could be placed where their skills would be most useful. Robert Yerkes and his doctoral student, Arthur Otis, created a multiple-choice test, the Army Alpha Test (Rothman, 1995) which measured nearly two million potential soldiers' mental abilities. In 1919, Terman expanded the Army Alpha Test into National Intelligence Tests for school children and sold over 400,000 copies during the first year (Hanson, 1993). According to Hanson, the war validated the use of standardized tests and provided a legitimate way for making decisions about ability and aptitudes of people.

However, the expansion and use of the Stanford-Binet Test by Terman and the Army Test under the guidance of Yerkes carried a stigma of controversy, racial overtones, and cultural biases which have been debated since their introduction (Gould, 1981, 1982). Yerkes and Terman believed that environmental factors did not affect the outcomes of the candidate's performance. Rather, they believed that the tests measured native intellectual ability and that this test would help Americans create a superior race. That was not what Binet anticipated when he created his IQ test. He believed that children who needed extra support could be identified and provided extra environmental support to improve their IQs (Minton, 1988).

Gould (1981, 1982) identified some multiple-choice questions on the Army Alpha Test which reflected a cultural bias. He concluded that recent immigrants and those not

exposed to that type of information would probably not be able to answer items correctly.

Two examples from Gould (1982) follow:

Crisco is a: patent medicine, disinfectant, toothpaste, food product

Christy Mathewson is famous as a: writer, artist, baseball player, comedian (p. 349).

Gould also noted that Yerkes found that average test scores for immigrants consistently increased with years of residence in the United States. Gould (1981, 1982) believed that this validated environmental lack of opportunity as the controlling factor in score differences rather than innate intelligence and that this result was correlated with cultural biases on the tests.

The demand and controversy concerning the use of standardized tests continued to flourish. Gallagher (2003) wrote that social factors such as large urban population enrollments and industrial efficiency pressured schools to sort students by test results and that because of this the creation of the Stanford Achievement Test in 1923 was welcomed. According to Minton (1984), the debate of nature versus nurture was once again fueled when the University of Iowa conducted a series of studies in 1929 which revealed an increase in IQ scores among children who had interventions such as preschool, academic interventions, and adoptive home placements. However, Terman and his followers argued that the Iowa research had methodological flaws and that the research was inconclusive (Minton, 1984).

Concurrently, in 1923, The College Entrance Examination Board (CEEB) gathered to oversee the adoption and administration of a standardized college entrance

and intelligence test for students pursuing admission to college (Gallagher, 2003). In 1925, Brigham of Princeton University refined the CEEB Test which later become known as the Scholastic Aptitude Test (SAT). The American College Test (ACT) was created in 1959 as an alternative to the SAT, and both tests have been revised numerous times for internal validity and reliability (Walsh & Beltz, 1995).

In May of 1954, the U. S. Supreme Court declared in *Brown v. The Board of Education of Topeka, Kansas* that separation of the races in schools was illegal. The decision was met with resistance but was considered to be a milestone which would eventually affect desegregation in schools around the country. Suburban districts were left out of these plans, and it caused many of the higher income white families to leave the desegregated schools for the suburbs (Reber, 2005).

The Elementary and Secondary Education Act (Title I) of 1965 promoted the use of standardized tests by requiring schools to administer them and submit their results to the federal government in order to meet the criteria for federal funds (Chapman, 1988; Thorndike & Lohman, 1990). The federal government expanded testing initiatives through the National Assessment of Educational Progress (NAEP) in 1969 by collecting test data on student performance from almost every state and comparing district, state, national and international results which acquired the nickname the “Nation’s Report Card” (Berliner & Biddle, 1995; Gould, 1981).

According to Sacks (1999), the 1960s civil rights movement spotlighted a heightened awareness of possible testing inequities on the basis of social class and racial/cultural background. Conversely, it presented a dilemma for civil rights activists.

If they eliminated the testing practices that allegedly favored White, middle-class schools, they risked losing evidence of inadequacies in predominantly minority schools (Berliner & Biddle, 1995). The National Center for Education Statistics (NCES) commissioned a study, The Coleman Report, in 1966 to examine concerns of inequality and equity among racially and ethnically diverse student populations. In this study, the most important predictor of school achievement was identified as the student's "general social context" or home background and environment. Critics of these findings concluded this meant that schools had little impact on academic performance, and that home environments and biased standardized assessments were responsible for lower test scores for certain populations. On the contrary, additional investigation of the Coleman Report resulted in discovery of design and data analysis flaws (Berliner & Biddle, 1995). The System of Multicultural Pluralistic Assessment was advocated by Mercer in 1978 continuing the allegations that standardized tests penalized women, students with limited English-proficiency, and racial minorities because of cultural, language or gender differences (Kincheloe, Steinberg, & Gresson, 1996).

Schools, colleges and universities were encouraged to implement more rigorous standards and higher expectations for student performance with the National Commission on Excellence in Education release of *A Nation at Risk* in 1983 (Rothman, 1995). In 1991, President George H. W. Bush monitored state by state progress toward six national education goals and encouraged parents to become informed by reviewing a school's publicized test scores (Gallagher, 2003). In support, President Bill Clinton's Goals 2000: Educate America Act called for a nationwide administration of standardized tests and

outlined what was expected of teachers and students to contribute to higher achievement scores on these tests (Heubert & Hauser, 1999).

21st Century Education Reform and High-Stakes Testing

On January 8, 2002, President George W. Bush signed into the law the No Child Left Behind Act (NCLB) of 2001. The Act contained four basic education reform principles: stronger accountability for results, increased flexibility and local control, expanded options for parents, and an emphasis on teaching methods that have been proven to work. The goal of NCLB was to have 100% of the nation's students proficient in reading, mathematics, and writing by 2014, thus closing the achievement gap for all students. As viewed by some (Gorey, 2009; Sherman, 2008), this new law contained the most sweeping changes to education since the Elementary and Secondary Education Act was enacted in 1965. A number of researchers and authors have expressed their belief that the NCLB effort to close the achievement gap could potentially surpass Brown V. the Board of Education in terms of widening access to high quality public education for students of color and the children of poverty (Hursh, 2007; Noguera & Wing, 2006; Poplin & Soto-Hinma, 2005; Roach, 2005; Rothstein & Jacobsen, 2006; Sherman, 2008).

Rothstein and Jacobsen (2006) observed that the mandate by NCLB for reporting student academic achievement by racial, ethnic, and economic subgroups cast a spotlight on the persistent achievement gap using the measure of adequate yearly progress (AYP). NCLB's bureaucratic complexity and over dependence on standardized tests have been criticized as unfairly punishing schools that serve high percentages of minority students

and students living in poverty because of penalties and loss of federal funds resulting from not making AYP (Delpit, 2006; Kahn, 2008; Kozol, 2005; Noddings, 2005).

According to Noguera and Wing (2006), even the most adamant critics of NCLB have been forced to acknowledge that despite its many shortcomings, NCLB has moved the national conversation about race and education forward. For the first time in the nation's history, schools have been required to produce evidence that they can serve all students.

In 2009, President Barack Obama introduced his education reform with the Race to the Top initiative which also was poised to provide waivers to states releasing them from the mandate of NCLB to have all students on grade level by 2014 (U.S. Department of Education, 2012). This initiative focused on increasing accountability and expectations for all students with an emphasis on providing more funding for education. Those goals included higher standards for early childhood, elementary, and high school programs as well as increased teacher pay and recruitment, and increased college aid. The reform called for longer school days and a longer calendar year. In addition, the goals warranted less difference between the 50 states in standardized testing and innovated and effective approaches to turn around struggling schools. Awards for the Race to the Top Grants were designed for States that were leading the way with ambitious yet reachable plans for implementing coherent, compelling education reform (U.S. Department of Education, 2012). Florida received one of these grants.

As part of the NCLB waiver, Florida had to revamp its teacher evaluation process to also ensure that 40% to 50% of the evaluation was based on student learning growth (Florida Department of Education, 2011a). The state convened a special committee

called the Student Growth Implementation Committee (SGIC) and contracted with the American Institutes for Research (AIR) in 2011 to provide technical expertise. The AIR helped them define the factors that should be included and to identify the type of value-added model (VAM) that would be appropriate for use in the state. Value-added models are supposed to help level the academic playing field by taking into account the proficiency and characteristics of students assigned to teachers. The U.S. Department of Education (2012) defined value-added models (VAMs) as:

a specific type of growth model in the sense that they are based on changes in tests scores over time. VAMs are complex statistical models that generally attempt to take into account student or school background characteristics in order to isolate the amount of learning attributable to a specific teacher or school. Teachers or schools that produce more than typical or expected growth are said to 'add value'. (p. 19)

The U. S. Department of Education (2011a) listed the controlling variables for the value added estimates, e.g., including up to two prior years of achievement scores which was thought to be the strongest predictor of student growth, the number of subject-relevant courses, disability/gifted status, ELL, mobility/attendance, and class size. Socioeconomic status and race were not permitted to be considered under Senate Bill 736.

The American Institutes for Research (2011) explained how the combination of teacher effect and school effect would combine to reach a Teacher VAM number. The teacher effect was defined as that portion of student growth attributed to the classroom teacher. The school effect was that number estimated that considered the school in comparison to others across the state in terms of performance. The teacher effect was

combined with one-half of the school effect to give an overall VAM number (VAM = 2/3 Teacher Effect + 1/3 School Effect) (American Institutes for Research, 2011).

President Obama also launched the “Educate to Innovate” campaign which aims to increase science, technology, engineering and mathematics (STEM) literacy. Federal grants in the amount of \$4.35 billion were offered to schools who can innovate in STEM education and the private sector is stepping up with an additional \$260 million in related funding and programs. The goal is to increase STEM literacy for our students and improve our national standing from below and average to the top (U.S. Department of Education, 2012).

History of Florida Comprehensive Assessment Test (FCAT)

In an effort to meet the requirements of NCLB, the state of Florida has required that all school districts administer the Florida Comprehension Assessment Test (FCAT), a high-stakes timed test given to students in third through 11th grades each year which was developed a few years before NCLB became a requirement. The FCAT measures overall student achievement in reading, math, writing and science as well as for the various subgroups mandated by NCLB for Adequate Yearly Progress and has increased its accountability measures over the years (Florida Department of Education, 2001).

According to the FCAT Briefing Book (Florida Department of Education, 2001), the Florida Comprehensive Assessment Test (FCAT) was initiated in 1972. Over the years, it has gone through many changes and legal challenges as Florida’s state-wide assessment program, but it has prevailed. The first cycle of tests was used to measure

students' acquisition of certain minimum competency skills and was known as the minimum competency testing program. The state implemented a new accountability act in 1976 that required state-wide assessments for students in grades 3, 5, 8, and 11 to measure mastery of basic skills. This promoted Florida's first high school required graduation test which was highly debated in the federal landmark case known as *Debra P. v. Turlington* (1983).

The State Board of Education adopted recommendations from the Florida Commission on Education Reform and Accountability and coined them as the Comprehensive Assessment Design in June 1995, which outlined four expansive assessment areas in reading, writing, mathematics, and creative and critical thinking (Florida Department of Education, 2001). This Design provided a foundation for the development of Florida's curriculum frameworks known as the Sunshine State Standards (SSS) which established a course of action for a state-wide system that encompassed assessment, accountability, and in-service training components which were approved by the State Board of Education in 1996. That same year, the SSS were acknowledged by the Florida Legislature as law in support of the academic standards for Florida students, and the field testing of FCAT in Grades 4, 5, 8, and 10 was authorized for February 1997. These components were implemented to promote higher educational expectations for students and increase their global competitiveness for jobs (Florida Department of Education, 2001). In January of 1998, the reading and mathematics test was given to students in Grades 4, 5, 8, and 10. In 2001, with legislative approval, the FCAT was expanded to include all Grades 3-10.

According to the Florida Department of Education (2010a) the high stakes school grades associated with the FCAT were first calculated in 1999 and were designed to support the A+ Plan for Education. This grading system continued to expand in 2001 and encompassed the FCAT results of the students in grades 3-10. After discussion with various education committees, student learning gains were included in 2002. In 2005, all students were included in the calculation of school grades. The inclusion of additional measures for FCAT science, reading and mathematics learning gains for students in the lowest 25% and the performance of high school students retaking the FCAT began to play an important role in the overall school grade.

The Next Generation Sunshine State Standards were adopted in 2007 to effectively engage, communicate, and compete globally with students around the world. These standards incorporate important skills such as critical thinking, problem-solving, creativity, innovation, collaboration and communication. The NGSSS have significantly changed the standards and benchmarks students are expected to know and be able to do at each grade level. The NGSSS require concepts and content to be taught in far greater depth than the Sunshine State Standards and were designed to promote a deeper understanding of the subject matter with increased rigor. FCAT 2.0 was an updated version measuring student success with the NGSSS and included assessments in reading (Grades 3-10) and mathematics (Grades 3-8) and was implemented in the 2010-2011 school year. The FCAT 2.0 science assessments (Grades 5 and 8) began in the spring of 2012 (Florida Department of Education, 2010a).

High school principals across the state voiced their concerns on the challenges faced trying to make the school grade. As a result, beginning in 2010, high schools had the additional components of graduation rate, student performance, participation in accelerated course work, end of course exams and college readiness included in the calculation of their school grade. The End-of-Course (EOC) Assessments are computer-based tests designed to measure student achievement of the NGSSS for specific middle- and high-school level courses as outlined in their course descriptions. The Algebra 1 EOC Assessment was administered for the first time in May 2011. The Biology 1 and Geometry EOC Assessments were administered for the first time in May 2012. The U.S. History EOC Assessment was scheduled to be administered for the first time in 2013 followed by the Civics EOC Assessment to seventh graders for the first time in 2014 (Florida Department of Education, 2010a).

On July 27, 2010, the Florida State Board of Education adopted the new Common Core State Standards (CCSS) which were planned to eventually replace the NGSSS. The CCSS were defined as the knowledge and skills students should have within their K-12 education careers. The goal was that students would graduate from high school able to succeed in entry-level, credit-bearing academic college courses and in workforce training programs. States were asked to adopt the CCSS in their entirety and include the core in at least 85% of the state's standards in English language arts and mathematics. This was the first step in a process timeline that would lead to complete implementation in all content areas in all grades during the 2013/14 school year, and new CCSS aligned

assessments to replace FCAT 2.0, in the 2014/15 school year (Florida Department of Education, 2010b).

The Florida Department of Education (2012) set a higher bar for the grading scale for the FCAT 2.0, and 20% more of the third-grade students across the state of Florida were expected to score a Level 1 on the FCAT 2.0 and face possible retention. These changes were made to better prepare elementary and middle school students to meet the rigorous standards that are being put in place for high school students. The third-grade reading scores for the students in the state of Florida were released on May 24, 2012. The higher grading scale was evident with the state average for reading proficiency moving from 72% to 56% (Florida Department of Education, 2012).

No Child Left Behind Act of 2001 and Adequate Yearly Progress (AYP)

The Federal Elementary and Secondary Education Act (ESEA), reauthorized as the No Child Left Behind (NCLB) Act of 2001 (2002), required states to evaluate the performance of all students in public schools. The purpose was to determine whether schools, school districts, and the state have made adequate yearly progress (AYP) towards enabling all students to meet the state's academic achievement standards (Florida Department of Education, 2010c). The state of Florida submitted its AYP implementation process which measured overall student achievement in reading, mathematics, writing, and science as well as a number of subgroups mandated by NCLB for AYP (Florida Department of Education, 2010c).

Based on the federal No Child Left Behind Act, schools must meet criteria in 39 different areas to achieve adequate yearly progress. Districts must meet the same criteria as schools except that school grades are not taken into consideration. If one or more of the 39 criteria are not met, the school has not made adequate yearly progress under the federal accountability plan (Florida Department of Education, 2010d).

Table 1

Adequate Yearly Progress (AYP): Subgroup-Based Components

Subgroup-Based Components	Reading 95% Tested	Math 95% Tested	Reading Criteria Met	Math Criteria Met
Total	YES/NO	YES/NO	YES/NO	YES/NO
White	YES/NO	YES/NO	YES/NO	YES/NO
Black	YES/NO	YES/NO	YES/NO	YES/NO
Hispanic	YES/NO	YES/NO	YES/NO	YES/NO
Asian	YES/NO	YES/NO	YES/NO	YES/NO
American Indian	YES/NO	YES/NO	YES/NO	YES/NO
Economically Disadvantaged	YES/NO	YES/NO	YES/NO	YES/NO
Limited English Proficient	YES/NO	YES/NO	YES/NO	YES/NO
Students with Disabilities	YES/NO	YES/NO	YES/NO	YES/NO

Note. 36 Subgroup-based components measure reading participation, math participation, reading proficiency, math proficiency for each of the nine subgroups.
Source: Florida Department of Education (2010).

According to the Florida Department of Education (2010c), as outlined in an AYP Technical Assistance Paper,

An additional criterion that is included in NCLB requires the state definition of AYP to include a graduation rate and at least one additional academic indicator as determined by the state. In Florida, the writing assessment is used as the additional indicator and school grades will be used as an additional condition. Thus, in addition to the three criteria listed above, schools must meet three other criteria: All schools must demonstrate a 1% improvement in the percentage of

students meeting state standards in writing. For purposes of AYP determination, students scoring at level 3.0 and above on the FCAT or at level 4 and above on the SWD alternate assessments are meeting state standards. The writing target is also met if the school has a writing performance rate of 90% or better. High schools must demonstrate a 2% improvement in their graduation rate. The target is also met if a school attains a rate of 85% or better in the current year. The school is not a D or an F: The A+ School Grades are calculated prior to AYP. If a school receives a D or an F, that school does not make AYP. (pp. 2-3)

The 2010 AYP Technical Assistance Paper also outlined two other ways that schools can make AYP:

A school that has met the requirements for participation as well as the state's other indicators (writing, graduation rate, and school grade) but has not met the reading and/or mathematics proficiency targets can still make AYP through a provision in NCLB called Safe Harbor. Safe Harbor applies only to those subgroups that did not meet the reading or mathematics targets. In Safe Harbor, the percentage of non-proficient students must be decreased by at least 10% from the prior year in the subject being evaluated. In addition, the subgroup must meet AYP requirements in writing proficiency and the graduation rate. A school that has met the requirements for participation as well as the State's other indicators (writing, graduation rate, and school grade) but has not met the reading and/or mathematics proficiency targets can still make AYP through a provision in NCLB called the Growth Model. The Growth Model applies only to those subgroups that did not meet the reading or mathematics targets through the status model or safe harbor. (p. 3)

The data used for the foundation of AYP incorporates the assessment results in Grades 3-10 from the Florida Comprehensive Achievement Test (FCAT), alternate assessments given to students with disabilities (SWDs), and records showing participation in the Comprehensive English Language Learning Assessment (CELLA) for English language learners (ELLs). Not making AYP does not mean that a school has failed. It means that the school has not met a certain standard for at least one group of students (Florida Department of Education, 2010c, p. 1). The following table shows the annual objectives

that the state has set for reading and math proficiency based on the ultimate goal to have 100% of all students proficient in both reading and mathematics by 2013-14.

