

2007

# Third grade retention and Florida's pupil progression plan: Individual and school characteristics associated with long-term outcomes in reading performance

Heather A. Powell  
*University of South Florida*

Follow this and additional works at: <http://scholarcommons.usf.edu/etd>

 Part of the [American Studies Commons](#)

---

## Scholar Commons Citation

Powell, Heather A., "Third grade retention and Florida's pupil progression plan: Individual and school characteristics associated with long-term outcomes in reading performance" (2007). *Graduate Theses and Dissertations*.  
<http://scholarcommons.usf.edu/etd/2326>

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact [scholarcommons@usf.edu](mailto:scholarcommons@usf.edu).

Third Grade Retention and Florida's Pupil Progression Plan:  
Individual and School Characteristics Associated with Long-Term Outcomes  
in Reading Performance

by

Heather A. Powell

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
Department of Psychological and Social Foundations  
College of Education  
University of South Florida

Major Professor: Michael J. Curtis, Ph.D.  
George Batsche, Ed.D.  
Kathy L. Bradley-Klug, Ph.D.  
Constance Hines, Ph.D.  
Richard Marshall, Ph.D.

Date of Approval:  
July 11, 2007

Keywords: assessment, achievement, remediation, academic, policy

© Copyright 2007 , Heather Powell

To my father, Jon Lillemoe,  
who taught me that learning is a lifelong endeavor

### Acknowledgements

The author would like to express her gratitude to individuals and organizations that were critical to the completion of this project. I extend my sincere appreciation to the Florida Department of Education for the privilege of accessing the statewide student database, to each of my committee members for their keen eyes and astute comments, and to Bethany Bell-Edison for her statistical expertise. Special thanks are extended to Mike Curtis, for his unwavering support and constant encouragement through my graduate program. His academic expertise and holistic approach to mentoring have been invaluable at each step, and I hope to emulate his dedication to students and colleagues in my own career as a school psychologist. Finally, I would like to thank my husband, Michael, for his patience, flexibility, adventurous spirit, and constant belief in my dreams.

## Table of Contents

List of Tables	iii
List of Figures	iv
Abstract	v
Chapter I: Introduction	1
Purpose of the Study	9
Research Questions	10
Significance of Study	11
Definition of Terms	12
Chapter II: Literature Review	15
Enactment of Federal and State Legislature	15
Recent Trends in Grade Retention	21
Characteristics of Retained Children	22
Evidence of Positive Outcomes of Retention	23
Evidence of Negative Outcomes of Retention	29
Factors that Influence Retention Outcomes	36
Effects of Retention Policy	38
Methodological Issues to Consider	45
Future Directions in Retention Research	46
Chapter III: Method	49
Purpose of the Study	49
Sample	50
Measures	51
Procedures and Data Preparation	53
Data Analysis	54
Research Questions	54
Research Question 1	54
Research Question 2	54
Research Question 3	54
Research Question 4	55
Research Question 5	55
Chapter IV: Results	64

Description of Sample	64
Research Questions	66
Research Question 1	66
Research Question 2	67
Research Question 3	74
Research Question 4	74
Research Question 5	74
Model 1	78
Model 2	82
Model 3	82
Model 4	83
Model 5	84
Chapter V: Discussion	87
Summary of Findings	88
Limitations	95
Delimitations	99
Implications for Practice and Future Research	99
Conclusion	103
References	106
Appendices	114
Appendix A	115
About the Author	End page

## List of Tables

Table 1	Dummy Coding for Student- and School- variables	57
Table 2	Students Scoring Level 1 on the 2003 FCAT-Reading and Retention/Promotion Status by Gender, Race/Ethnicity, and SES	66
Table 3	Performance of Retained Students on the 5 <sup>th</sup> grade FCAT-Reading in 2006 by Level	67
Table 4	Number and Percent of Students Promoted to 4 <sup>th</sup> Grade due to Good Cause Exemption by Type of Exemption	68
Table 5	Performance of Students Promoted through Good Cause on the 5 <sup>th</sup> grade FCAT-Reading in 2005 by Level	69
Table 6	Level 1 Promoted Students' Performance on the 2005 5 <sup>th</sup> grade FCAT-Reading by Good Cause Exemption	70
Table 7	Performance of Students Promoted through Good Cause on the 6 <sup>th</sup> grade FCAT-Reading in 2006 by Level	72
Table 8	Level 1 Promoted Students' Performance on the 2006 6 <sup>th</sup> grade FCAT-Reading by Good Cause Exemption	73
Table 9	Means and Standard Deviations for Student- and School-Level Variables	76
Table 10	Pearson Product Moment Correlations for Student- and School-Level Variables	78
Table 11	HLM Means as Outcomes Models: Selected School Characteristics as Predictors of Variation Between Schools in Performance on 5 <sup>th</sup> Grade FCAT-Reading	80

## List of Figures

Figure 1	Retention Status X Gender Effects on 5 <sup>th</sup> grade FCAT-Reading Score	85
----------	---	----

THIRD GRADE RETENTION AND FLORIDA'S PUPIL PROGRESSION PLAN:  
INDIVIDUAL AND SCHOOL CHARACTERISTICS ASSOCIATED WITH  
LONG-TERM OUTCOMES IN READING PERFORMANCE

Heather A. Powell

ABSTRACT

Literacy is a growing national concern that has resulted in federal legislation (e.g., the No Child Left Behind Act) instituting higher accountability for states and schools with regard to reading instruction and remediation. As a result, Florida's statewide measure of achievement, the reading portion of the Florida Comprehensive Assessment Test (FCAT-Reading) is now tied to retention decisions for students in the third grade as part of the pupil progression plan for the state. In its first year of implementation (2003), 23% of third-grade students failed the FCAT and over 28,000 were retained. Though failure rates are decreasing, (i.e., 6% of third grade students failed in 2006), tremendous numbers of students continue to be affected by this policy.

Research has consistently shown retention to be a negative experience for children; even when academic gains are made, their subsequent achievement is equal to or lower than that of both same-grade and same-age regularly promoted students within two to three years. However, these findings cannot be generalized to the current student progression plan in Florida, which mandates remediation activities and diligent progress monitoring during and after the retention year. Therefore, holding negative beliefs about the third grade retention policy in Florida is premature as only preliminary research exists

to date evaluating the outcomes of the plan.

The present study examined the student progression plan in Florida as it relates to performance on the FCAT-Reading and mandated third-grade retention. More specifically, this study examined the relationship between reading performance outcomes and various student characteristics (e.g., retention status, gender, SES, race/ethnicity) as well as school-related variables (e.g., school-wide SES status, school size, Reading First status). This study also explored fifth and sixth grade performance on the FCAT-Reading of low-performing students who were promoted through good cause exemptions.

Descriptive analyses revealed that of 12,685 third-grade students retained in 2003, 40% scored at Level 1 in 2006. With regard to students who were promoted due to a good cause exemption, findings indicated that a higher proportion of those who demonstrated reading proficiency through an alternative assessment procedure (67%) or through student portfolios (58%) achieved success in 2006 compared to those who did not demonstrate proficiency (13%-19%). In addition, retention status was significantly associated with scores on the 5<sup>th</sup> grade FCAT-Reading, but that association varied by student gender.

## CHAPTER ONE

### Introduction

Literacy is a growing national concern, justified by statistics reporting that only 32% of our nation's fourth grade children read proficiently (United States Department of Education, 2002). In addition, there is a growing achievement gap, such that the performance of top achievers has increased over time while that of the poorest achievers has declined (United States Department of Education, 2002). These developments have been alarming, provoking new federal legislation (e.g., No Child Left Behind Act) to address these issues, the effects of which have proven to be far-reaching (United States Department of Education, 2003a).

Since the No Child Left Behind Act was enacted on January 8, 2002, initiatives have been undertaken across the country to restructure many aspects of the American educational system (NCLB, 2002). The law was designed to address the country's reading crisis; its overriding goal is for every child to be reading at a proficient level by the 2013-2014 school year (United States Department of Education, 2003b). To this end, the law includes requirements regarding school and state accountability, including procedures to measure student progress. The law mandates the demonstration of Adequate Yearly Progress (AYP) in student performance by every state; thus, each state is required to set annual goals for its schools and propose a way to systematically measure progress toward attainment of those goals (United States Department of

Education, 2003c). In order to address the widening achievement gap, states are no longer allowed to aggregate achievement data across economic background, race/ethnicity, English proficiency or disability status; schools now must disaggregate performance data and demonstrate that all students are learning and that those who are disadvantaged or at-risk are not being “left behind”.

Although the Act has been criticized in some quarters for the failure to financially support its mandates, funds have been allocated to support some specific programs. For example, funds are available to assist states in the implementation of “empirically-based reading programs”, though research has shown that the specific components of such programs need to be delineated (e.g., classroom management, skills instruction, literature emphasis, scaffolding; Pressley et al., 2001). States apply for the funds through the “Reading First” program and funding is granted based on the number of children aged 5-17 years who are considered low-income in each state (United States Department of Education, 2003d).