Table 2

Adequate Yearly Progress Benchmarks in Florida

School Year	Reading	Mathematics
2001-02	31	38
2002-03	31	38
2003-04	31	38
2004-05	37	44
2005-06	44	50
2006-07	51	56
2007-08	58	62
2008-09	65	68
2009-10	72	74
2010-11	79	80
2011-12	86	86
2012-13	93	93
2013-14	100	100

Note: Annual Objective for AYP.

Source: Florida Department of Education, 2010.

President Barack Obama’s education goals increased accountability and expectations for all students. Those goals included higher standards for early childhood, elementary, and high school programs as well as increased teacher pay and recruitment, and increased college aid (Jackson, 2009). He also advocated for world class standards in mathematics and science including skills that focused on problem-solving, creativity, critical thinking, and entrepreneurship. His plan called for longer school days and a longer calendar year. In addition, his plan called for less difference between the 50 states

in standardized testing and additional funding linked to specific requirements in competitive Race to the Top grants (Jackson, 2009).

Title I Programs

On April 11, 1965, President Lyndon B. Johnson signed into law the Title I Elementary and Secondary Education Act (ESEA) coining his declaration of War on Poverty which served as a symbol of the federal government's commitment to educational equality and opportunity for all (Robelen, 2005). Borman (2000) stated,

along with the emerging system of social programs of the 1960s, Title I was implemented as the major educational component designed to close the achievement gap between poor children and their more advantaged peers, and ultimately, to break the vicious cycle of poverty. (p. 27)

This was a historic event allowing the federal government to play a vital role in policy decisions that affect public schools. As stated by Baptiste, Orvosh-Kamenski, and Kamenski (2004),

When looking at social injustice and the oppression of "others" in our country, one can look no further than the political leadership of our government to take the moral and ethical responsibility to eradicate such injustices. (p. 34)

At the time of the present study, over 50,000 public schools in the United States benefited from Title I funds. These funds provided additional academic support, interventions and learning opportunities that help low-achieving children master challenging curricula and meet state standards (U.S. Department of Agriculture, 2010). The U.S. Government has allowed states flexibility in distributing Title I funds, but the schools must have at least 40% of their enrolled students meet eligibility for free or reduced lunch. As a result, the

national achievement gap has also been reported and analyzed in terms of race and free or reduced lunch eligibility status. Table 3 provides information on the most current federal poverty guidelines for eligibility for monetary assistance with the National School Lunch Program (U.S. Department of Agriculture, 2010).

Table 3

Annual Income Eligibility Guides (Effective July 1, 2010 to June 30, 2011)

Household Size	Reduced Price	Free Meals
<i>48 Contiguous States, District of Columbia, Guam, and Territories</i>		
1	\$20,036	\$14,079
2	\$26,955	\$18,941
3	\$33,874	\$23,803
4	\$40,793	\$28,665
5	\$47,712	\$33,527
For each additional family member, add. . .	\$6,919	\$4,862
<i>Alaska</i>		
1	\$25,031	\$17,589
2	\$33,689	\$23,673
3	\$42,347	\$29,757
For each additional family member, add. . .	\$8,658	\$6,084
<i>Hawaii</i>		
1	\$23,051	\$16,198
2	\$31,006	\$21,788
3	\$38,961	\$27,383
For each additional family member, add. . .	\$7,955	\$5,590

Source. U.S. Department of Agriculture, 2010.

Economically disadvantaged students or Title I students who qualify for free or reduced meals have had the lowest scores on standardized tests according to the 2011

National Report Card. When educators talk about closing the achievement gap, they are trying to research ways to help economically disadvantaged students catch up with their more affluent peers. This is the reason why Title I students were selected for this study. According to Nieto (2004), “Standardized test scores correlate highly with family income, thus exposing the myth of equality regardless of social class and race” (p. 99). Table 4 displays the median combined SAT scores by family income level.

Table 4

Median Combined SAT Scores by Family Income (400-1600 Scale)

Family Income	Median Score
More than \$100,000	1130
\$80,000 to \$100,000	1082
\$70,000 to \$80,000	1058
\$60,000 to \$70,000	1043
\$50,000 to \$60,000	1030
\$40,000 to \$50,000	1011
\$30,000 to \$40,000	986
\$20,000 to \$30,000	954
\$10,000 to \$20,000	907
Less than \$10,000	871

Note. These figures are based on the test results of 1,302,903 SAT takers in 1999.
Source: *Race, Class and Gender in The United States* (p. 176) by P. S. Rothenberg, 2001, New York: Worth.

The Achievement Gap

This section contains information about the results of the 2011 National Report Card, the 2011 State FCAT, and the 2011 Orange County Public School District FCAT. Student performance comparison data for Black, Hispanic, and White ethnic groups are presented. In addition, student performance of those students who qualified as economically disadvantaged and those students who did not qualify as economically disadvantaged is compared and discussed. A summary of these categories is included with regard to the achievement gap for each report. Also included in this section is a review of the literature on 14 identified factors which have been documented as being strongly associated with the achievement gap. These factors are not an exhaustive list, but they validate that a variety of measures need to be considered when looking at student performance and the achievement gap.

2011 National Reading Report Card

The National Report Card provided 2011 assessment results for students across the nation. These results indicated that the average reading scores for fourth-grade students had remained unchanged since 2009, and the average reading scores for eighth-grade students showed a one-point increase. Significant academic achievement gaps remained between minority students and White students and between students who did and did not receive free or reduced meals.

The results for reading showed that fourth-grade White students earned an average score of 231, but fourth-grade Black students earned an average score of 205 and

fourth-grade Hispanic students increased their average score by one point to 206.

National results for fourth-grade students in reading by eligibility for the National School Lunch Program or the economically disadvantaged Title I students indicated that those students who did not qualify for free or reduced lunch earned an average score of 235 in contrast to the 218 average score of those eligible for reduced meals and the 206 average of those eligible for free meals.

Similarly, the national scores for eighth-grade students continued to reveal the significant racial and socio-economic achievement gap. Eighth-grade White students earned an average score of 274 in reading, Black students earned an average score of 249, and Hispanic students earned an average score of 252. The national reading results for eighth-grade students as compared by their eligibility for the National School Lunch Program reported those not eligible earned an average score of 275, those eligible for reduced meals earned an average score of 261, and those eligible for free meals earned an average score of 250.

The results of the 2011 National Report Card indicated, therefore, a correlation of the achievement gap in student performance on standardized tests with eligibility criteria for the National School Lunch Program. The results showed an additional achievement gap for students who participated in the National School Lunch Program between those who qualified for reduced meals and those who qualified for free meals. These findings continue to validate the need for sustained research in ways to help close the achievement gap and level the academic playing field for economically disadvantaged students.

According to the 2011 National Report Card, female students scored an average of 225 points in reading which was 7 points higher than male students' average reading score of 218. However, this was not significantly different from the reading score gap in 2009 with females scoring 224 and males scoring 218.

2011 State FCAT Reading Results

Florida students' performance on the 2011 State FCAT was consistent with that indicated on the 2011 National Report Card. The significant achievement gap between Whites and minorities in addition to those who qualified for free or reduced meals and those who were not eligible was evident. The third-grade 2011 State FCAT Reading results revealed that 83% of the White students performed at Levels 3 and above, but only 67% of the Hispanic students and even fewer Black students at 56% met that criterion. Of the third-grade students who did not qualify for free and reduced meals, 88% performed at Level 3 and above, whereas only 63% of those who did qualify for free and reduced meals performed at this same level.

The fourth-grade 2011 State FCAT Reading results validated the achievement gap showing an 81% performance at Level 3 and above for White students, 68% at the same level for Hispanic students, and only 55% for Black students. Of the fourth-grade students who did not qualify for free or reduced lunch, 86% scored a Level 3 and above in reading in contrast to 63% who qualified for free or reduced meals.

The same pattern was continued in the 2011 eighth-grade State FCAT Reading results. A total of 67% of the White students met the Level 3 and above criterion. Of the

Hispanic eighth-grade students, 51% met the criterion, and Black eighth graders performed much lower at 36%. For eighth-grade students who did not qualify for free or reduced lunch, 72% of students performed at Level 3 and above. For those who did qualify for free and reduced lunch, their total of 43% was much lower. A closer look at these data solidifies a relationship between student performance on standardized tests and their socio-economic status (Florida Department of Education, 2011b).

2011 Orange County District FCAT Reading Results

The 2011 FCAT results for students in the Orange County Public Schools were similar to the 2011 National Report Card results and the 2011 State FCAT results for achievement gaps between the same groups of students. The third-grade White students once again outscored the other students on the District FCAT report with 85% performing at Level 3 and above. Only 64% of Hispanic students and 58% of Black students performed at Level 3 and above. A total of 87% of third-grade students who were not eligible for free or reduced meals scored at Level 3 and above, but only 62% of third-grade students who were eligible for free or reduce meals were able to meet that criterion.

The fourth-grade district reading results showed that the achievement gap was even wider between Black and White students than at the state level. Of the fourth-grade White students, 85% scored a Level 3 and above in reading, but only 65% of Hispanic and 57% of Black students reached this level. The achievement gap between those who received free and reduced lunch and those who did not was once again highlighted. Of

the fourth-grade students who were not eligible for free and reduced lunch, 87% performed at Level 3 and above. Of those eligible, only 62% met that standard.

The eighth-grade students for the district did not perform as well overall in reading, but a consistent gap was still revealed. A total of 71% of the eighth-grade White students earned a Level 3 and above, but there were lower scores for Hispanics at 45% and Blacks at 35% who reached the Level 3 mark. A gap of 32% existed in meeting the Level 3 and above criterion between those not eligible for free and reduced meals earned 72% and those who were eligible earned 40%.

2011-2012 Orange County Benchmark Reading Results

Orange County Public Schools has administered a Benchmark examination twice each year. This test, untimed, is given to help guide teacher instruction and make student predictions on the FCAT. These results have continued to follow the same pattern as the results from the standardized tests that were shared in the previous sections. The White students consistently outperform the Black and Hispanic students except the achievement gap is not as wide as it on the FCAT.

The third-grade Benchmark reading test results revealed that White students had an average score of 63%, and Black students earned an average score of 48% with Hispanic students earning an average score of 51%. Fourth-grade students' overall Benchmark reading results were higher than those of third graders, but white students still earned higher scores than did Black and Hispanic students. White students had an average score of 76%. Hispanic students came in next with an average score of 65% and

Black students last with an average score of 63%. White eighth-grade students scored an average of 68% on the Benchmark reading examination. Black students scored an average of 57%, and Hispanic students scored an average of 59%.

The achievement gap for third-grade students was smaller on the Benchmark test which was untimed than on the FCAT which was timed. The third-grade gap between White and Hispanic on the FCAT was 16 points and on the Benchmark test it was only 12 points. The third grade gap between White and Black on the FCAT was 27 points and on the Benchmark only 15 points. The gap was also consistently smaller on the Benchmark test than it was on the FCAT for fourth-grade students. On the FCAT, the gap between White and Hispanic students was 13 points and on the Benchmark test it was 11 points. The FCAT gap between White and Black students was 26 points. On the Benchmark test, it was 13 points. Similarly, the eighth-grade Benchmark scores revealed smaller gaps between White students and the Black and Hispanic students than the FCAT. The gap between White students and Hispanic students was 9 points on the Benchmark test and 16 points on the FCAT. The gap between the White students and the Black students was 11 points on the Benchmark test and 31 points on the FCAT.

The untimed Benchmark test revealed that the gap was smaller when this standardized test was untimed. The gap on the FCAT test was wider when the standardized test was timed. In addition, the gap became even wider between White and Black students on the FCAT test than it was between White and Hispanic students.

Identified Achievement Gap Factors

Ford (2006) conducted an extensive literature review and identified the following 14 factors as being strongly associated with the achievement gap: (a) rigor of curriculum, (b) teacher quality and preparation, (c) teacher experience and attendance, (d) class size, (e) technology-assisted instruction, (f) school safety, (g) birth weight, (h) lead poisoning, (i) hunger and nutrition, (j) parent availability, (k) parent participation, (l) reading to young children, (m) television watching and (n) student mobility. Further review of the research on these factors validated their consideration in the present research as being appropriate.

Ford (2006) found that though rigor of curriculum was the strongest predictor of achievement, Blacks and Latinos had the least amount of access to challenging courses such as advanced placement and gifted. Though wealthier students have been able to enjoy a rich and varied curriculum and many opportunities to engage in the arts, many of their less advantaged counterparts have devoted their days to routine worksheet assignments in an effort to boost standardized test scores (Clotfelter, Ladd, & Vigdor, 2006; Kincheloe & Steinberg, 2004; Meir & Wood, 2004). As determined by numerous researchers and authors, providing students with a culturally relevant curriculum that considers the cultural racial and ethnic, social class, linguistic, and religious backgrounds of students to plan inclusive and relevant instruction is what is needed (Cicchelli & Cho, 2007; Davis, Ramahlo, Beyerabac, & London, 2008; Delpit, 2006; Gay, 2004; Griner & Lue-Stewart, 2011; Kozol, 2005; Ladson-Billings, 1994; Nieto & Bode, 2008; Payne, 2008; Schultz, 2007).

Some researchers also shared that finding a way to bring the curriculum to the students through a method called Universal Design was a more effective way of having at-risk learners show what they know on large scale assessments (Acrey, Johnstone & Milligan, 2005). The elements of Universal Design that Thompson, Johnston and Thurlow (2002) proposed are as follows:

1. Inclusive assessment population
2. Precisely defined constructs
3. Nonbiased items
4. Amenable to accommodations
5. Simple, clear, and intuitive procedures
6. Maximum readability and comprehensibility
7. Maximum legibility (Executive Summary section, para. 2)

In addition, Universal Design Assessments (UDA) proposed by the National Center on Educational Outcomes (NCEO) are one method for flexibility in large-scale assessments: “Universally designed assessments are designed and developed from the beginning to allow participation of the widest possible range of students, and to result in valid inferences about performance for all students who participate in the assessment” (Thompson et al., 2002, p. 6).

In regard to teacher quality and preparation, it has been widely known that students who are most in need of well trained teachers tend not to get them (Ford, 2006). Specifically, teachers with the fewest credentials often teach in the lowest performing classrooms, too many of which are comprised of Black and Latino students (Aud, Fox, & KewalRamani, 2010; Kozol, 2005; Orfield & Lee, 2004). According to Houston (2005), highly qualified teachers have not been required to be good teachers pedagogically. They simply need to show they have passed the subject area examinations required by their

state. However, it is vital that teachers of urban students are given an opportunity to learn more about the cultural environments of their urban students, so they can design culturally relevant curriculum and teaching (Gay, 2004; Ladson-Billings, 1994). In addition, data indicated that urban schools have often had higher rates of teacher absences and turnovers (Machtinger, 2007). This lack of consistency has consequently negatively affected the quality of the students' education (Fine & Wise, 1999; Kozol, 2005; Miller, Murnane, & Willett, 2008).

Classroom management, time on task and opportunities for individualized attention have often been compromised for urban schools where class size may be larger due to lack of funding (Ford, 2006; Kozol, 1991). Many students misbehave and display a "whatever attitude" for a number of reasons including poverty, cultural differences, peer pressure, lack of self-confidence, teacher indifference, and low expectation of African American students (Delpit, 2006; Futrell, 2004; Nieto & Bode, 2008). Ford, Granthan, and Gilman (2008) wrote that this attitude may lend itself to another body of work which focused on social variables impacting gifted Black students' achievement. This included discussion of how accusations of "acting White" may cause them to shut down and not contribute to their achievement. When students are viewed as "acting White" by their minority peers, they are thought to act as though they are smarter and better than everyone else.

Nieto and Bode (2008) stated,

A recent survey of 32,000 students in 108 urban schools found that 24 percent of the students surveyed said they felt uncertain about their safety in school, fully

half indicated that they had seen other students being bullied at least once a month (p. 138)

Educators know that learning is more optimal when students feel safe; students who feel unsafe or threatened in any way are not likely to perform at optimal levels (Ford, 2006). Many White parents have enrolled their children in parochial schools because of their concern about violence and the lack of discipline in urban schools (Fine & Weis, 1999).

The income levels of the family either provide the student with advantages or disadvantages which correlate to the school settings and affect the student's academic performance (Paulson & Marchant, 2009). In 2007, the percentage of children living in poverty was higher for Blacks (34%), American Indians/Alaska Natives (33%), Hispanics (27%), and Native Hawaiians or Other Pacific Islanders (26%) than for children of two or more races (18%), Asians (11%), and Whites (10%) (Aud et al., 2010). Evans and Schamberg (2009) shared that the chronic psychological stress of living in poverty can negatively affect the working memory and physical health of a child.

Many economically disadvantaged students come to school hungry. They do not arrive at school early enough to take advantage of the National School Lunch Program which also provides breakfast for those students who qualify. Jensen (2005) emphasized the importance of good nutrition for optimum brain development during the early years. Nutrition was, therefore, determined to have a significant impact on student learning and performance. Gaps in student achievement have also been influenced by more subtle resources such as social capital where the benefits are derived from connection to networks with power and influences which have been associated with the wealthy White

families (Fine & Weis, 1999; Noguera & Wings, 2006). For example, more affluent families would be able to provide their children with better family vacations, extracurricular activities, tutors, and technology (Kozol, 2005; Schultz, 2007).