The acquisition of Reading First funds is a large part of Florida’s response to the No Child Left Behind Act. Florida has established goals according to which 31% of its students are to be reading proficiently by the 2003-04 school year, 48% by 2006-07, 65% by 2009-10, 82% by 2012-13, and 100% by 2013-14 (Florida Department of Education, 2003a). With Reading First funds (\$52 million annually, for six years), schools around the state will be instructing children using research-based reading programs (Florida Department of Education, 2003b).

In addition to preventative programs such as Reading First, the State of Florida

also enacted a new retention policy, effective January 7, 2003 (The Florida Senate, 2003). Designed to align Florida policies with federal law, this new policy places heavy emphasis on remediation of reading deficiencies and is strongly tied to the standardized test used to quantify student performance, the Florida Comprehensive Assessment Test (FCAT). The reading portion of the FCAT is administered every year to students in grades 3-10 and purports to measure students' progress towards attainment of predefined academic standards in reading (i.e., Sunshine State Standards) according to five levels, defined using the following scaled scores: Level 1:100-258, Level 2: 259-283, Level 3: 284-331, Level 4: 332-393, and Level 5: 394-500 (Florida Department of Education, 2003c). Students who score at a Level 1 are assumed to be having "little success with the challenging content of the Sunshine State Standards." At Level 2, students are having "limited success" with the same content. With regard to reading, it is imperative that students master the basic skills at the lower grade levels (grades K-3), as in fourth grade and beyond, students are expected to utilize reading skills in all academic areas. Without basic reading skills, the chances of student success are severely limited, since the approach to teaching reading shifts from basic skills to more advanced applications of those skills at this critical juncture (Just Read, Florida, 2003). The state contends that third grade students who are achieving at Level 1 have not mastered the most basic reading skills, indicating that they are not ready to move on to the more challenging material of the fourth grade; consequently, students who score at Level 1 on the FCAT - Reading and are not eligible for promotion according to one of six predefined "good cause exemptions" are to be retained in third grade (The Florida Senate, 2003). Though

perceived by some as pejorative, the retention policy in this context is viewed as an effort towards the remediation of current academic difficulties and the prevention of additional difficulties in the students' future academic careers.

Despite the policy's emphasis on remediation through retention, the State of Florida recognizes that grade retention may not be the most appropriate decision for each student scoring at a Level 1 on the FCAT-Reading. For this reason, six guidelines were developed, indicating those circumstances under which a student scoring at Level 1 on the FCAT-Reading can still be promoted to the fourth grade. These "good cause exemptions" include "[a]limited English proficient (LEP) students with less than two years in an English for Speakers of Other Languages (ESOL) program, [b]students with disabilities whose individual educational plan (IEP) indicated that participation in the FCAT was not appropriate, [c]demonstration of an acceptable level of performance on an alternate assessment...or scoring at the 51st percentile or higher on the norm-referenced test portion of the FCAT, [d]demonstration of proficiency in Sunshine State Standards through a student portfolio, [e]students with disabilities who participate in the FCAT...(but)...still demonstrate a deficiency in reading after more than two years of intensive remediation and were previously retained in kindergarten, first, second, or third grade, and [f]students who still demonstrate a deficiency in reading after two or more years of intensive remediation and were previously retained in kindergarten, first, second, or third grade for a total of two years" (Florida Department of Education, 2003e, p. 1). Notably, of these good cause exemptions, only two (i.e., alternate assessment and student portfolio) indicate that a student has displayed proficiency in reading skills at a third-

grade level.

The 2003 FCAT was administered several months after the new retention policy went into effect, resulting in a total of 28,028 (14.6%) third grade students being retained in grade for the 2003-2004 school year (Florida Department of Education, 2003d). In 2004, the numbers were similar, with 23,283 (11.0%) third grade students being retained in grade (Florida Department of Education, 2004a). The tremendous numbers of students who continue to be affected by the policy combined with longstanding belief that grade retention is detrimental to students' academic progress call for a careful re-examination of the literature regarding grade retention, as well as empirical examination of the impact of this new policy.

The practice of grade retention is not a new phenomenon; in fact, retention literature spans several decades. Despite prevailing negative beliefs, some studies (e.g., Alexander, Entwisle, & Dauber, 1994; Gottfredson, Fink, & Graham, 1994; Mantzicopoulos, 1997) have reported positive effects of retention, to varying degrees. In some cases, small academic gains may be seen (Mantzicopoulos, 1997), particularly when comparing retainees with younger, same-grade peers (Alexander et al., 1994). However, a larger number of studies have found that benefits are short-term and that within two to three years, the achievement of retainees is equal to or lower than that of both same-grade and same-age peers (Jimerson, 1999; Jimerson, Carlson, Rotert, Egeland, & Sroufe, 1997; Mantzicopoulos, 1997, Mantzicopoulos & Morrison, 1992; McCoy & Reynolds, 1999). In addition, it is interesting to note that increases in performance are most often found only in mathematics; reading remains a consistent

deficit in retained children (Jimerson et al., 1997; Mantzicopoulos, 1997).

Despite a few studies in which positive results have been reported, the overwhelming majority of research indicates that grade retention does not have a remedial effect; rather, it operates to produce the direct opposite of desired effects. This is particularly true for several groups of students who can be considered “at-risk” for being retained or experiencing negative outcomes subsequent to the retention year. These risk factors include being male, a member of an ethnic minority group, or living in poverty (e.g., McCoy & Reynolds, 1999; Reynolds, 1992). Reported negative outcomes have included lower reading achievement as measured by standardized tests (Jimerson, 1999; McCoy & Reynolds, 1999), increased social and behavior problems (Pagani, Tremblay, Vitaro, Boulerice, & McDuff, 2001), and increased dropout rates (Jimerson, 1999). These types of results were consistent across many studies using a variety of designs and analyses.

In addition to examining student variables as possible mediators of retention outcomes, it is important to consider ecological factors as well. Previous research has shown that school characteristics significantly influence general student achievement (Ma & Willms, 2004), suggesting that these same characteristics may have the potential to impact achievement following a retention year as well. Specifically, schoolwide student socioeconomic (SES) level has been found to have a large effect on student achievement (e.g., Rumberger & Palardy, 2005); in one study, school SES was found to have a larger effect on achievement than did individual SES (Ma & Willms, 2004). School policies and practices have also been implicated in at least one study as having a significant effect on

student achievement, such that when these are accounted for, the effects of school SES are rendered insignificant (Rumberger & Palardy, 2005). These results are notable in light of current educational policies in the State of Florida, including statewide retention laws as well as federally-funded Reading First programs, which have been established in some, but not all, of Florida's schools.

Though limited in scope and number, studies that examine retention specifically within a larger context of state or school policy have begun to emerge. Given the vast impact of recent federal legislation on educational practices, it will become increasingly important for studies to adopt this wider view of retention. In their 2005 study, Hong and Raudenbush offered an example of this expansion by considering not only the effect of retention on retained students, but also the average effect of a retention policy on all students, particularly on those who would be promoted under a schoolwide retention policy (i.e., high achievers). Similarly, Powell (2005) and Porter (2005) have contributed early results on student achievement using data from the State of Florida, the site of the current study. These mixed results, along with previous research, highlight the importance of utilizing a longitudinal design when examining retention, as true effects may not be seen within the first year after a student is retained in grade. Results from the former study (Powell, 2005) indicated a qualitative difference in subsequent performance among students promoted through the six good cause exemptions in place in Florida, specifically between the two exemptions requiring students to demonstrate proficiency through alternative methods and the remaining four exemptions, which do not require proficiency. Taken together, these studies provide clear examples of the direction

retention research must take in light of the current educational climate and legislation.

Studies describing negative outcomes and warning against the practice of retention are plentiful; however, it is imperative that these outcomes be interpreted with caution, as it is very difficult to design a methodologically sound study when looking at outcomes of grade retention. The grade retention literature abounds with methodological problems, including the absence of comparison groups (McCoy & Reynolds, 1999; Meisels & Liaw, 1993), a lack of consideration of socio-economic variables (Mantzicopoulos, 1997; Mantzicopoulos & Morrison, 1992), and a lack of longitudinal data (Buntaine & Costenbader, 1997). Perhaps most relevant to the current study, all previously published studies have failed to indicate what, if any, remediation efforts were made in conjunction with grade retention (e.g., Armistead, Kempton, Lynch, & Forhand, 1992; Gottfredson et al., 1994; Jimerson, 2001). This raises questions about the application of past results to the educational climate in Florida, as Florida's policy does not allow a child to repeat a grade without the provision of additional, intensive remediation in place.