In addition, students who have poorer reading skills have been at higher risk for special education referral and placement, including such categories as learning disabilities, mental retardation, and developmental delays (Raymond, 2004). Many economically disadvantaged students have been read to less often than their peers who are not economically disadvantaged, because watching television has been more widely utilized as a family activity (Ferguson, 2006; Ford, 2006). Researchers Anderson, Wilson and Fielding (1988) found there was a strong correlation between the numbers of minutes children read per day in relationship to their academic performance and percentile rank on standardized tests. If prior learning has not created sufficient skill and knowledge for the student to be ready to cope with the assigned instructional task, it is much more difficult for learning to occur (Ausubel, 1968; Driscoll, 2005; Nieto & Bode, 2008). The economically disadvantaged child has been viewed as more likely to transfer to different schools during the school year. These transfers may prevent students from establishing trusting and caring relationships with their teachers which are vital to academic success (Payne, 2008).

Theories Impacting Student Performance

Current barriers to closing the achievement gap include definitions, instruments, curriculum, policies and procedures, as well as expectations for diverse students (Hyland,

2005; Metz, 1987; Nieto & Bode, 2008; Weiner, 1993, 2003). Deficit theory and cultural mismatch theory both highlight important research that needs to be considered as the nation moves forward in its efforts to close the educational achievement gap. Yet, researchers have drawn different conclusions on which theory would be the most effective in achieving this goal.

Deficit Theory

Some researchers have concluded that society views minorities through a “deficit theory” lens which would assume that some children because of genetic, socio-economic, cultural, or experiential differences are inferior to other children and they have deficits that must be overcome if they are to learn (Hyland, 2005; Herrnstein & Murray, 1994; Nieto & Bode, 2008; Weiner, 1993, 2000, 2003, 2006; Winfield & Manning, 1992). In this case, teacher education normally provides research that correlates academic failure and socioeconomic status, academic failure and cultural difference, and academic failure and single-parent households (Delpit, 2006; Gay, 2004; Kincheloe & Steinberg, 2004; Kozol, 2005, 2007; Ladson-Billings, 1994, 2001; Schultz, 2007). Embedded in deficit theory is heredity cultural capital which consists of the values, tastes, languages, dialects, and cultures that have the most status and are associated with the dominant group (Noguera & Wing, 2006). However, when teachers help students embrace their native language or dialect in the classroom, students are inspired to participate and engage in learning rather than shut down (Delpit, 2006; Ladson-Billing, 1994; Nieto & Bode, 2008). This is an important factor in validating that teacher preparation programs must

improve, so that all teachers will become culturally competent and hold higher expectations for all students to aid in closing the achievement gap (Ford, 2006).

Cultural Mismatch Theory

Another potential theory that attempts to explain the achievement gap is the cultural mismatch theory (Gay, 2004; Ladson-Billings, 1994; Nieto & Bode, 2008). According to this theory, school failure results because of the cultural clash between the urban student's culture at home which is normally very different from the school's culture and the teacher's culture (Bergeron, 2008; Pransky & Bailey, 2002/2003). Urban students normally have prior knowledge and background experiences that are not always compatible with those of middle class European American students or the normative expectations, programs, and practices of mainstream classroom teachers as spotlighted in *Brown vs. the Board of Education* (Ladson-Billings, 1994; Gay, 2004). Therefore, it is important to consider the varied experiences, values, skills, expectations, and lifestyles that students have when they enter school and what affect these differences may have on their academic achievement (Haberman, 1991; Nieto & Bode, 2008). This theory supports the idea that the more inconsistent the home and school cultures are the less success the student will experience. Providing the students with culturally relevant curriculum that considers the cultural, racial and ethnic, social class, linguistic, and religious backgrounds of students to plan inclusive and relevant instruction is vital for the urban student's academic success (Cicchelli & Cho, 2007; Davis et al., 2008; Gay, 2000, 2004; Kozol, 2007; Ladson-Billings, 1994, 2001; Nieto & Bode, 2008).

Theoretical Controversy

Researchers have disagreed as to which theory--deficit or cultural mismatch--best describes the needs of an urban learner and which issue is most important to address. Ryan (1972) and Heath (2004) criticized deficit theory for blaming the victim and claimed that it was the schools that were to blame. This would basically support the cultural mismatch theory. They expressed the belief that the school's negative perceptions of the language, culture and socio-economic status can adversely affect students' self-esteem and motivation to learn. Deficit Theory has also been considered to be classist and racist and provide an excuse for the detrimental effects of structural inequality, racism, and poverty on student learning (Neito & Bode, 2008). Some researchers have posited that cultural mismatch theory does a more convincing job of explaining the achievement gap than deficit theory and is more promising because it assumes that teachers can create culturally relevant environments to help all students achieve success (Gay, 2000, 2004; Freire, 1998; Ladson-Billings, 1994). Even though cultural mismatch theory is more comprehensive than deficit theory and does not have the same classist and racial overtones, it still falls short. According to Delpit (2006), Ladson-Billings (1994), and Nieto and Bode (2008), not enough attention has been given to biculturation and immigrants, and cultural mismatch theory is in danger of stereotyping students from particular cultural groups.

As a result, addressing one of these theories, or both, still leaves a large area of the academic playing field unlevelled for urban students. The most important way for teachers to help bridge the cultural divide, according to numerous researchers, is to

always look for opportunities that value and celebrate each student with culturally relevant learning experiences (Gay, 2004; Jensen, 2005; Kozol, 2007; Ladson-Billings, 1994; Nieto & Bode, 2008; Noguera & Wings, 2006).

Extended Time Accommodations and Student Performance

The need to analyze the fairness of the use of fixed time limits and accommodations on standardized tests expanded in the late 20th and early 21st centuries. The 1965 legislation for the Elementary and Secondary Education Act (ESEA) as reauthorized under the No Child Left Behind Act of 2001 (2002), which challenged schools to show learning gains for all students, has enhanced this focus. The resulting high stakes testing has brought to the surface new concerns about fairness in testing and a need for closer examination of timed versus untimed testing conditions which could provide advantageous to some students and not others (Case, 2003). For the purposes of this study, the following provides a definition of untimed tests:

an untimed test will be defined as a test that allows any additional time a student needs or is given beyond the predetermined scheduled set time to finish a standardized test. It may also be used interchangeably with the term extended time or extended time accommodation. Extended time can range from time and half, double time or unlimited time on the assessment (Cawthon et al., 2009, p. 3).

Following are reviews of issues related to timed versus untimed testing conditions: (a) federal laws which require accommodations for certain groups of students and (b) two studies which reveal both advantages and disadvantages for lifting the time limits for all students on standardized testing including those without disabilities.

Federal Laws and Accommodations

Thurlow and Bolt (2001) stated that an accommodation is a change in assessment materials or procedures that addresses aspects of disabilities for students that may interfere with the valid assessment of their knowledge and skills. Several federal laws, including Section 504 of the Rehabilitation Act of 1973, the Americans with Disabilities Act of 1990, Title I of the Elementary and Secondary Education Act of 1965, and the Individuals with Disabilities Education Act Amendments (IDEA) of 1977, require that accommodations be provided to individuals with disabilities on statewide and district-wide assessments, so that students with disabilities can demonstrate what they truly know and can do (Thurlow & Bolt, 2001).

School districts in the state of Florida have been required to include all accommodations on an ESE student's Individualized Education Plan. Students with disabilities in Florida may also participate in the John McKay Scholarship program (Orange County Public Schools, 2012). This program provides funding for the ESE student to attend another OCPS public school or a participating private school in Florida. The Florida Senate (2011) approved the expansion of the McKay scholarships to include students with 504 Plans effective July 1, 2011.

The inclusion of accommodations for English language learners (ELL) on statewide assessments was extended in the mid-1900s to help fulfill the Florida Consent Decree Settlement as written in the Florida Statute 233.058 (Florida Department of Education, 1990). This document is also known as the Multicultural Education and Training Advocacy (META) Consent Decree and it addresses the civil rights of ELL

students including their rights to equal access to all education programs.

Accommodations have been provided for students with limited English proficiency because many of them (a) have not yet become proficient in the English language, (b) may take longer to process the language of the test, and (c) may encounter difficulty accessing the content of test items due to unfamiliar language and cultural references (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). Several accommodations have been used to help level the academic playing field for students with disabilities and ELL students. However, the Extended Time Test Accommodation (ETTA) has been the most commonly used, because it allows students additional time for processing test items when the language is challenging or unfamiliar (Stretch & Osborn, 2005; Willner et al., 2009).

Timed Versus Untimed Timed Studies

Marquart (2000) conducted research on the use of extended time as an accommodation on a standardized mathematics test with 97 eighth-grade students with and without disabilities. Included in this study were students identified as at-risk with low academic skills in mathematics and those who were on or above grade level in mathematics. An analysis of the average number of items answered correctly in the extended time conditions revealed that (a) all groups increased the number of correct answers, and (b) students with disabilities and students at-risk in mathematics showed higher gains than students without disabilities. The findings were deemed noteworthy because they also showed that students at-risk in mathematics who did not normally

receive any accommodations had greater increases in their academic performance than did the other groups (Marquart, 2000).

In 2002, a study was conducted under the auspices of Pearson Education of timed versus untimed testing conditions to determine if administering the Stanford 10 under both conditions would yield different results in the areas of reading, mathematics, science and social studies (Case, 2003). There were approximately 390,000 students tested in Grades 1-10. The results revealed that the differences in average raw scores for the students tested under untimed versus those tested under timed conditions amounted to less than one raw score point. In addition, students tested under untimed conditions had an improved performance in Grades 1-6, but students above sixth grade actually performed slightly better under timed conditions. A closer look at the third-grade scores revealed that those students did have a significant increase in the Language and Spelling subtests with a 3.1 average raw score increase in Language and a 1.7 average raw score increase in Spelling.

Pearson Education thought it was important to note that allowing additional testing time would not boost the scores of non-disabled students over students with disabilities (Brooks, Case, & Young, 2003). Even though the results of this study did not display a score increase sufficiently significant to validate lifting the time limits on the Stanford 10, Pearson Education chose to do so for several reasons. First, 48 of the 50 states required the administration of high-stakes assessments that, rather than testing speediness, allowed students to show what they know and could do when measured against criterion-referenced standards. Second, research reviewed and summarized by

Tindal and Fuchs (2000) examined two factors in assessment, ability and rate of answering, and found ability to be more valid. Third, Pearson conducted its own empirical study, examining timed versus untimed testing conditions, and found that the amount of time allowed to complete the test had little bearing on student performance. Finally, a focus on accommodated, standards-based assessments has been supported by the implementation of the No Child Left Behind Act of 2001(2002), the Individuals with Disabilities Education Act Amendments of 1997 (IDEA), and Title II of the Americans with Disabilities Act (ADA) of 1990 (Case, 2003, p. 2).

According to Case (2003), emerging from this study were some examples of students who might benefit from additional testing time, e.g., bright students because they tend to be perfectionists. However, it was noted that any longer than 30 minutes of additional time could work against these students because they tend to second guess themselves and change answers that are correct. Case also suggested that students who performed in the lower quartile may benefit from a significant amount of additional testing time. In addition, students using any type of accommodations such as Braille tests might also require additional time.

Radville (2008) conducted research to evaluate the effectiveness of two intervention programs compared to scientifically based classroom instruction for improving reading fluency and reading comprehension for third-grade students. This was a yearlong study that focused on 57 third-grade students who attended an urban charter elementary school. Data for the 2007-2008 school year indicted that 73.5% of the

students were classified as low-income and that 71.0% were Black, 23.8% were Hispanic, 6.7% were Caucasian, and 2.9% were Multi-racial.

The data analyses involved different reading skills in standardized test scores.

Radville (2008) highlighted that:

Notably, eight students from the Read Naturally (RN) group were designated as English language learners (ELLs) during this study. However, statistical analyses indicated no significant differences between this subgroup compared to other RN group participants for all pretest scores and gain scores for fluency and comprehension outcomes measures. Notably, the majority of the participants in this study speak non-standard dialects of English, including African-American English (AAE) and Hispanic American English (HAE). This may, in part, account for the lack of significant differences between the ELL students and other study participants. (p. 10)

Radville (2008) shared the analyses of the Gates-MacGinitie Reading Test (GMRT) in relation to how the scores compared in a timed setting and an untimed setting. The GMRT measured the students overall comprehension skills and was administered with a multiple-choice assessment format. For the timed setting, only the RN group had a significant gain in scores. The untimed setting led to significant gains in scores for all groups tested. The difference in the scores between timed and untimed conditions was suggested because a student may be conflicted regarding when to apply quick versus slow, careful analysis strategies in reading. Radville (2008) emphasized, “However, it is encouraging that, when provided with unlimited time, the students in this study were able to successfully apply comprehension strategies to novel material” (p. 24).

In summary, the Pearson study did not reveal any significant correlation in student performance between students who took the test under timed or untimed conditions. In contrast, Marquart (2000) and Radville (2008) revealed research that all

groups of students tested increased their number of correct answers with extended time. An investigation of the performance of Title I third-grade students on timed versus untimed tests, as was conducted in this research, was intended to reveal insights as to whether or not extended time or untimed time testing conditions would benefit economically disadvantaged students and help narrow the achievement gap for them on standardized tests.

Anxiety Concerns of High Stakes Testing

As the NCLB accountability measures have increased, there has been a growing concern about the level of anxiety and stress that these high-stakes tests have placed on students, their families, teachers, counselors, other school personnel, administrators and additional stake holders (Duffy, Giordano, Farrell, Paneque & Crump, 2008). These authors illuminated the pressures of high-stakes testing for all concerned in their statement that “mandated assessments and high-stakes tests affect the progress and future of individual students: the reputations, salaries, and careers of teachers; and the reputations and status of schools as critical institutions within local communities” (p. 53).

According to Lee (1995), test anxiety can come from many different sources including psychological, physiological, environmental, emotional and social. He posited that the imposition of time limits introduces anxiety for some students. If students are anxious about their capability of finishing the test within the time limit, this concern could compete for limited working memory capacity and destructively affect their performance on the test. To explain in more detail, “Test relevant and irrelevant

information competes within the limited working memory capacity and causes poor performance in highly anxious subjects” (Lee, 1995, p. 1). The following research studies on test anxiety synthesized some alarming implications for students and educators.

Anxiety Studies

Field studies have found that stress is normally increased for students when they are taking tests such as intelligence tests and course-related examinations that are important to their future (Tobias, 1985). Lee (1995) suggested that highly anxious people have less working memory capacity available because that space has been overloaded with worry and interferes with a person’s cognitive processing abilities.

An earlier study by Tobias (1985) was conducted to examine the contrasting viewpoints accounting for the poor test performance of high-anxious students. One school of thought was that anxiety interferes with the retrieval of prior learning within the inference model. The other premise is that one of two types of deficits in study or testing skills may be the reason for poor test performance. According to Wine (1971), with respect to the inference model, students may experience high anxiety because they need to divide their attention between task demands and person concerns composed of negative outcomes. The students who have low anxiety are able to devote more thought and attention to task demands instead of negative thought patterns. In the study-skills deficit camp, Kirkland and Hollandsworth (1979) questioned “whether anxiety interferes with effective test-taking behavior or whether the lack of effective study skills resulted in

anxiety” (p. 435). According to Tobias (1985), less cognitive capacity may be needed by students who have effective study skills because they may be better organized and have a reduction on the cognitive demands of the task.

Brutch, Juster, and Kaflowitz (1983) found that test-taking skills had an important effect on student performance with regard to essay and multiple-choice tests but were not as important on the performance of a mathematics test. Tobias (1985) concluded, “There are a number of advantages to interpreting deficit and interference effects in terms of cognitive capacity” (p. 140). Tobias also suggested that treatments to help students cope with or reduce anxiety in both areas were beneficial.

In a more recent study, Segool (2009) explored differences in test anxiety on high-stakes standardized achievement testing and classroom testing among 335 students in Grades 3-5 in the state of Michigan. These students were administered surveys at two different times to analyze their self-reported test anxiety in relation to these two different testing environments. The results revealed that students as a whole reported significantly more test anxiety in relation to their state standardized tests versus any classroom testing. Also examined was the relationship between test anxiety and student demographic characteristics. The analyses indicated that student gender and grade level significantly predicted student test anxiety, but student ethnicity, educational verification and socioeconomic status did not. To further explain, females indicated more anxiety towards standardized tests than did males, and the higher the grade level, the less test anxiety was self-reported by students. In addition, teachers were surveyed and reported

that they believed that students experienced significantly heightened anxiety relative to high-stakes testing (Segool, 2009).

Both Segool (2009) and Tobias (1985) suggested that test anxiety may negatively affect student performance. “Interventions to reduce test anxiety may be appropriate for 60-70% of the elementary school student population who are classified as moderately to highly test anxious in relation to high-stakes testing,” (Segool, 2009, p. 125). Both researchers agreed that it is imperative for educators to identify students with test anxiety and develop strategies to help them reduce test anxiety for better academic performance (Segool, 2009; Tobias, 1985). The detrimental outcomes correlated with test anxiety are extensive, including impaired cognitive functioning, scholastic underperformance, and limited occupational educational outcomes (Zeidner, 1998).

Summary

The administration of standardized testing in public schools began in the 1800s and has continued into the 21st century as the primary high-stakes method for measuring students’ academic performance in K-12 schools (Budd, 2009). Educators, at the time of this study, were faced with a very serious mandate placed into law by the No Child Left Behind Act of 2001 (2002) requiring all students to be reading on grade level by the year 2014 in an effort to close the achievement gap. President Obama has granted waivers to several states waivers, thus allowing them to escape the harsh sanctions of the NCLB goal by 2014. Those states must, however, provide stricter accountability plans that

include improving academic achievement for 15% of their most troubled schools (Feller & Hefling, 2012; McNeil, 2011).

The National Report Card provided the 2011 assessment results indicating that the average reading scores for fourth-grade students had remained unchanged since 2009, and the average reading scores for eighth-grade students had increased by only one point. For the same time period, academic achievement gaps have continued between economically disadvantaged students who receive free or reduced meals and those whose families make too much money to qualify for the free or reduced meals programs. An immense amount of research and literature has been published on the reasons for the achievement gap between these groups of students. Many lower income families have tended to read less to their children than higher income families, and this may have put these students at a disadvantage on standardized tests with regard to reading fluency, vocabulary, comprehension, and prior knowledge (Ausubel, 1968; Driscoll, 2005; Ford, 2006; Gay, 2004; Ladson-Billings, 1994; Neito & Bode, 2008). For more than three decades, critics of standardized assessments have questioned the validity of their interpretive judgments for children from low-income families because the questions require a set of knowledge and skills more likely possessed by children from privileged backgrounds (Armstrong, 2010; Kincheloe & Steinberg, 2004; Payne 2008).

There is also a body of research that reveals the negative effect that test anxiety may have on student performance (Cizek & Burg, 2006; Segool, 2009; Tobias, 1985). The detrimental outcomes correlated with test anxiety are extensive, including impaired cognitive functioning, scholastic underperformance, and limited occupational and

educational outcomes (Zeidner, 1998). In addition, it has been determined that the neural systems of poor children may develop differently from those of middle-class children, affecting language development and executive function, or the ability to plan, remember details, and pay attention in school. More alarming is that certain brain functions of some economically disadvantaged nine and 10 year olds pale in comparison to those of wealthy children with the difference being almost equivalent to a stroke (Kishiyama, Boyce, Jimenez, Perry, Knight, 2008).

This study would contribute information to the body of knowledge that (a) places an emphasis on closing the achievement gap for Title I students and (b) advances research on timed versus untimed tests. Many Title I students have a multitude of challenges facing them on a daily basis. A closer look at timed versus untimed testing would allow educators to evaluate whether or not an untimed test would help to level the academic playing field for these students. As shared by Paulson and Marchant (2009), this could be vital since standardized tests are high-stakes tests which could result in a student's retention, a teacher's termination, a principal's demotion, or the replacement of a superintendent. The FCAT is especially important for students in third grade, because a score of a Level 1 in reading could mean retention for them. Third-grade students gain their first exposure to high-stakes tests. There are multiple implications for them, their teachers, and their schools if they do not perform well on these standardized tests.

CHAPTER 3 METHODOLOGY

Introduction

On January 9, 2002, President George W. Bush signed into law the No Child Left Behind Act of 2001 (NCLB). This mandate required that all students read on grade level by the year 2014 in an effort to close the achievement gap. The 2011 National Report Card, however, validated a significant gap remained between minority and economically disadvantaged students and their more affluent White peers (National Report Card, 2011).

The purpose of this study was to investigate the extent to which the performance of Title I third-grade students in a central Florida School District differed on standardized tests administered under timed and untimed conditions. The standardized test that was used in this study was the 2006 Florida Comprehensive Assessment Test (FCAT). In addition, this study was conducted to add to the research and literature focusing on closing the achievement gap.

This chapter identifies the research questions and describes the design, population and sample, instrumentation, data collection procedures, data analysis process and ethical considerations used in conducting this study.

Research Questions

This quantitative research study was guided by the following research questions:

1. To what extent do gender differences and students who are Black, who receive free lunch, who are English language learners, and who are in exceptional student education moderate the relationship between timed and untimed testing conditions and standardized test scores?
2. To what extent does reading achievement differ on standardized tests administered under timed and untimed conditions when controlling for students who are Black, who receive free lunch, who are English language learners, who are in exceptional student education and have gender differences?
3. To what extent does reading achievement differ on standardized tests administered under timed and untimed conditions for domain specific reading tests for Words and Phrases in Context, Main Idea, Plot and Purpose, Comparisons and Cause and Effect, and Reference and Research when controlling for the same covariates in Question 2?

Population and Sample

Previous studies have found that the effect sizes of timed versus untimed tests ranged between $d = .202$ and $d = .440$; therefore, the researcher needed between 166 and 772 participants for sufficient power (Marquart, 2000; Onwuegbuzie & Seaman, 1995; Tsui & Mazzocco, 2007). Once the sample size was determined, and the study's status

had been determined to be exempt from Institutional Review Board approval process (Appendix A), the researcher met with the District's Title I Director for input on the selection of possible Title I schools that were included in this study. Eight Title I - principals were emailed a recruitment letter (Appendix B). Six of the principals agreed to allow their third-grade students to participate in the study. The Title I schools were well above the 75% free and reduced meal percentage which schools needed to qualify for Title I funding in the targeted school district. Five of the schools had between 95% and 97% students receiving free and reduced cost meals, and 86% of the students participated in this program for the sixth school. Two of the principals declined the invitation to participate. The principals who declined expressed concern about their AYP status and their potential FCAT grade for the current school year and that participation in the study would be too much of a burden at that time.

The researcher communicated with each of the six principals who agreed to participate. The testing procedures were reviewed along with the Test Administrator Scripts for both timed (Appendix C) and untimed (Appendix D) tests. The researcher explained that parent permission had to be obtained before any of the third-grade students could be considered for participation in the study, and each school was provided parent permission slips (Appendix E) for all of their current third grade students with a few extra to allow for student mobility. In addition, a student assent script (Appendix F) was provided for each third- grade student. Principals participating in the study were informed that it was important to make certain that only the students who wanted to participate were involved. A student information data sheet (Appendix G) was designed

and provided to collect demographic information about each student as part of the descriptive statistics and to be included in the analyses as covariates and moderators. The data specific variables for which data were collected were gender, race, ESE status, free or reduced lunch status, ELL status, and whether the student took a timed or untimed test. The dependent variable was the FCAT scores and included overall reading performance as well as the reading performance on the domain specific tests for Words and Phrases in Context, Main Idea, Plot and Purpose, Comparisons and Cause and Effect, and Reference and Research. The independent variable was the testing conditions, either timed or untimed.

A combined total of approximately 450 students from the six participating schools were provided with parent permissions slips to take part in the study, and 194 parent permission slips were returned. The test score data, therefore, were gathered from 194 third-grade students in six Title I schools in a large urban school district in Florida who had signed parent permission slips and documented student assent. The third-grade students were randomly selected and placed into either the timed group or the untimed group. Any student who needed special one-on-one accommodations and could not take the test independently was excluded from the sample. In addition, students could be excluded if the principal thought that a student should be excluded from the study. A request was not made by the researcher to gather data on how many students fell into this category. This study was focused on the number of students who met qualifications of parent permission slip and student assent scripts and did participate.

Instrumentation

As outlined by the Florida Department of Education (2007):

The 2006 Released FCAT Reading Test was designed to assess student achievement of the Sunshine State Standards (SSS). Two types of scale scores are reported on the FCAT: (1) scale scores for each grade level (100-500 points), and (2) developmental scale scores (DSS) that span all grade levels tested (0-3000). In addition, the five levels of achievement are reported. Level 1 is low and Level 5 is high. (p. 37)

The 2006 Released FCAT Reading Test was administered in both timed and untimed settings to Title I third-grade students in selected schools. This was a paper and pencil test which measured overall reading comprehension at the third-grade level. The students had to read five different passages and answer the questions that followed each passage. The test had a total of 45 questions with questions that focused on: Words and Phrases in Context; Main Idea, Plot and Purpose; Comparisons and Cause/Effect; and Reference and Research.

According to the Assessment and Accountability Briefing Book (2007), the FCAT meets all professional standards of psychometric quality usually connected with standardized achievement tests. Internal consistency reliabilities for the FCAT are reported using two methods: Cronbach's Alpha and Item Response Theory (IRT) marginal reliabilities. A value of one indicates perfect consistency. The 2006 FCAT had a 0.89 reliability coefficient using Cronbach's Alpha and a 0.92 reliability coefficient using the IRT, both of which are highly reliable. The 2006 FCAT had a validity coefficient of $r = 0.84$ when correlated with the Stanford 9 test (Florida Department of Education, 2007), indicating good concurrent validity. These coefficient ratings support

the claim that the FCAT is technically sound and meets or exceeds the professional standards for standardized achievement tests (Florida Department of Education, 2007).

The Florida Department of Education (2007) implemented the following steps for all of the items included on the FCAT to ensure high content validity:

Educators and citizens judged the standards and skills acceptable.

Item specifications were written.

Test items were written according to the guidelines provided by the item specifications.

The items were pilot tested using randomly selected groups of students at appropriate grade levels.

All items were reviewed for cultural, ethnic, language and gender bias and for issues of general concern to Florida citizens.

Instructional specialist and practicing teachers reviewed the items.

The items were field tested to determine their psychometric properties.

The tests were carefully constructed with items that met specific psychometric standards.

The constructed tests were equated to the base test to match both content coverage and test statistics. (p. 40)

Data Collection Procedures

The participating schools were allowed to administer the 2006 FCAT Reading Test between September 1, 2011 and October 28, 2011. The test was given over a period of two days in two different settings. The students were randomly assigned into either a timed setting or an untimed setting. Test Administrators had a script (Appendixes C and D) that they read to the students indicating that they were either allowed (a) 60 minutes to take the test each day or (b) as much time as they needed during the two days. The students were told to bubble in the correct answers on their bubble sheets and not to mark in their test booklets.

The researcher worked with the District's Senior Director for Accountability, Research, and Assessment in regard to the scoring of the tests. The answer key for the 2006 Reading FCAT was entered into the district's data management system, thus simplifying the task of scoring the tests for the participating schools.

Data Analysis Process

The results of the analyses are reported in Chapter 4 and have been displayed using tables supported by accompanying narratives to show any differences in reading performance for timed and untimed groups. The students' 2006 Reading FCAT overall percentage of correct answers and percentages of correct answers for each of the domain specific tests were used for this analysis. For Grade 3 reading, the following approximate number of correct answers/points needed for each achievement level were as follows: level 1 (low), 0-17; level 2, 18-23; level 3, 24-34; level 4, 35-43, level 5 (high) 44+ (Florida Department of Education, 2007).

Ethical Considerations

Though the study was granted exempt status by the University of Central Florida's Institutional Review Board (IRB) (Appendix A), the IRB protocol was followed strictly to ensure confidentiality of participants' responses. The student data had identifying characteristics removed prior to analysis and were not included in any part of this report. The researcher gained permission from the Florida Department of Education to conduct this research with some copyright limitations (Appendix H). The 2006 FCAT

Released was allowed to be copied and administered to collect the data, but it could not be published as part of the dissertation because the Department of Education does not have full copyrights for all of the passages used in the 2006 FCAT Released. Written permission was given to conduct this research by the selected urban school district and principals of the selected Title I Schools.

Summary

This chapter has presented the methods and procedures that were used to conduct the study. The research questions which guided the study were restated, and the sample of participant schools was described. Instrumentation, data collection procedures, and the processes used to analyze the data were discussed. Ethical considerations were also addressed. Chapter 4 contains a presentation of the results of the data analysis organized around the three research questions which guided the study.

CHAPTER 4 RESULTS

Introduction

The purpose of this study was to investigate the extent to which the performance of Title I third-grade students in a central Florida School District differed on standardized tests administered under timed and untimed conditions. Additionally, demographic variables of interest, i.e., gender, race, exceptional student education (ESE) status, free or reduced lunch status, and English Language Learner (ELL) status, were investigated as part of the data analyses.

Demographic Characteristics for Participating Schools

A total of 194 students in six schools agreed to participate in the study. Frequencies and percentages, provided in Table 5, describe the total sample and provide information by school. Of the 194 students, 85(43.8%) were boys, and 109 (56.2%) were girls. There were 14 (7.2%) exceptional education students and 25 (12.9%) ELL students in the sample. A total of 188 (96.9%) qualified for free lunch. In regard to race, 154 (74.9%) students were Black. The next highest number was 28 (14.4%) Hispanic students.

Table 5

Demographic Characteristics of Participating Schools

Characteristics	Participating Schools by School Number						Total n (%)
	1 n	2 n	3 n	4 n	5 n	6 n	
Boys	25	6	19	18	7	10	85 (43.8)
Girls	32	13	24	9	11	20	109 (56.2)
ESE	2	0	5	3	1	3	14 (7.2)
ELL	1	5	4	8	0	7	25 (12.9)
Free Lunch	56	19	43	27	13	30	188(96.9)
Full Price	1	0	0	0	5	0	6 (3.0)
Black	52	18	38	14	11	22	154 (79.4)
Hispanic	3	1	3	10	5	6	28 (14.4)
White	0	0	1	2	2	1	6 (3.0)
Multiracial	0	0	1	1	0	0	2 (1.0)
American Indian	0	0	0	0	0	1	1 (0.5)
Asian	0	0	0	0	0	1	1 (0.5)
Other	2	0	0	0	0	0	2 (1.0)

Note. ESE = exceptional student education; ELL = English language learners.

Data Analysis for Research Question 1

To what extent do gender differences and students who are Black, who receive free lunch, who are English language learners, and who are in exceptional student education moderate the relationship between timed and untimed testing conditions and standardized test scores?

A factorial ANOVA revealed that there were no significant two-way interactions between the covariates (Black, free lunch, ELL, ESE and gender) and test conditions for FCAT scores. However, when school was included as a moderator, there was a statistically significant interaction between test conditions and schools on FCAT scores.

The relationship between testing conditions and FCAT scores varied for each individual school. The results of the analysis are presented in Table 6.

Table 6

Interactions Between Covariates and Test Conditions

Source	<i>N</i>	<i>F</i>	<i>p</i>
TimedUntimed*Gender	194	1.33	.250
TimedUntimed*Exceptional Student Education (ESE)	14	2.40	.123
TimedUntimed*English Language Learner (ELL)	25	0.83	.363
TimedUntimed*Free Lunch	188	0.18	.671
TimedUntimed*Black	154	2.08	.150
TimedUntimed*School	194	5.62	<.001

Although there were no interactions between the covariates and testing conditions, there were statistically significant differences in main effects for free lunch status, and ESE on FCAT scores as shown in Table 7.

Table 7

Main Effects for Covariates

Source	<i>N</i>	<i>F</i>	<i>p</i>
Black	154	3.39	.067
Free Lunch	188	4.69	.032
English Language Learner (ELL)	25	1.19	.277
Exceptional Student Education (ESE)	14	6.98	.009
Gender	194	0.00	.983

Those who were ESE students ($M = 39.07$, $SD = 27.58$) had lower FCAT scores than those who were not ESE students ($M = 54.70$, $SD = 62.75$). Those students who received free lunch ($M = 36.99$, $SD = 47.24$) had lower FCAT scores than those who did not receive free lunch ($M = 56.78$, $SD = 23.06$). However, there were not statistically significant differences in FCAT scores between students who were and were not Black, English language learners, or female.

Data Analysis for Research Question 2

To what extent does reading achievement differ on standardized tests administered under timed and untimed conditions when controlling for students who are Black, who receive free lunch, who are English language learners, who are in exceptional student education and have gender differences?

A factorial ANCOVA was performed for this research question. School was included as a moderator to analyze if there was a difference in test results and testing conditions among schools. Table 8 indicates that the mean differences between students who took the timed and untimed 2006 FCAT Reading Test varied from school to school after accounting for the covariates $F(5, 172) = 5.62$, $p < .001$. For Schools 1 and 4, those students who took the untimed tests scored higher than those who took the timed test. In contrast, those students who took the untimed tests scored lower than those students who took the timed test for School 2. There was no statistically significant difference for Schools 3, 5, and 6.

Table 8

Means, Standard Deviations, and p Values: Timed and Untimed FCAT Reading Tests

School	Timed	Untimed	Mean Difference	<i>F</i>	<i>p</i>	N
1	33.21 (18.78)	45.88 (18.00)	12.67	6.28	.015	57
2	60.35 (18.56)	27.17 (19.00)	33.18	21.85	<.001	19
3	48.70 (18.57)	40.63 (18.49)	18.07	1.40	.243	43
4	33.05 (18.81)	51.15 (19.36)	18.10	4.90	.038	27
5	55.41 (19.42)	65.67 (20.10)	10.26	0.97	.344	18
6	34.29 (18.54)	41.07 (19.12)	6.78	1.68	.207	30

Data Analysis for Research Question 3

To what extent does reading achievement differ on standardized tests administered under timed and untimed conditions for domain specific reading tests for Words and Phrases in Context, Main Idea, Plot and Purpose, Comparisons and Cause and Effect, and Reference and Research when controlling for the same covariates in Question 2?

A factorial multivariate analysis of covariance (MANCOVA) was used to compare reading performance on the 2006 Reading FCAT between the timed and untimed groups on domain specific tests for Words and Phrases in Context, Main Idea, Plot and Purpose, Comparisons and Cause/Effect, and Reference and Research after accounting for the covariates mentioned in Research Question 2. School 5 did not submit data for the domain specific tests and was excluded from this analysis. School was included as a moderator to analyze if there was a difference in test results and testing conditions among schools. Wilks' Lambda indicated that there was an interaction

between testing condition and school $F(20,489) = 2.21, p = .002$. Therefore, the relationship between testing condition and FCAT scores for each domain specific test varied depending on the individual school. Table 9 displays the F ratios of the interactions.

Table 9

Interactions Between Testing Conditions and School for Each Domain Test

Source	N	F	p
Words and Phrases	166	6.63	$\leq .001$
Main Idea	166	3.80	.006
Plot and Purpose	166	5.51	$\leq .001$
Comparison and Cause and Effect	166	3.17	.015
Reference and Research	166	3.60	.008

Table 10 shows the statistical results for the domain specific test for Words and Phrases. For Schools 1 and 4, those students who took the untimed test scored higher than those who took the timed test. Those students who took the untimed test scored lower than those students who took the timed test for School 2. There was no statistically significant difference between untimed test students' scores and timed tests students' scores in Words and Phrases for Schools 3 and 6.