Until July 2006, Florida law mandated remediation in the form of an Academic Improvement Plan (AIP) for each child who was performing below grade level, which would have included those who scored at Levels 1 or 2 on the FCAT-Reading. These plans were to consist of instructional modifications as well as clear and measurable academic goals that related to individual skill deficiencies. Examples of instructional modifications included pull-out services, one-on-one tutor instruction, peer tutors, and the employment of reading coaches. AIPs also included periodic evaluation to determine if

retained students were making progress towards their academic goals. While state law was quite clear regarding remediation requirements, specific data verifying treatment integrity of AIPs was lacking. Nevertheless, most of the students in this study were required to have AIPs in place following their FCAT-Reading performance in 2003, thus holding negative beliefs about the efficacy of the current student progression plan in Florida is premature as research evaluating the outcomes of the plan was limited. It should be noted that in July 2006, legislation replaced the AIP with a mandated Progress Monitoring Plan (PMP) that places greater emphasis on the use of data to monitor student performance; however, similar implementation and integrity issues need to be addressed (Florida Department of Education, 2006).

#### *Purpose of the Study*

The purpose of this study was to evaluate the longitudinal effects of third grade retention practices in the State of Florida, as measured by scores on the state mandated standardized test (FCAT-Reading). Retention as practiced in the State of Florida (i.e., within the context of a state mandated remediation program) is largely unstudied. Though exploratory data on the short-term academic outcomes of third grade retention for one cohort exist (Porter, 2005; Powell, 2005), this study will extend those findings by examining retained students' performance in fifth grade (i.e., three years post-retention) and comparing it to the fifth grade performance of low-achieving third grade students who were promoted to the fourth grade in 2003 through one of the two good cause exemptions that indicate proficiency. This study will also examine the relationship between subsequent performance on the FCAT-Reading and various student

characteristics (i.e., gender, race/ethnicity, socioeconomic status, retention status) and school variables (i.e., Reading First status, school size, school socioeconomic status), which likely play a large role in academic outcomes, regardless of individual student characteristics.

### *Research Questions*

In order to explore the effectiveness of third-grade retention, this study utilized FCAT-Reading scores from the 2003, 2004, and 2005 administrations to address the following research questions:

1. For Florida students who scored at Level 1 on the 3<sup>rd</sup> grade FCAT-Reading in 2003 and were retained in third grade during the 2003-2004 school year, what proportion scored at Level 2 or higher on the 5<sup>th</sup> grade FCAT-Reading in 2006?
2. For Florida students who scored at Level 1 on the 3<sup>rd</sup> grade FCAT-Reading in 2003, but were promoted to the fourth grade for the 2003-2004 school year through one of six good cause exemptions, what proportion for each exemption scored at Level 2 or higher on the 5<sup>th</sup> grade FCAT-Reading in 2005, and the 6<sup>th</sup> grade FCAT-Reading in 2006?
3. To what extent do differences in performance on the 5<sup>th</sup> grade FCAT-Reading exist between students who scored at Level 1 in 2003 and were retained in third grade during 2003-2004 and those who scored at Level 1 in 2003 but were promoted through one of the two good cause exemptions requiring a demonstration of proficiency?
4. To what extent are observed differences on the 5<sup>th</sup> grade FCAT-Reading between

students who were retained in 2003 and students who were promoted in 2003 through a good cause exemption indicating proficiency moderated by the following student characteristics: gender, race/ethnicity, and socioeconomic status?

5. To what extent are observed differences in performance on the 5<sup>th</sup> grade FCAT-Reading between students who were retained in 2003 and students who were promoted in 2003 through a good cause exemption indicating proficiency moderated by the following school characteristics: Reading First status, school size, schoolwide student socioeconomic status?

### *Significance of Study*

The present investigation was designed to function as a follow-up study to preliminary results (Powell, 2005) and addressed several methodological issues that plague the retention literature. First, it involved the analysis of long-term data, allowing for an examination of reading achievement of retained students three years after the retention. In addition, these data allowed direct comparison of low-achieving students who experienced retention and low-achieving students who were promoted due to a good cause exemption, as each have taken the fifth grade FCAT-Reading, albeit one year apart. Another important contribution of this study comes in its inclusion of school-level variables (i.e., Reading First status, school size, school socioeconomic status), which provided a context within which to interpret student outcomes.

Finally, there has been a consistent lack of focus on potential academic modifications implemented during the retained year. It is largely unknown whether the

students included in most previous studies received instruction during the retained year that was quantitatively and/or qualitatively different from that of the year prior. Because the State of Florida mandates a systematic plan of remediation for every child who is retained in the third grade (Florida Department of Education, 2003f), the current study aimed to examine the longitudinal outcomes of retention among students who were required to receive intense remedial instruction during the retained year. The findings of this study make an important contribution to the literature on grade retention by addressing an area that has heretofore gone unexamined.