Table 10

Means and Standard Deviations: F Ratios Between Timed and Untimed Conditions for FCAT Reading Section on Words and Phrases

School	Timed	Untimed	Mean Difference	F	P	N
1	29.66(25.37)	49.33(25.27)	19.67	7.16	.010	51
2	65.57(24.94)	29.33(25.23)	36.24	8.97	.010	19
3	51.22(27.90)	39.15(27.91)	12.07	1.94	.171	43
4	35.81(17.82)	64.65(17.99)	28.84	13.50	.002	23
6	32.40(22.23)	46.48(22.84)	14.08	2.41	.133	30

The statistical results for the domain specific test for Main Idea are shown in Table 11. For School 1, those students who took the untimed tests scored higher than those who took the timed test. In contrast, those students who took the untimed test scored lower than those students who took the timed tests for School 2. There was not a statistically significant difference in timed vs. untimed tests for the domain specific tests for Main Idea for Schools 3, 4 and 6.

Table 11

Means and Standard Deviations: F Ratios Between Timed and Untimed Conditions for FCAT Reading Section on Main Idea

School	Timed	Untimed	Mean Difference	F	P	N
1	33.57 (19.65)	44.80 (19.58)	11.23	3.89	.055	51
2	58.87 (19.67)	30.68 (20.46)	28.19	8.73	.010	19
3	43.21 (24.04)	35.12 (24.03)	8.09	1.81	.284	43
4	36.84 (26.61)	56.21 (26.86)	19.37	2.732	.117	23
6	37.09 (24.73)	36.89 (25.41)	0.20	0.000	.985	30

On the Plot and Purpose section, students in Schools 4 and 6 who took the test without a time restriction scored higher than those who took the timed test. However, the students in School 2 who took the test with a time restriction scored higher than those who took the untimed test. There was not a statistically significant difference in this section for Schools 1 and 3. The results of the analysis for Plot and Purpose are displayed in Table 12.

Table 12

Means and Standard Deviations: F Ratios Between Timed and Untimed Conditions for the FCAT Reading Section on Plot and Purpose

School	Time	Untimed	Mean Difference	F	P	N
1	37.44 (24.04)	46.68 (23.95)	9.24	1.77	.192	51
2	65.05 (27.81)	16.80 (28.14)	48.25	12.80	.003	19
3	47.06 (31.26)	41.44 (31.25)	5.62	.377	.565	43
4	28.57 (32.69)	55.36 (33.00)	26.79	3.46	.080	23
6	27.81 (25.48)	52.09 (26.20)	24.28	5.45	.028	30

Table 13 contains the statistics for the domain specific test for Comparison and Cause and Effect. For School 1, those students who took the untimed tests scored higher than those who took the timed test. For School 2, those students who took the untimed tests scored lower than those students who took the timed tests for school 2. There was no statistically significant difference for Schools 3, 4 and 6 with regard to Comparison and Cause and Effect.

Table 13

Means and Standard Deviations: F Ratios Between Timed and Untimed Conditions on FCAT Reading Section for Comparison and Cause and Effect

School	Timed	Untimed	Mean Difference	F	P	N
1	30.63 (32.51)	47.96 (32.36)	17.33	3.611	.026	51
2	71.58 (24.53)	26.21 (24.82)	45.37	12.368	.003	19
3	46.48 (31.01)	41.68 (31.06)	4.80	0.141	.709	43
4	35.80 (19.16)	43.36 (19.34)	7.56	0.208	.654	23
6	30.32 (25.86)	47.18 (26.55)	16.86	1.374	.253	30

In the data analysis for the Reference and Research section, students in Schools 1 and 4 who took the test without a time restriction scored higher than those who took the timed test. Conversely, the students in School 2 who took the test with a time restriction scored higher than those who took the untimed test. As shown in Table 14, there was no statistically significant difference between the scores of students taking the test under timed conditions and those taking the test under untimed conditions for Schools 3 and 6.

Table 14

Means and Standard Deviations: F Ratios Between Timed and Untimed Conditions for FCAT Reading Section on Reference and Research

School	Timed	Untimed	Mean Difference	F	P	N
1	30.66 (31.46)	52.38 (31.34)	21.72	5.314	.026	51
2	54.17 (26.71)	22.40 (27.03)	31.77	7.136	.018	19
3	45.50 (41.26)	42.94 (41.24)	2.56	0.070	.791	43
4	30.78 (37.61)	54.98 (37.97)	24.20	8.219	.011	23
6	36.67 (35.22)	42.44 (36.22)	5.77	0.300	.589	30

Summary

In this chapter, the purpose of the study was restated, and the demographic characteristics of students in the six schools who agreed to participate in the study were presented. Tables and accompanying narratives were used to present a summary of the data analysis organized around the three research questions which guided the study. Chapter 5 contains a summary and discussion of the findings along with implications for policy and practice and recommendations for future research.

CHAPTER 5 SUMMARY, DISCUSSION AND RECOMMENDATIONS

Introduction

This chapter contains a summary and discussion of the findings as they relate to the research questions and review of the literature in this study. Implications for policy and practice and recommendations for future research will be offered. In addition to a discussion of the statistical results, the insights and first hand experiences of the researcher as they relate to the achievement gap will be shared. The researcher served as principal of an elementary Title I school from January 1998 until June 2008 and was the principal of a higher socioeconomic elementary school at the time of the study.

Purpose of the Study

The purpose of this study was to investigate the extent to which the performance of Title I third-grade students in a central Florida school district differed on standardized tests administered under timed and untimed conditions. Gender differences, students who were Black, who received free lunch, who were English language learners, and who were in exceptional student education comprised the additional variables of interest.

Discussion of Findings

Research Question 1

To what extent do gender differences and students who are Black, who receive free lunch, who are English language learners, and who are in exceptional student education moderate the relationship between timed and untimed testing conditions and standardized tests scores?

The factorial ANOVA revealed that there were no significant two-way interactions between the covariates (Black students, free lunch students, ELL students, ESE students and gender differences) and test conditions for FCAT scores. However, there was a statistically significant interaction found when school was included as a moderator. This meant that the relationship between testing conditions and student FCAT scores varied for each individual school, and a consistent conclusion could not be determined for this relationship.

The study conducted by Pearson Education on student performance on the SAT 10 and timed and untimed conditions had similar results to this study (Case, 2003). It was concluded that the results did not sufficiently increase scores to validate lifting of the time limits on the Stanford 10 (Case, 2003). Pearson Education correlated the FCAT with the Stanford 9 test and found that the FCAT had a high concurrent validity to the Stanford 9. Therefore, it is not surprising that the results found for the FCAT generalize to other measures produced and scored by Pearson Education. In the present study, it was revealed that students in Schools 1 and 4 who took the test untimed scored higher

than those who took the test timed which suggests that some students may perform better on standardized tests in an untimed setting. Pearson Education proposed that students who performed in the lowest quartile on the SAT 10 could significantly increase their standardized test scores with extended time (Case, 2003). In contrast, Marquart (2000) and Radville (2008) found in their research that all groups of students tested increased their number of correct answers with extended time. This inconsistency is a problem because some states are still under the NCLB law mandating schools to have all students reading on grade level by 2014 or there would be tough sanctions (No Child Left Behind, 2002).

Previous studies have produced results indicating that Black students have consistently scored lower than White students and Hispanic students on standardized tests, but the reasons for this are not clear (Florida Department of Education, 2011b; National Report Card, 2011). Ford, Grantham, and Gilman (2008) observed that some Black students may shut down and not contribute to their achievement because they do not want to be ridiculed as “acting White” by their minority peers.

Although there were slight differences between Black students and those of other races for their performance on the 2006 FCAT, the results in this study were not statistically significant at $p < .05$, but was at $p = .067$. It is possible that this effect would have been significant if the sample sizes between the ethnic groups had been equivalent. Of the 194 third-grade students who participated in the study, 154 were Black, 28 were Hispanic, 6 were White, and the remaining 6 students represented the other races that participated in this study.

Previous studies on the 2011 FCAT and the 2011 Benchmark scores revealed that Black students scored lower on both of these standardized tests than White students and Hispanic students, but this was not supported by the findings in this study (Florida Department of Education, 2011b; Orange County Public Schools, 2012b). The academic achievement gap between the Black students and the White students and the Black students and the Hispanic students were smaller on the Benchmark test, which is untimed, than the FCAT which is timed. Furthermore, the gap between Black students and White students on the FCAT was wider than it was between White students and Hispanic students (Florida Department of Education, 2011b).

This broadening of gap for Black students could be, in part, due to the accommodations that the ELL students are given for extended time on the FCAT to help meet the META Consent Decree requirements. Some of the Black students who speak English as a first language and are struggling with the speaking, reading, and writing of the English Language are not automatically given the extended time accommodation as are the ELL students as mandated by the META Consent Decree (Florida Department of Education, 1990).

However, in this study, it was not found that ELL students scored significantly lower than students who were not ELL students, nor did English language learners benefit more from lifted time restrictions. There were only 25 ELL students of the 194 students who participated in this study. The findings may have been different if there were a larger number of ELL students who had participated.

Students who were categorized as ESE were found to perform significantly lower than students who were not categorized as ESE in this study. Many ESE students have a variety of challenges that they face in order to acquire the knowledge they need to pass standardized tests. They have time extensions because they take longer to process information and there are several federal laws that have been passed to help level the academic playing field for these ESE students. Section 504 of the Rehabilitation Act of 1973, the Americans with Disabilities Act of 1990, Title I of the Elementary and Secondary Education Act of 1965, and the Individuals with Disabilities Education Act Amendments (IDEA) of 1977 require that accommodations be provided to individuals with disabilities on statewide and district-wide assessments, so that students with disabilities can demonstrate what they truly know and can do (Thurlow & Bolt, 2001). Although there have been a variety of accommodations provided for ESE students on standardized tests, the Extended Time Test Accommodation (ETTA) has been the most commonly used because it allows students additional time for processing test items when the language is challenging or unfamiliar (Stretch & Osborn, 2005; Willner et al., 2009).

It was also found in the present study that students who were categorized as free or reduced lunch had statistically significant lower FCAT scores than those students who were not categorized as free or reduced lunch students. These findings are congruent with the 2011 National Report Card student performance results as well as the 2011 FCAT student performance results. Evans and Schamberg (2009) shared that the chronic psychological stress of living in poverty can negatively affect the working memory and physical health of a child. Jensen (2005) emphasized the importance of good nutrition

for optimum brain development during the early years. The students who qualify for free or reduced meals are eligible to eat breakfast and lunch for free or at a reduced price, but they may not have much to eat at home which could result in their being hungry and unable to concentrate on their homework assignments. Nieto (2004) found that students who qualified for free and reduced lunch programs had a strong correlation to lower academic performance.

Some researchers blame NCLB for the remedial curriculum that has been taught in some high poverty settings and believe that it has increased the achievement gap (Clotfelter et al., 2006; Ford, 2006; Kozol, 2005). The current researcher has observed that many schools use FCAT preparation programs to drill their students and have a certain amount of time carved out for that each day. In addition, some of the schools believe that the students must memorize huge quantities of factual information, complete stacks of worksheets, and sit behind computers using tedious programs to increase their performance on the FCAT. Ford (2006) found that rigor of curriculum was the strongest predictor of increased student achievement. The students would probably retain more information if the curriculum was presented in a more stimulating way.

Gaps in student achievement have also been influenced by more subtle resources such as social capital where the benefits are derived from connection to networks with power and influences which have been associated with the wealthy White families (Fine & Weise, 1999; Noguera & Wings, 2006). Many families that live in poverty are not able to provide their children with a wide variety of experiences that provide valuable prior knowledge which can help students more quickly comprehend the reading passages

on standardized tests. If prior learning has not created sufficient skill and knowledge for the student to be ready to cope with the assigned instructional task, it is much more difficult for learning to occur (Ausubel, 1968; Driscoll, 2005; Nieto & Bode, 2008). For example, one of the passages on the 2006 FCAT was about the change in seasons and talked about the difference between spring and winter. The more affluent families have probably had the opportunity to visit a different part of the country during the spring and winter. The current researcher has found that many of the students living in poverty have not even had the opportunity to travel out of the state of Florida and experience different seasons. Urban students normally have prior knowledge and background experiences that are not always compatible with those of middle class European American students or the normative expectations, programs, and practices of mainstream classroom teachers as spotlighted in *Brown vs. the Board of Education* (Ladson-Billings, 1994; Gay, 2004). In addition, the researcher has observed that as soon as the more affluent parents realized that their child was struggling in reading they quickly provided a one-on-one tutor which the parents in poverty may not be able to afford. Many of these tutors are classroom teachers who are highly trained and skilled on how to effectively prepare the students for standardized tests, thus increasing the odds for success for the more affluent student.

In this study, there was no significant difference in student performance on the 2006 FCAT when gender was considered despite similar sample sizes of boys ($n = 89$) and girls ($n = 109$). This is in contrast with the 2011 National Report Card results which indicated that fourth-grade female students outperformed fourth-grade male students in reading by 7 points. In addition, the National Center for Education Statistics (2011b)

found that the retention rate for boys was 14% but only 9% for girls. This suggests that the females and males in this study may have had similar reading skills and abilities, since there was not a significant difference found as with the 2011 National Report Card.

Research Question 2

To what extent does reading achievement differ on standardized tests administered under timed and untimed conditions when controlling for students who are Black, who receive free lunch, who are English language learners, who are in exceptional student education and have gender differences?

The factorial ANCOVA results indicated that the mean differences between students who took the timed and untimed 2006 FCAT Reading Test varied from school to school after accounting for the covariates. School was also included as a moderator to determine if there were any significant differences based on testing conditions and student performance for each individual school. Only two of the schools had an increase in scores for the extended time group. Students in another school had lower scores when placed in the extended time group.

The reasons for the inconsistencies in the FCAT scores from school to school are not clear. The researcher's first response was that perhaps the data had been entered incorrectly. However, after thoroughly reviewing the data and the entries, it was found that the data had been entered correctly. The researcher also doubled checked with the schools to make certain that the students had been randomly selected and assigned to testing conditions and that proper procedures had been followed in administering the

2006 FCAT. All schools assured the researcher that the students had been randomly selected and assigned and proper procedures had been followed. This could have just been a case where more of the students who had the stronger reading skills were inadvertently clustered in one group versus the other, and that could have caused variations in student FCAT scores from school to school. It is also possible that the students in the timed group for School 2 took the test more seriously because it was a timed test, and the students in the untimed group did not take it as seriously or became fatigued with time. In addition, some testing groups could have had more distractions than other testing groups, and this could have interfered with the students' concentration and impacted their performance. It could also be possible that students in the untimed group at School 2 had a higher mobility rate than the students in the timed group. The economically disadvantaged child is more likely to transfer from school to school, and these transfers prevent students from establishing trusting relationships with their teachers which are crucial to their academic success (Payne, 2008).

Research Question 3

To what extent does reading achievement differ on standardized tests administered under timed and untimed conditions for domain specific reading tests for Words and Phrases in Context, Main Idea, Plot and Purpose, Comparisons and Cause and Effect, and Reference and Research when controlling for the same covariates in Question 2?

The researcher wanted to take a closer look at the domain specific tests to compare the performance between the timed and untimed groups to see if any insights or patterns were revealed that might help Title I students improve their FCAT scores. The factorial MANCOVA results for the reading performance on the 2006 FCAT between the timed and untimed groups on domain specific tests for Words and Phrases in Context, Main Idea, Plot and Purpose, Comparisons and Cause/Effect and Reference and Research were examined. School was also included as a moderator to determine if there were any significant differences based on testing conditions and student performance for each individual school and a statistically significant interaction was found. After accounting for the covariates, there was only an effect for testing condition for certain schools and the effect of time restrictions varied depending on the domain specific test. Therefore, it could not be concluded from these analyses that testing conditions would consistently result in increases or decreases of student performance on standardized domain specific tests.

The comparison for School 1 revealed that using the factorial ANCOVA, a statistically significant difference was found based on the testing conditions. The untimed group outperformed the timed group after controlling for the covariates. However, that statistical significance was not consistent across all of the domain specific subtests. The Reference and Research subtest for School 1 had the largest mean difference which was 21.72 and was statistically significant with the untimed group outperforming the timed group. The students in the untimed group for Words and Phrases and Comparison and Cause and Effect also outperformed the students in the

timed group. The effects for these three domain tests were consistent with the impact of removing the timed restrictions on the overall composite scores of the FCAT Reading Test.

School 2 had a statistically significant difference in students' scores for all of the domain specific tests with the students in the timed setting consistently outperforming the students in the untimed setting. The treatment effects from the domain specific results were congruent with the overall composite results. For School 2, those in the timed condition scored an average of 48.25 points higher than those in the untimed conditions for the Plot and Purpose subtest. This finding suggests that the students in the timed setting may have had more prior knowledge and experience that related to the passages and that this helped them to answer more of these questions correctly. This prior knowledge could have been, in part, because of real life prior experiences that were related to these passages (Gay, 2000; Ladson-Billings, 2001; Nieto & Bode, 2008). In addition, the students in the timed setting for School 2 may have had less test anxiety than students in the untimed setting because they were better able to utilize specific test taking strategies that helped them navigate through the test more quickly (Segool, 2009; Tobias, 1985; Zeidner, 1998). In addition, a larger number of students with higher reading and comprehension skills may have inadvertently been clustered in the timed group for School 2 even though they were randomly assigned.

The factorial MANCOVA revealed that none of the domain specific subtests were statistically significant for School 3 when comparing reading performance between the timed and untimed groups after accounting for the covariates mentioned in Question 2.

There were no significant improvements in student tests scores when the time was lifted for School 3 which suggests that time restrictions do not affect test scores. This may suggest that the students in both the timed and untimed groups had similar background knowledge and prior experiences. In addition, the students in both groups could have been similar in reading and comprehension skills as they related to student ability, and this could have resulted in similar student performance on the tests. These findings may also suggest that the teachers in School 3 had worked more closely together to plan the pacing and the delivery of culturally relevant curriculum and test taking skills. The teachers for both groups of students could have had similar years of experience and been provided staff development which would help their students yield comparable results on standardized tests.

Using the factorial ANCOVA, it was found that the students attending School 4 who took the untimed test scored higher than students who took the test timed. However, this effect was not found consistently across all of the domain specific subtests. Student performance on subtests for Main Idea and Comparison and Cause and Effect was not found to be statistically significant between the timed and untimed conditions. However, School 4 showed the greatest treatment effect for Words and Phrases, in which students in the untimed group outperformed the time group by 28.84 points. Students in the untimed group also outperformed students in the timed group on the Plot and Purpose and the Reference and Research subtests.

School 5 did not submit data for the domain specific subtest, so there was not an analysis performed using the factorial MANCOVA for this school. The researcher

consulted with the school principal and found that the school had experienced some extenuating circumstances which prevented them from submitting the domain specific subtest data.

The factorial ANCOVA for School 6 did not reveal a statically significant difference in overall composite reading performance between the timed and untimed groups after controlling for the covariates. However, the factorial MANCOVA did reveal that there was a statistical difference in student performance on the domain specific subtest for Plot and Purpose with the untimed group outperforming the timed group. This subtest may have been more challenging for the third-grade students at School 6 because they may not have had as much exposure to Plot and Purpose questions as they might have had for the type of questions in the other subtests. In addition, students need to be able to use higher order reasoning skills to determine the plot and purpose of a passage, and this could take much longer than other types of questions on the FCAT. According to Goleman (1995), students may find different kinds of tests easy or particularly difficult because they each have their own baselines of anxiety and comfort.

A closer look at the average mean scores for the domain specific subtests revealed that students in the timed group at School 2 outperformed all of the other schools in each domain specific subtest. In contrast, students in the untimed group for School 2 had the lowest average mean score of all the other participating schools on every domain specific subtests. A closer look at the domain specific subtests validated the original findings that it could not be concluded from these analyses that testing conditions would consistently result in increases or decreases of student performance on standardized tests. It may,

however, be important to learn from School 2 what factors contributed to the consistency of the students in the timed group outperforming students in the untimed group on overall FCAT reading performance as well as on the domain specific subtests.

Even though the findings were not consistent in all areas, this study provides valuable information on testing conditions and standardized test scores. These results can be used to help further research on the academic achievement gap.

Conclusion

The 2011 National Report Card found that students who fall under the categories of Black students, free lunch students, ELL students, and ESE students did have significant gaps in their performance on standardized tests and perform lower than other students. In addition, the 2011 National Report Card revealed that there was a significant difference in student performance for gender differences with females outperforming males. These findings are in contrast to those in this study for Black students, ELL students and gender differences which did not reveal any significant differences. The significance differences may not have been found for these covariates because the sample population for this study was not as diverse or as large as the sample population for the 2011 National Report Card. However, if these groups of students were found to perform significantly lower than other groups, policy should be reviewed to make certain that each category is getting the necessary accommodations.

The students in this sample were all considered Title I students. A total of 188 of the students qualified for free or reduced lunch, and only 6 did not. Therefore, it is likely

that the students, for the most part, had similar prior knowledge and background experiences. The researcher believes that the greatest influence on the variations in student performance for this study had to be based on what was happening at the schools and in the classrooms. This validates the importance of providing students with rigorous curriculum that is also culturally relevant (Ford, 2006; Gay, 2000; Griner & Lue-Stewart, 2011; Ladson-Billings, 1994).

These results once again validated the importance of what was happening at each individual school and in each classroom with the students. A closer look at the domain specific subtest using the MANCOVA revealed that one school scored significantly higher in the timed setting. However, two of the schools that were found to score significantly higher in the untimed setting based on their overall composite scores did not show that same consistency across all of the domain specific tests. This finding suggests that some of the schools might have more thoroughly covered some of the domain specific subtests than other subtests and presented them in a way that was more culturally relevant than some of the other schools. Nonetheless, the researcher concluded that since the results revealed that some students did score significantly higher in the untimed setting that further attention should be devoted to when and how to allow this accommodation on standardized tests. In addition, it is also recommended that each school have a plan in place to make certain that all areas of the domain specific subtests are thoroughly taught and reviewed before the students take the FCAT each year.

In summary, it could not be concluded from this study and these analyses that testing conditions would consistently result in increases or decreases of student

performance on standardized tests. The achievement gap is based on a wide variety of variables as presented in the literature review. However, if there is even a slim possibility that allowing students extra time on the FCAT could mean the difference between their being retained or being promoted, graduating from high school or not, and their teachers keeping or losing their jobs, it is certainly worth consideration as an accommodation that has the potential to benefit some students and help reduce the achievement gap. Most importantly, educators play a vital role in the academic success of their students and must make certain that the curriculum is rigorous and culturally relevant in order to help their students reach their fullest potential.

Limitations

There were several limitations identified in this study. Participants in the study were limited to third-grade students in six selected Title I schools who could only participate in the study if their parents or guardians returned permission slips. As a result, the statistical analysis was based on a limited number of participants from a much larger potential sample. Also, the six Title I schools selected and used in the study were chosen because of the large percentage of Title I students for whom English was the first language. The majority of the participants were identified as Black, and this limited the ability to make comparisons based on individual racial performances. This resulted in a less diverse student sample than the actual population at the various Title I schools. This could have contributed to some of the differences or lack thereof at different schools.

Another limitation of this study was that the researcher was dependent on the participating schools to follow procedures, as outlined for them, in the collection of data. The researcher was not able to personally monitor the process at the participating Title I schools and relied on school staff as directed by their principals to follow the written directions provided for them. The statistical analyses in this study were used to test for differences in FCAT scores under timed and untimed conditions. If differences varied based on the covariates, the analyses cannot explain why differences exist. Therefore, causal explanations described in this discussion are purely speculative. Finally, because this research was limited to one district and six Title I schools, the results of this study are only generalizable to Title I schools with similar demographics.

Implications for Policy and Practice

The achievement gap will continue to remain a topic of focus for educators and the nation's leaders because the United States is becoming more diverse each year (Nieto & Bode, 2008). This will have a profound influence on the increasing challenges that students will bring with them as they enter the nation's classrooms. The NCLB Act has been instrumental in shining a spotlight on those students who have not been achieving as high as other students on the standardized tests. Those students are primarily those from racially, culturally, and linguistically marginalized and poor families (Nieto & Bode, 2008).

The Race to the Top waivers that have been granted to some states permitting them to avoid some of the NCLB mandates could ultimately put schools and educators

under a more intense spotlight, allowing swifter and tougher consequences for schools, educators and students than NCLB (Feller & Hefling, 2012; McNeil, 2011). According to McNeil (2011), the parameters of the waivers keep schools under the microscope to close the achievement gap, and low performing schools will continue to struggle to achieve higher scores on standardized tests. The waivers require a stronger accountability for teachers as between 40% and 60% of their overall evaluations have been required to be based on their students' standardized test scores (Florida Department of Education, 2011a).

Payne (2008) has shared that students of poverty have very different cultural backgrounds than their classroom teachers, and this contributes to lower academic performance. In this study, students who were on free or reduced lunch scored lower than students who were not on free or reduced lunch. According to Cultural Mismatch Theory, school failure results because of the cultural clash between the urban student's culture at home that is normally very different from the school's culture and the teacher's culture (Bergeron, 2008; Pransky & Bailey, 2002/2003). It may be helpful if school districts provide their teachers and staff with professional development in the area of multicultural education to equip them with the knowledge and teaching strategies needed to help level the academic playing field and bridge the cultural divide for students of poverty. This is an important factor in validating that teacher preparation programs must improve so that all teachers will become culturally competent and hold higher expectations for all students, thus aiding in closing the achievement gap (Ford, 2006).

Implications in relation to standardized high stakes test scores have become significantly important for students with the mandatory retention requirements and pending policy on end-of-the-year course examinations. The waiver as outlined in Senate Bill 736 ties educators' evaluations and their job security to their students' standardized test scores. This research continued to validate that lower test scores are linked to students who live in poverty. This correlation could potentially make it more difficult to find teachers and administrators who want to work at Title I schools. Thus, the result could be less experienced teachers and a higher teacher turnover rate in Title I schools because this is often where the openings are. Ford (2006) found that a high teacher turnover rate can negatively influence the performance of students on their standardized tests. In addition, as the bar continues to rise for student performance on standardized tests, high teacher turnover rates could increase the retention rates and dropout rates for students of poverty.

The value added model (VAM) places an even higher priority on student performance on the FCAT for the teacher evaluation system than in previous years. The results of the analysis for main effects for this study indicated that students who were on free or reduced lunch scored significantly lower than students who were not on free or reduced lunch. The VAM for the new teacher evaluation does not include any extra consideration for students of poverty even though the 2011 National Report Card and this study validated those students who received free or reduced lunch scored lower than their more affluent peers.

Consequently, educators must consider all factors when teaching struggling students as well as what testing conditions would allow each student to perform at their fullest potential on standardized tests. This research revealed that there was a statistically significant difference in main effects for students who received free lunch and students who did not receive free lunch with students who received free lunch scoring lower. The majority of students who attend Title I schools receive free lunch, so they may benefit from additional time accommodations.

The META Consent Decree (1990) may need to be revisited to address the needs of any student having sufficient difficulty speaking, reading, writing, or listening to the English language. This would afford these students the same accommodations and civil rights even if they speak English as a first language. Consideration should be given to extend these rights and accommodations on standardized tests for those students who speak English as a first language if there are groups of students who are found to score significantly lower than other groups of children as did students of poverty in this study.

The nation must work collectively to examine all possible avenues in closing the achievement gap and helping all students reach their fullest potential. District officials and educators should meet with local legislators to make them aware of the growing concerns of the effects of high-stakes testing on students, educators, parents, schools and their communities.

Recommendations for Future Research

Future studies should concentrate on high poverty schools and districts that have been successful to learn more about their actions in regard to closing the achievement gap despite their challenges. Successful patterns and strategies in those schools and school districts should be documented and researched, so they can be duplicated to help others who have not been as successful.

In addition, this study should be replicated to include a larger sample size with more schools and other districts. This replication could include research into the extent to which the performance of Title I third-grade students differs on standardized tests administered under timed and untimed conditions for the subjects of mathematics and writing. Future research should also be conducted with students in Grades 3-10 to determine if the relationship differs by grade level.

A more diverse study should be conducted to make certain that there are enough participants in each ethnic group to permit the examination of differences in student performance by ethnicity on timed and untimed standardized tests. This study should also include gifted students as one of the covariates.

The need for teachers to make certain that the curricula they are teaching is culturally relevant was one of the factors identified as important in lessening the achievement gap. Teachers could take part in a specially designed staff development to help them become more culturally aware of the strategies that would be beneficial to their students, and students' standardized test scores could be studied to investigate the benefits accrued.

Data could also be collected in regards to the ability of students who speak English as a first language to pass the Comprehensive English Language Learning Assessment (CELLA). This test is given to students who are English language learners to monitor their progress in learning the English language and provide them with accommodations on standardized tests. If students do not pass the CELLA test, they could be provided the same accommodations as an ELL student to see if their performance increases on the FCAT. This test is given to ELL students to monitor their progress in learning the English language and provide them with accommodations on standardized tests.

A study should be conducted on the difference in anxiety levels for students who receive free or reduced lunch and those that do not. Those results should then be analyzed using the testing conditions of timed and untimed tests with their performance on standardized tests.

The Florida Department of Education (FLDOE) should contact Pearson Education about their study and ask them to replicate that study using the FCAT instead of the SAT 10. The FLDOE should discuss with Pearson Education why they decided to lift the time limits on the SAT 10. In addition, the FLDOE should inquire about the suggestions by Pearson which state that students who score in the lowest quartile would benefit significantly from extra time on the test. They should also analyze if lifting the time limits on the FCAT would help to close the achievement gap.

Researcher's Reflections

This research directed me to reflect on my first hand practical experiences as a principal. These experiences have led me to believe that some of the challenges of poverty can make it much harder for a student to be successful on high-stakes standardized tests than for a student who does not have those same challenges. As a principal of a Title I school from January of 1998 until July of 2008, I worked in a school that had free and reduced rates at 90% and above and ethnic minority rates of 99% and above. Under my leadership, the school earned every letter grade possible on the FCAT from “F” to “A.” For the 2001-2002 school year, the school earned an F on the FCAT. I asked the district office to find another school in the state of Florida that had more than 90% free and reduced lunch and more than 90% Black that had earned an A so that I might identify strategies used to achieve that level of success and implement them in my school. No school was found that met both criteria.

I did learn of one school that was 90% Black and had earned a letter grade of “B,” but they had a very low free and reduced lunch rate and were not a Title I school. This confirmed for me that the majority of the students who attended the school probably came from families who had the financial resources and support to provide them with a solid foundation for academic success (Neito & Bode, 2008; Paulson & Marchant, 2009).

I was able to identify a neighboring Title I school in Orange County, with 90% Black students and an approximate 85% free and reduced lunch rate that had earned a B. This was very inspiring because of the similar challenges the two schools shared. In speaking with the principal of this Title I school, a variety of factors were identified

things that contributed to the school's success in this particular year—there was no magic solution. The principal believed that the FCAT B grade was earned due, in part, to (a) the more experienced teachers who had been placed in the FCAT grades; (b) the alignment of the curriculum to the FCAT rigorous standards; (c) extended tutoring programs, (d) increased parental involvement; and (e) technology-assisted instruction. These are some of the achievement gap factors identified by Ford (2006) as making a difference in student achievement. However, the increased student, teacher, and administrator mobility percentages, among other factors, made it impossible for the school to repeat its B performance the following year and for several years thereafter. The school was able to earn an A grade for two consecutive years during 2008-2009 and 2009-2010 with a 99% minority population and more than 95% students on free and reduced lunch.

My mission, for the 2002-2003 school year was to earn at least a B grade on the FCAT and never have my Title I school branded with the scarlet letter of F again. In earning the F grade, a D was missed by only three points. The average FCAT reading and FCAT mathematics scores were higher than some schools that had earned Cs. My school's downfall rested in the total number of points earned on learning gains in reading and mathematics.

My staff and I continued to analyze the school's data and found that several teachers in the FCAT grades (third, fourth, and fifth) had significantly lower or higher class averages than did other classes at their respective grade levels. In the analysis, the impact of revolving substitutes was observed in that classes with lower averages had

revolving substitutes (Ford, 2006; Kozol, 2005). It was also determined that, in some of the classes with higher class averages, teachers consistently provided more rigorous curriculum for their students and did not remediate as much as did other teachers or as suggested in the literature review conducted for this study (Ford, 2006). For example, one fourth-grade teacher required her students to read chapter books and requested sixth-grade curriculum for some of her higher achieving students. As I monitored this particular teacher more closely, I also found that she utilized every possible moment to keep her students engaged in learning with culturally relevant curriculum (Delpit, 2006; Gay, 2004; Kozol, 2005; Ladson-Billings, 1994; Nieto & Bode, 2008; Payne, 2008). This teacher and others who exercised similar rigor and had high expectations for their students became models in demonstration classrooms where the capabilities of the school's Title I students were validated in terms of achievement (Ford, 2006; Kozol, 2005).

I worked diligently with my staff, parents, community, students, and the district to increase parental involvement, extend tutoring hours, implement technology-assisted instruction, provide a more rigorous curriculum, and motivate and encourage all stakeholders with the vision of educational excellence (Ford, 2006; Meir & Wood, 2004). During that period, my staff and I realized there were struggling students who needed more time to finish FCAT simulations. Though these students did not fall under the ESE or ESOL categories for an extended time accommodation, we were able to write 504 plans for students whom we could document needed additional time to be successful on the FCAT. We knew this was a very important piece of the puzzle for their academic

success. The results revealed improved test scores for the majority of the students that were able to qualify for additional time on the FCAT. However, it was still very difficult to document the need for all the students that we felt would really benefit from additional time.

In 2002-2003, the school earned a B grade on the FCAT. Setting the target even higher resulted in an A grade for the 2003-2004 school year. Both of these goals were accomplished because of the hard work of dedicated teachers, staff, parents, community, partners, volunteers, district support. One partner in education, Universal Studios, made a significant difference by adopting the school in the year of the F grade. Universal contributed numerous resources which provided for additional tutoring and incentives for the school's students and their parents. We also began to work closely with our local churches, the Boys and Girls Club, and our community library to provide additional learning opportunities for our students.

Unfortunately, we were unable to sustain that A grade and dropped to a C grade in 2004-2005 and 2005-2006 due, in part, to increased student, teacher, and administrator mobility and other factors similar to those of the neighboring school previously mentioned. However, the school was able to earn an A grade once again for the 2006-2007 school year and did not decline to the F level during my tenure. We continued to work diligently to help level the academic playing field for our students of poverty and provide extended time to those students for whom we could document the need.

In July of 2008, I was assigned as principal of a school located in a more affluent community with very different demographics. This school had earned an A grade on the

FCAT since the 2001-2002 school year. The free and reduced lunch rate ranged from 26% to 41%, and the minority rate ranged from 50% to 59%. This was much different from my previous experience.

In this school, the majority of the students came to school ready to learn, and many of the teacher's children attended the school. Many of the students came from families whose parents were doctors, lawyers, engineers, college professors, business owners, etc. According to Paulson and Marchant (2009), the income levels of the family either provide the student with advantages or disadvantages which correlate to the school settings and affect the student's academic performance. Parental support was high. Parents enjoyed coming to school and providing classroom presentations which validated for the students the importance of a good education. The PTA was strong and provided wonderful activities that the families could enjoy together. The Young Men's Christian Association (YMCA) held a before- and after-school program on campus for students of working parents. There were always more than enough volunteers when needed. If substitute teachers were used, the majority of them lived in the community and already knew many of the children (Fine & Weis, 1999; Noguera & Wings, 2006).

Most of the teachers were veteran teachers or had taught for at least five years. They seemed to have a strong bond and a family type atmosphere prevailed in the school. Teachers had done a great job of aligning the curriculum. The curriculum was particularly rigorous in reading and writing and needed only a little attention in mathematics and science.

In contrast, the majority of the teachers at my Title I school had taught less than three years. Teachers frequently accepted positions at that school in order to get established in the Orange County Public School System. High teacher and student turnover made creating a family-like atmosphere difficult.

The more affluent school seemed to have the majority of the structures, resources and supports in place for their students as researched by Ford (2006). Many of the students who needed tutoring had families that were prepared to pay for it, and students were often assisted in their homes by one of teachers for a fee. We did work together to put intensive interventions in place for the students who needed them. Several of those students transferred from other schools and were on free or reduced lunch. Struggling students, whenever possible, were accommodated with more time on the FCAT when it was possible to justify and provide it.

The teachers at both schools are basically being held to the same standards with the new Instructional Personnel Evaluation System. They know that the results of the FCAT will be analyzed and their 2011-2012 teacher evaluations will be based (40%) on their students' FCAT scores. This underscores the importance of looking at all possible variables to help close the achievement gap including extended time for standardized tests for every student. At both schools, I have had students so sick they could not come to school on the first day of FCAT, came in crying, or became physically ill during the test. Regardless of the different challenges, I know that teachers at both schools care about their students and will continue to do their best to meet the needs of all children.

Summary

This study has added to the literature and research by analyzing and examining the relationship between third-grade Title I students and testing conditions and their academic performance on standardized tests. The results lend support to the growing body of evidence that educators must continue to pursue all avenues to help level the academic playing field for students who come from racially, culturally, and linguistically marginalized and poor families as it relates to closing the achievement gap and standardized testing.

APPENDIX A
APPROVAL OF EXEMPT HUMAN RESEARCH



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

Approval of Exempt Human Research

From: **UCF Institutional Review Board #1
FWA00000351, IRB00001138**
To: **Ruth E. Haniff and Co-PI: Martha L. Lue**
Date: **July 29, 2011**

Dear Researcher:

On 7/29/2011, the IRB approved the following activity as human participant research that is exempt from regulation:

Type of Review: Exempt Determination
Project Title: The Impact of Timed Versus Untimed Standardized Tests on Reading Scores of Third Grade Students in Title I Schools
Investigator: Ruth E. Haniff
IRB Number: SBE-11-07785
Funding Agency:
Grant Title:
Research ID: N/A

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

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

On behalf of Kendra Dimond Campbell, MA, JD, UCF IRB Interim Chair, this letter is signed by:

Signature applied by Joanne Muratori on 07/29/2011 09:19:46 AM EDT

IRB Coordinator

APPENDIX B
RECRUITMENT LETTER

Ruth Haniff
PO Box 780824
Orlando FL 32878

July 11, 2011

Dear Educator:

My name is Ruth Haniff and I am a doctoral candidate in Education at the University of Central Florida. I am working under the supervision of Dr. Martha Lue-Stewart. The purpose of this study is to investigate the extent to which the performance of Title I students differs, if at all, on standardized tests administered under timed and untimed conditions. This study will add to the research on closing the achievement gap.

I am contacting you to ask you to participate in my study because I need a select group of Title I schools to help me conduct my research. This would entail administering the 2006 Released FCAT Reading Test to your third grade students in the fall. The students would need to be randomly selected and put into two heterogeneous groups. One group should be given the test with timed conditions and the other with untimed conditions. The timed test will take two days with 60 minutes provided for each day of testing. The untimed test will also take two days with 120 minutes provided for each day of testing for those students who need that length of time. You may exclude any third grade student from the test that would normally require a one on one test administrator or whom you think would not be a good participant for this study. You may also use the results of the data to help guide curriculum decisions at your school if you choose.

If you decide to participate, you will be asked to meet with me for approximately 20 minutes, so I can review the process with you. I will come to your work location at a time that is convenient for you, or we can schedule a phone conference. You should take part in this study only because you want to. There is no penalty for not taking part, and you will not lose any benefits. You have the right to withdraw from the process at any time. There are no expected risks for taking part in this study.

This study has been approved by Orange County Public Schools' Research Department. If you have questions about the study, you can contact me at 321-278-4434 or by email at ruthie.haniff@ocps.net. In addition, you may contact Dr. Martha Lue-Stewart, my Faculty Supervisor, at the College of Education at 407-823-2036, or by email at Martha.Stewart@ucf.edu.

If you agree to allow your third grade students to participate in this research, please sign and return a copy of this letter to me. If you choose not to participate, please also let me know. You may scan this letter and send it back by email.

Thank you for consideration of my request.

Ruth Haniff

Yes, I agree to participate in the study.

No, I do not wish to participate in the study.

_____/_____
Signature **Date**

APPENDIX C
TEST ADMINISTRATOR SCRIPT FOR TIMED TESTS

Timed Test Script

Day One

1. Tell students they have exactly 60 minutes to work on the test today.
2. Tell them you will give them another 60 minutes to work on the test tomorrow.
3. Tell them they should not write in the booklet and you want them to bubble in their answers on the bubble sheet.
4. Write on the board to choose **J on the answer sheet** when they want to choose **I as an answer**.
3. Tell them to do their best.

Day Two

1. Tell students they have exactly 60 minutes to work on the test today.
2. Tell them they must start where they left off and are not allowed to go back to what they did yesterday.
3. Remind them not to write in the test booklet and to circle in their answers on the bubble sheet.
4. Write on the board to choose **J on the answer sheet** when they want to choose **I as an answer**.
4. Tell them to do their best.

APPENDIX D
TEST ADMINISTRATOR SCRIPT FOR UNTIMED TESTS

Untimed Test Script

Day One

1. Tell students you will give them as much time as they need to finish the test.
2. Tell them what they do not finish today that you will give them time to finish tomorrow.
3. Tell them they should not write in the booklet and you want them to bubble in their answers on the bubble sheet.
4. Write on the board that they should choose **J on the answer sheet** when they want to choose **I as an answer**.
3. Tell them to do their best.

Day Two

1. Tell students you will give them as much time as they need to finish the test today.
2. Tell them they must start where they left off and are not allowed to go back to what they did yesterday.
3. Remind them not to write in the test booklet and to circle in their answers on the bubble sheet.
4. Write on the board that they should choose **J on the answer sheet** when they want to choose **I as an answer**.
4. Tell them to do their best.

APPENDIX E
PARENT PERMISSION



Research Study Title

The Impact of Timed Versus Untimed Standardized Tests on Reading Scores of Third Grade Students in Title I Schools

]

Principal Investigator: Ruth E. Haniff, Ed.S.

Faculty Supervisor: Martha Lue-Stewart, Ph.D.

Investigational Site(s): Recruited Title I Schools That Agree To Participate
Orange County Public Schools

How to Return This Consent Form: Please return this form to your child's classroom teacher.

Introduction: Researchers at the University of Central Florida (UCF) study many topics. To do this we need the help of people who agree to take part in a research study. You are being asked to allow your child to take part in a research study which will include about 250 third grade students who attend a Title I School in the Orange County Public School System.

The person doing this research is Ruth Haniff who is a doctorate student at UCF in the Department of Education. Because the researcher is a graduate student, she is being guided by Dr. Martha Lue-Stewart, Ph.D. an UCF faculty supervisor in the Education Department.

What you should know about a research study:

Someone will explain this research study to you.

A research study is something you volunteer for.

Whether or not you take part is up to you.

You should allow your child to take part in this study only because you want to.

You can choose not to take part in the research study.

You can agree to take part now and later change your mind.

Whatever you decide it will not be held against you or your child.

Feel free to ask all the questions you want before you decide.

Purpose of the research study: The purpose of this study is to investigate the extent to which the performance of Title I third grade students differ on standardized tests

administered under timed and untimed conditions. The No Child Left Behind Act is mandating that all students need to be reading on grade level by 2014. The 2009 National Report Card shows that there continues to be significant academic achievement gaps between ethnic groups as well as between students who receive free or reduced lunch and those who do not. This study hopes to help with closing those achievement gaps by investigating if students would do better on the Florida Comprehensive Assessment Test (FCAT) and other standardized tests by having as much time as they need to answer the questions rather than only a limited amount of time. It will also add to the research and literature focused on closing the achievement gap.

What your child will be asked to do in the study: Your child will be asked to take the 2006 Released FCAT Reading Test for third grade students. Your child will either be placed in a timed group which would be for a 60 minute period each day for two days. Or, your child may be randomly selected to take the FCAT test in an untimed group which will allow your child as much time as they need to finish up to a maximum of 120 minutes each day for two days. The test will be administered between September 1, 2011 and October 28, 2011 by teachers and staff at your child's elementary school in a classroom. Children who participate in the study will not miss any work that needs to be made up. Children who do not take part in the study will read independently during the reading block time.

Location: The test will be administered at your child's elementary school in a normal classroom setting.

Time Required: We expect that your child will be in this research study for either a 60 minute period each day for two days in a timed setting or a 120 minute period maximum each day for two days in an untimed setting. The FCAT Reading Test normally takes place during the reading block of time during the school day which is 90 minutes. If additional time is required, the Principal at the school will determine what time during the day the rest of the test will be completed.

Risks:

There are no expected risks for taking part in this study.

Benefits:

We cannot promise any benefits to you, your child, or others from your child taking part in this research. However, possible benefits include giving your child additional practice on taking the FCAT. It may also provide your child's school with additional information on how to help your child to better prepare for the actual FCAT that they are required to take to pass third grade.

Compensation or payment: There is no compensation, payment or extra credit for your child's part in this study.

Confidentiality: Your child will be assigned a pseudo name for data analysis purposes to prevent your child from being identified in the research study results. No names will be used in the report.

Efforts will be made to limit your child's personal information to people who have a need to review this information such as your child's elementary school personnel. We cannot promise complete secrecy. Organizations that may inspect and copy your child's information include the IRB and other representatives of UCF.

Study contact for questions about the study or to report a problem: If you have questions, concerns, or complaints, or think the research has hurt your child talk to Ruth Haniff, Graduate Student, Principal Investigator, College of Education at 321-278-4434 or Dr. Martha Lue-Stewart, Ph.D., Faculty Supervisor, College of Education (407) 823-2036 or by email at Martha.Stewart@ucf.edu.

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

Your questions, concerns, or complaints are not being answered by the research team.

You cannot reach the research team.

You want to talk to someone besides the research team.

You want to get information or provide input about this research.

Withdrawing from the study:

You may decide not to have your child continue in the research study at any time without it being held against you or your child. If you decide to have your child leave the study, please contact the Principal of your child's elementary school and they will honor your request.

The Principal at your child's elementary school can remove your child from the research study without your approval. Possible reasons for removal include the student not following directions or disturbing the testing environment.

Your signature below indicates your permission for the child named below to take part in this research.

DO NOT SIGN THIS FORM AFTER THE IRB EXPIRATION DATE BELOW

Name of participant

Signature of parent or guardian

Date
Parent
Guardian
(See note
below)

Printed name of parent or guardian

Assent

Obtained

Note on permission by guardians: An individual may provide permission for a child only if that individual can provide a written document indicating that he or she is legally authorized to consent to the child's general medical care. Attach the documentation to the signed document.

APPENDIX F
STUDENT ASSENT SCRIPT

STUDENT ASSENT SCRIPT

***Note: The parent of the child must always give consent before you invite the child to participate in the study.**

Please read the following script to obtain assent from the students who wish to participate in the study once you have received parent permission.

1. Hi, (Child's Name).
2. My name is _____, and we are trying to learn more about how to increase students' scores on the FCAT.
3. We would like you to take a practice FCAT and see how well you do.
4. So you want to do this?
(If the child does not indicate affirmative agreement, you cannot continue with this child.)
5. Do you have any questions before we start? (Clarify answer if necessary.)
6. If you want to stop at any time just tell me.

APPENDIX G
STUDENT INFORMATION DATA SHEET

STUDENT CODE _____

Please place a checkmark next to the categories that pertain to this student.

_____ Timed _____ Untimed

_____ Male _____ Female

_____ ESE (Not Gifted) _____ ELL

_____ Reduced Lunch _____ Free Lunch

Race

_____ White _____ Black _____ Hispanic

_____ Asian _____ American Indian _____ Other

APPENDIX H
FLORIDA DEPARTMENT OF EDUCATION COPYRIGHT PERMISSION

FLORIDA DEPARTMENT OF EDUCATION



Dr. Eric J. Smith
Commissioner of Education

STATE BOARD OF EDUCATION

T. WILLARD FAIR, *Chairman*

Members

DR. AKSHAY DESAI

MARK KAPLAN

ROBERTO MARTÍNEZ

JOHN R. PADGET

KATHLEEN SHANAHAN

SUSAN STORY



November 29, 2010

Ruth E. Haniff
ruthie.haniff@ocps.net

Dear Ms. Haniff:

Your request for copyright permission to use the 2006 released Grade 3 FCAT Reading assessment posted at <http://fcats.fldoe.org/fcatrelease.asp> for the purposes of your dissertation has been reviewed. Permission is hereby granted solely for the reprinting of the Grade 3 FCAT Reading items (questions and answers) in your dissertation and for use of the full versions of the released tests for your research purposes.

The following copyright statement must accompany reprinted FCAT items:

FCAT content appears by permission of the Florida Department of Education, Office of Assessment, Tallahassee, Florida 32399-0400.

The copyrights for FCAT reading passages are secured for the Florida Department of Education by our FCAT contractor for specific purposes, i.e., print, Web posting, and for a specific period of time. You may use the reading passages during your research processes, but you do not have permission to reprint any of the reading passages in your dissertation. To obtain copyright permission for reprinting individual reading passages, please use the copyright information that follows each reading passage for contacting the copyright holder.

If you have any questions about this information, please let me know. I may be reached at Judy.Kent@fldoe.org or (850) 245-0759.

Sincerely,

Judy T. Kent
Product Information Consultant

SHARON KOON, PH.D.
ASSISTANT DEPUTY COMMISSIONER
ACCOUNTABILITY, RESEARCH, AND MEASUREMENT
OFFICE OF ASSESSMENT

325 W. GAINES STREET • SUITE 414 • TALLAHASSEE, FL 32399-0400 • (850) 245-0513 • www.fldoe.org

LIST OF REFERENCES

- Abrams, L., Pedulla, J., & Madaus, G. (2003). Views from the classroom: Teachers' opinions of statewide testing programs. *Theory into Practice, 42*, 18-29.
- Acrey, C., Johnstone, C., Miligan, C. (2005). Using universal design to unlock the potential for academic achievement of at-risk learners. *Teaching Exceptional Children, 38*(2), 22-31.
- American Institutes For Research. (2011). *Florida's Value-Added Technical Assistance Workshop*. Retrieved from:
<http://www.fldoe.org/committees/pdf/august12tampres>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Anderson, R., Wilson, P., & Fielding, G. (Summer, 1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly, 23*(3), 285-303.
- Armstrong, A. (2010). Myths of poverty. *The Education Digest, 75*(8), 49-53.
- Artiles, A., Kozleski, E., Trent, S., Osher, D., & Oriz, A., (2010). Justifying and explaining disproportionality, (1968-2008): A critique of underlying views of culture. *Exceptional Children, 76*(3), 279-299

- Aud, S., Fox, M., & KewalRamani, A. (2010). *Status and trends in the education of racial and ethnic groups* (NCES 2010-15). U.S. Department of Education, National Center for Education Statistics. Washington, DC. Retrieved from <http://nces.ed.gov/pubs2010/2010015.pdf>
- Ausubel, D. (1968). *Educational psychology: A cognitive view*. New York: Holt, Rinehart, & Winston.
- Baptiste, H., Orvosh-Kamenski, H., & Kamenski, C. (2004). American presidents and their attitudes, beliefs, and actions surrounding education and multiculturalism: A series of research studies in education policy. Fifth installment--Examining presidents Thomas Jefferson, Abraham Lincoln, and Lyndon B. Johnson. *Multicultural Education*, 12(2), 34-47.
- Barksdale-Ladd, M., & Thomas, K. (2000). What's at stake in high-stakes testing: Teachers and parents speak out. *Journal of Teacher Education*, 51, 384-397.
- Bergeron, B. (2008, January). Enacting a culturally responsive curriculum in a novice teacher's Classroom. Encountering disequilibrium. *Urban Education*, 43(1), 4-28.
- Berliner, D., & Biddle, B. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools*, Reading, MA: Addison-Wesley.
- Black, P. (1998). *Testing: Friend or foe? The theory and practice of assessment and testing*. London: The Falmer Press.
- Black, S. (2004). Second time around. *American School Board Journal*, 191(11), 40-42.

- Bolt, S., & Thurlow, M. (2004). Five of the most frequently allowed testing accommodations in state policy: Synthesis of research, *Remedial and Special Education*, 25(3), 141-52.
- Borman, G. (2000). Title I: The evolving research base. *Journal of Education for Students Placed at Risk*, 5(1, 2), 27-45.
- Brooks, T., Case, B., & Young, M. (2003). *Timed versus untimed testing conditions and student performance*. San Antonio, TX: Harcourt Educational Measurement. Pearson Education.
- Brutch, M., Juster, H., & Kaflowitz, N. (1983). Relationships of cognitive components of test anxiety to test performance: Implications for assessment and treatment. *Journal of Counseling Psychology*, 30, 527-536.
- Buddy, J. (2009). Standardized testing review 101. *School Library Monthly*, 26(4), 18-21.
- Case, B. (2003). *It's about time: Stanford achievement tests series, tenth edition Stanford 10*. San Antonio, TX: Harcourt Educational Measurement. Pearson Education.
- Cawthon, S., Ho, E., Patel, P., Potvin, D., & Trundt, K. (2009). Multiple constructs and the effects of accommodations on standardized test scores for students with disabilities. *Practical Assessment, Research & Evaluation*, 14(18). Retrieved from <http://pareonline.net/getvn.asp?v=14&n=18>.
- Chapman, P. (1988). *Schools as sorters*. New York: New York University Press.
- Cicchelli, T., & Cho, S. (2007). Teacher multicultural attitudes: Intern/teaching fellows in New York City. *Education and Urban Society*, 39(3), 370-381.

- Cizek, G., & Berg, W. (2006). *Addressing text anxiety in a high-stakes environment: Strategies for classrooms and schools*. Thousand Oaks, CA: Corwin Press.
- Clotfelter, C., Ladd, H., & Vidgor, J. (2006) The achievement gap in grades 3 to 8. NBER Working Papers 12207, National Bureau of Economic Research.
- Dallas Indicators. (2010). *Economically disadvantaged students*. Retrieved from www.dallasindicators.org/Education/Equityin Education.
- Davis, R., Ramahlo, T., Beyersbach, B., & London, A. (2008). A culturally relevant teaching course: Reflecting preservice teachers' thinking. *Teaching Education*, 19(3), 223-234.
- Debra P. v. Turlington (1983). Retrieved from <http://www.fldoe.org/asp/hsap/hsap1983.asp>
- Delpit, L. (2006). *Other people's children: Cultural conflict in the classroom*. New York: The New Press.
- Driscoll, M. (2005). *Psychology of learning for instruction* (3rd ed.). Boston: Pearson Education.
- Duffy, M., Giordano, V., Farrell, J., Paneque, O., & Crump, G. (2008). No child left behind: Value and research issues in high-stakes assessments. *Counseling and Values*, 53, 53-66.
- Evans, G., & Schamberg, M. (2009). Childhood poverty, chronic stress, and adult working memory. *Proceedings of the National Academy of Sciences*, 106(16), 6545-6549.

- Feller, B., & Hefling, K. (2012, February 9). Official: 10 states given waiver on no child left behind learning laws. *MSNBC.COM*. Retrieved from <http://www.msnbc.msn.com/id/46323704/ns/polit>
- Ferguson, R. (2006, January). *Closing the achievement gap*. Paper presented at the meeting of the Vanderbilt University Learning Sciences Institute, Nashville, TN.
- Fine, M., & Weis, D. (1999). *The unknown city: The lives of poor and working class young adults*. Boston, MA: Beacon Press.
- Fletcher, D. (2009, December). A brief history of standardized testing, *Time Magazine*. Retrieved from <http://www.time.com/time/nation/article>
- Florida Department of Education (1990). *The META Consent Decree*. Retrieved from <http://www.fldoe.org/aala/cdpage2.asp>
- Florida Department of Education. (2001). *FCAT briefing book*. Retrieved from http://hsee.umn.edu/hsee_documentation/fl01_flde.pdf
- Florida Department of Education (2007). *Assessment and accountability briefing book*. Retrieved from fcata.fldoe.org/pdf/BriefingBook07web.pdf.
- Florida Department of Education. (2010a). *Bureau of curriculum and instruction: Next generation sunshine state standards*. Retrieved from <http://www.fldoe.org/bii/curriculum/sss/#read>.
- Florida Department of Education (July, 2010b). *Florida's adoption of the common core state standards: State board of education*. Retrieved from http://www.fldoe.org/board/meetings/2010_06/corepres.pdf

- Florida Department of Education. (July, 2010c). *2010 guide to calculating adequate yearly progress (AYP) technical assistance paper*. Retrieved from <http://schoolgrades.fldoe.org/pdf/0910/AYPAP2010>
- Florida Department of Education. (2010d). *No child left behind*. Retrieved from <https://www.fldoe.org/nclb/>
- Florida Department of Education. (2011a). *Florida's value-added technical assistance workshop*. Orlando, Florida. Retrieved from <http://www.fldoe.org/committees/sq.asp>
- Florida Department of Education. (2011b). *Student performance results: State reading demographic report*. Retrieved from <https://app1.fldoe.org/FCATDemographics/Selections>
- Florida Department of Education (2012). *Understanding FCAT 2.0 Reports*. Retrieved from <http://fcats.fldoe.org/fcat2/pdf/s12uf2r.pdf>
- Florida Senate (2011). *CS/SB 1656: McKay scholarships/students with disabilities*. Retrieved from <http://www.flSenate.gov/Session/Bill/2011/1656>
- Ford, D. (2006, Fall). Closing the achievement gap: How gifted education can help. *Gifted Child Today*, 29(4), 14-17.
- Ford, D., Grantham, T., & Whiting, G. (2008, March). Another look at the achievement gap: Learning from the experiences of gifted black students. *Urban Education*, 43(2), 216- 239.
- Freire, P. (1998). *Teachers as cultural workers: Letters to those who dare teach*. Boulder, CO: Westview Press.

- Futrell, M. (2004, Summer). Anticipating success; removing barriers to educational equity and equality. *Harvard Journal of African American Public Policy*, 10, 103-116.
- Gallagher, C. (2003). Reconciling a tradition of testing with a new learning paradigm, *Education Psychology Review*, 15(1), 83-99.
- Gay, G. (2000). *Culturally responsive teaching: Theory, research, & practice*. New York: Teachers College Press.
- Gay, G. (2004, Spring). Beyond Brown: Promoting equality through multicultural education, *Journal of Curriculum & Supervision*, 19(3), 193-216.
- Goleman, D. (1995). *Emotional Intelligence: Why It Can Matter More Than IQ for Character, Health and Lifelong Achievement*. Bantam Books, New York.
- Gorey, K. (2009). Comprehensive school reform: Meta-Analytic evidence of black-white achievement gap narrowing, *Education Policy Analysis Archives*, 17(25), 1-14.
- Gould, S. (1981). *The mismeasure of man*. New York: Norton.
- Gould, S. (1982, May). A nation of morons. *New Scientist*, 6, (349-352).
- Griner, A., & Lue-Stewart, M. (November, 2011). *Addressing the achievement gap and disproportionality through the use of culturally responsive teaching practices*. Paper presented at the 56th annual meeting of the Florida Educational Research Association, Orlando, FL.
- Haberman, M. (1991). The pedagogy of poverty versus good teaching. *Phi Delta Kappan*, 290- 294.

- Hammond, C., Linton, D., Smink, J., & Drew, S. (2007). *Dropout risk factors and exemplary programs*. Clemson, SC: National Dropout Prevention Center, Communities In Schools.
- Hanson, F. (1993). *Testing: Social consequences of the examined life*. Berkeley, CA: University of California Press.
- Heath, S. (2004). Ethnography in communities learning the everyday life of America's subordinated youth. In J. Banks, & C. A. McGee Banks (Eds.), *Handbook of research on multicultural education* (2nd ed.) (pp 146-162). San Francisco: Jossey-Bass.
- Herrnstein, R., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.
- Heubert, J., & Hauser, R. (Eds.) (1999). *High stakes; Testing for tracking, promotion, and graduation*. Washington, DC: National Academic Press.
- Hirsch, E. (1996). *The schools we need and why we don't have them*. New York: Doubleday.
- Holmes, T. (2006). Low Test Scores + High Retention Rates = More Dropouts. *Kappa Delta PI Record*, 42(2), 56-58.
- Houston, P. (2005). NCLB: Dreams and nightmares. *Phil Delta Kappan*, 86(6), 469-470.
- Hursh, D. (2007). Exacerbating inequality: The failed promise of the no child left behind act. *Race, Ethnicity and Education*, 10(3), 295-308.
- Hyland, N. (2005). Being a good teacher of black students? White teachers and unintentional racism. *Curriculum Inquiry*. 35(4), 429-459.

- Jackson, R. (2009, March 10). Obama urges education reform. *USA Today*. Retrieved from <http://www.usatoday.com>
- Jensen, E. (2005). *Teaching with the brain in mind* (2nd ed.). Alexandria, VA: ASCD.
- Kahn, M. (2008). Multicultural education in the United States: Reflections. *Intercultural Education, 19*(6), 527-536.
- Kincheloe, J., & Steinberg, S. (2004). *19 Urban questions: Teaching in the city*. New York: Peter Lang.
- Kincheloe, J., Steinberg, S., & Gresson, A. (1996). *Measured lies: The bell curve examined*. New York: St. Martins.
- Kirkland, K., & Hollandsworth, J. G. (1979). Test anxiety, study skills, and academic performance. *Journal of College Personnel, 20*, 431-435.
- Kishiyama, M., Boyce, W., Jimenez, A., Perry, L., & Knight, R. (2008). Socioeconomic disparities affect prefrontal function in children, *Journal of Cognitive Neuroscience, 21*(6), 1106-1115.
- Klingner, J., Artiles, A., & Barletta, L. (2006). English language learners who struggle with reading: language acquisition or LD? *Journal of Language Disabilities, 39*(2), 108-128.
- Kozol, J. (1991). *Savages inequalities*. New York: Harper Collins.
- Kozol, J. (2005). *The shame of a nation: The restoration of apartheid schooling in America*. New York: Random House.
- Kozol, J. (2007). *Letters to a young teacher*. New York: Crown.

- Ladson-Billings, G. (1994). *The dream-keepers: Successful teachers of African-American Children*. . San Francisco: Jossey-Bass.
- Ladson-Billings, G. (2001). *Crossing over to Canaan: The journey of new teachers in diverse classrooms*. San Francisco: Jossey-Bass.
- Lee, J. (1995, April). *Information processing model of text anxiety and its effect on the speed-accuracy trade-off (ACT or the spreading activation explanation)*. Paper presented at the Annual Meeting of the American Education Research Association, San Francisco, CA.
- Linden, K., and Linden, J. (1968). *Modern mental measurement: A historical perspective*, Boston, MA: Houghton Mifflin.
- Machtinger, H. (2007). What do we know about high poverty schools. Summary of the high poverty schools conference at UNC-Chapel Hill. *High School Journal*, 90(3), 1-8.
- Marquart, A. (2000). *The use of extended time as an accommodation on a standardized mathematics test: An investigation of effects on scores and perceived consequences for students of various skill levels*.(Doctoral dissertation). Retrieved from US: ProQuest Information & Learning.
- McNeil, M. (2011). Waiver plans would scrap parts of NCLB. *Education Week*, 31(13), 1-28.
- McNeil, M., Cavanagh, S., & Klein, A. (2012). Waiver hopefuls put through paces by review process. *Education Week*, 31(21), 1-27.
- Meier, D., & Wood, G. (Eds.). (2004). *Many children left behind*. Boston: Beacon Press.

- Metz, M. (1987). Teachers' pride in craft, school subcultures, and societal pressures. *Educational Policy*, 1(1), 115-134.
- Miller, R., Murnane, R., & Willett, J. (2008). Do teacher absences impact student achievement? Longitudinal evidence from one urban school district. *Educational and Analysis*, 30(2), 181-200.
- Minton, H. (1984). The Iowa child welfare research station and the 1940 debate on intelligence: Carry on the legacy of a concerned mother. *Journal of the History of the Behavioral Sciences*, 20, 160-176.
- Minton, H. (1988). *Lewis M. Terman: Pioneer in psychological testing*. New York: University Press.
- National Center for Education Statistics (2011a). *National assessment of educational progress (2011)*. Retrieved from <http://nces.ed.gov/nationsreportcard/studies/gaps/>
- National Center for Education Statistics (2011b). *Indicator 17. Retention, suspension, and expulsion rates*. Retrieved from http://nces.ed.gov/pubs2010/2010015/indicator4_17.asp
- National Center for Education Statistics (2011c). *Indicator 18. Dropout rates and graduation rates*. Retrieved from http://nces.ed.gov/pubs2010/2010015/indicator4_18.asp
- National Report Card (2009). Retrieved from <http://nationsreportcard.gov>
- National Report Card (2011). Retrieved from <http://nationsreportcard.gov>

- Nieto, S. (2004). *Affirming diversity: The sociopolitical context of multicultural education* (4th ed.). Pearson Education.
- Nieto, S., & Bode, B. (2008). *Affirming diversity: The sociopolitical context of multicultural education* (5th ed.). Boston: Pearson.
- No Child Left Behind (2002). Retrieved from <http://www.fldoe.org/nclb/>
- Noddings, N. (2005). What does it mean to educate the whole child? *Educational Leadership*, 63(1), 8-13.
- Noguera, P., & Wing, J. (Eds.). (2006). *Unfinished business: Closing the racial achievement gap in our schools*. Hoboken, NJ: Jossey-Bass.
- Orange County Public Schools (2012a). *McKay scholarships*. Retrieved from <https://www.ocps.net/cs/ese/Procedures/Pages/McKay.aspx>
- Orange County Public Schools (2012b). *Educational data warehouse (EDW)* [http://www.ocps.net/intranet/cc/css/ara/Pages/EDW%20\(Educational%20Data%20 Warehouse\).aspx](http://www.ocps.net/intranet/cc/css/ara/Pages/EDW%20(Educational%20Data%20Warehouse).aspx)
- Orfield, G., & Lee, O. (2004). *Brown at 50: Kings' dream Plessy's nightmare: The civil rights project*. Cambridge, MA: Harvard University.
- Onwuegbuzie, A., & Seaman, M. (1995). The effect of time constraints and statistics test anxiety on test performance in a statistics course, *Journal of Experimental Education*, 63(2), 115-124.
- Paulson, S., & Marchant, J. (2009). Background variables, levels of aggregation, and standardized test scores. *Education Policy Analysis Archives*, 17(23), 1-21.
- Payne, R. (2008, April). Nine powerful practices. *Educational Leadership*, 65(7), 48-52.

- Poplin, M., & Soto-Hinma, I. (2005). Taking off ideological blinders: Lessons from the start of a study on effective teachers in high-poverty schools. *Journal of Education, 186*(3), 41-44.
- Pransky, E., & Bailey, F. (2002/2003). To meet your students where they are, first you have to find them: Working with culturally and linguistically diverse at-risk students, *Reading Teacher, 56*(4), 370-383.
- Radville, K. (2008). *Relationships among cognitive-linguistic skills, fluency interventions and reading comprehension in third grade inner city children*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 1457946).
- Raymond, E. (2004). *Learning with mild disabilities: A characteristic approach* (2nd ed.). Needham Heights, MA: Allyn & Bacon.
- Reber, S. (2005). Court-order desegregation: Success and failures integrating schools since Brown v. The Board of Education. *The Journal of Human Resources, 40*(30) 559-590.
- Roach, R. (2005). Closing the achievement gap: Who's going to take the weight? *Black Issues In Higher Education, 22*(7), 32-39.
- Robelen, E. (2005). 40 years after ESEA, federal role in schools is broader than ever. *Education Week, 24*(31), 1, 42.
- Rothenberg, P. (2001). *Race, class, and gender in the United States*, New York: Worth.
- Rothman, R. (1995). *Tests of significance*. San Francisco: Jossey-Bass.

- Rothstein, R., & Jacobsen, R. (2006). The goals of education. *Phi Delta Kappan*, 88(4), 264-272.
- Ryan, W. (1972). *Blaming the victim*. New York: Vintage Books.
- Sacks, P. (1999). *Standardized minds*. Cambridge, MA: Perseus Books.
- Schlemer, L. (2007). *Test speededness and cognitive styles: A study of using a CAT version of the SAT*. Retrieved from US: ProQuest Information & Learning.
- Schultz, B. (2007). Living "savages inequalities." *Democracy & Education*, 17(1), 42-54.
- Segool, N. (2009). *Test anxiety associated with high-stakes testing among elementary school children: Prevalence, predictors, and relationship to student performance*. (Doctoral dissertation). Retrieved from US: ProQuest Information & Learning.
- Sherman, W. (2008). No child left behind: A legislative catalyst for superintendent action to eliminate test-score gap? *Educational Policy*, 22(5), 674-704.
- Stretch, L., & Osborne, J. (2005). Extended time test accommodation: Directions for future research and practice. *Practical Assessment, Research & Evaluation*, 10(8), 1-8.
- Thompson, S., Johnstone, C., & Thurlow, M. (2002). *Universal design of large scale assessments* (Synthesis Report No. 44). Minneapolis, MN: National Center on Educational Outcomes. Retrieved from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html>
- Thorndike, R., & Lohman, D. (1990). *A century of ability testing*. Chicago, IL: Riverside.

- Thurlow, M., & Bolt, S., (2001). *Empirical support for accommodations most often allowed in state policy* (Synthesis Report 41). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Tindal, G., & Fuchs, L. (2000). *A summary of research on test changes: An empirical basis for defining accommodations* (Rev. ed.) Lexington, KY: University of Kentucky.
- Tobias, S. (1985). Text Anxiety: Interference, defective skills, and cognitive capacity. *Educational Psychologist*, 20(3), 135-142.
- Tsui, J., & Mazzocco, M. (2007). Effects of math anxiety and perfectionism on timed versus untimed math testing in mathematically gifted sixth graders. *Roeper Review*. 29(2), 132-139.
- U.S. Department of Agriculture (2010). *Income eligibility guidelines*. Retrieved from <http://www.fns.usda.gov/cnd/Governance/notices/iegs/IEGs2010>
- U.S. Department of Education. (2002). Individuals with disabilities education act amendments of 1977. Retrieved from <https://www2.ed.gov/offices/OSERS/Policy/IDEA/index.html>
- U.S. Department of Education. (2010). Title I of the Elementary and Secondary Education Act of 1965. Retrieved October 10, 2010 from <http://www2.ed.gov/olicy/elsec/leg/esea02/pg1.html>
- U.S. Department of Education. (2012). *Race to the top Florida report year 1: School Year 2010-2011*. Retrieved from <http://www2.ed.gov/programs/racetothetop/performance/florida-year-1.pdf>.

- U.S. Department of Health and Human Services. (2006). *Your rights under section 504 of the rehabilitation act*. Retrieved October 10, 2010 from <http://www.hhs.gov/ocf/civilrights/resources/factsheets/504>
- U. S. Department of Justice. (2010). Americans with disabilities act of 1990, as amended. Retrieved October 10, 2010 from <http://www.ada.gov/pubs/ada.htm>
- Walsh, W., & Betz, N. (1995). *Test and assessment*. Englewood Cliffs, NJ: Princeton-Hall.
- Ward, A., & Murray-Ward, M. (1999). *Assessment in the classroom*. Albany, NY: Wadsworth.
- Weiner, L. (1993). *Preparing teachers for urban schools. Lessons from thirty years of school reform*. New York: Teachers College Press.
- Weiner, L. (2000). Research in the 90s: Implications for urban teacher preparation. *Review of Educational Research*, 70(3), 369-406.
- Weiner, L. (2003). Why is classroom management so vexing to urban teachers? *Theory Into Practice*, 42(4), 305-312.
- Weiner, L. (2006). Challenging deficit thinking, *Educational Leadership*, 64(1), 42-46.
- Willner, L., Rivera, C., & Acosta, B. (2009). Ensuring accommodations used in content assessments are responsive to English language learners, *The Reading Teacher*, 62(8), 696-698.
- Wine, J. (1971). Test anxiety and direction of attention. *Psychological Bulletin*, 6, 92-104.

Winfield, L., & Manning, J. (1992). Changing school culture to accommodate student diversity. In M.E. Dilworth (Ed.), *Diversity in teacher education* (pp. 181-214). San Francisco: Jossey-Bass.

Zeidner, M. (1998). *Test anxiety: The state of the art*. New York: Plenum Press.