### *Definition of Terms*

*Florida Comprehensive Assessment Test (FCAT).* The FCAT is a standardized test designed to measure student performance in reading, writing, and mathematics in accordance with the Sunshine State Standards. It also represents the foundation of the federally mandated and state created accountability system (Florida Department of Education, 2003c).

*Sunshine State Standards.* The Sunshine State Standards are a set of benchmarks developed by the Florida Department of Education and adopted in 1996. These standards identify the academic skills that the State of Florida wants its students to have attained at each grade level (Florida Department of Education, 2003c).

*Retention in grade.* This refers to the act of repeating a grade level. A student can be retained in grade but participate in curriculum not associated with grade level (e.g., retained in third grade due to reading difficulties but attend a fourth grade math class), or be promoted but return to previous curriculum for a problematic subject area (e.g.,

promoted to fourth grade, but attend a third grade reading class). While these types of gradations exist with regards to retention and promotion decisions, the present study focuses only on students who were retained in third grade after failing to score Level 2 or above on the FCAT. In Florida, retention is to be accompanied by intensive remediation efforts.

*Academic Improvement Plan (AIP).* An AIP is a set of formalized instructional modifications and related goals that are designed to address specific skill deficits in any child who is not meeting academic benchmarks in a timely manner; AIPs are state mandated in Florida (Florida Department of Education, 2002).

*Good cause exemptions.* These exemptions are six guidelines indicating the circumstances under which a student scoring Level 1 on the FCAT can still be promoted to the fourth grade. They include “[a]limited English proficient (LEP) students with less than two years in an English for Speakers of Other Languages (ESOL) program, [b]students with disabilities whose individual educational plan (IEP) indicated that participation in the FCAT was not appropriate, [c]demonstration of an acceptable level of performance on an alternate assessment...or scoring at the 51st percentile or higher on the norm referenced test portion of the FCAT, [d]demonstration of proficiency in Sunshine State Standards through a student portfolio, [e]students with disabilities who participate in the FCAT...(but)...still demonstrate a deficiency in reading after more than two years of intensive remediation and were previously retained in kindergarten, first, second, or third grade, and [f]students who still demonstrate a deficiency in reading after two or more years of intensive remediation and were previously retained in kindergarten,

first, second, or third grade for a total of two years” (Florida Department of Education, 2003e, p. 1).

*No Child Left Behind Act.* The No Child Left Behind Act is federal legislation passed on January 8, 2002, directly addressing the reading crisis in this country; it established strict guidelines for school and state accountability. The Act plans for every child to be reading proficiently by the year 2013-2014 (NCLB, 2002).

*Adequate Yearly Progress.* States are required by NCLB to set annual goals for adequate progress for their schools and to measure progress toward the attainment of those goals in a systematic way each year through standardized tests; funding is secured on the basis of states’ success in meeting those goals (United States Department of Education, 2003c).

*Reading First.* Reading First is a federal program established through NCLB that distributes additional funds to assist states in the implementation of empirically-based reading programs. Funding is based on the number of children aged 5-17 years who are considered low-income in each state (United States Department of Education, 2003d).

## CHAPTER TWO

### Literature Review

In this information age, with almost constant technological discoveries, it is astonishing to realize that only 32% of America's fourth graders read proficiently (United States Department of Education, 2002). Since the initial passage of the Elementary and Secondary Education Act in 1965, taxpayers have invested over \$321 billion in federal funds alone in public education; yet the average reading scores for 17-year-olds have not shown improvement since the 1970's (United States Department of Education, 2002). While scores of the top achievers have improved over time, those of the poorest performers have declined (United States Department of Education, 2002). Most of the children who are nonproficient in reading are minority students who live in poverty, creating a disturbing achievement gap. Recent federal legislation (e.g., No Child Left Behind Act) is reflecting acknowledgement of, as well as concern over this fact, by creating and implementing standards regarding student progression (United States Department of Education, 2003a). They are not only more rigorous for every child but also include new accountability procedures designed to ensure that substandard performance by any child is not tolerated as it was in the past.

#### *Enactment of Federal and State Legislation*

The No Child Left Behind Act (NCLB) was enacted on January 8, 2002 and has

changed the face of education across the country (NCLB, 2002). The law directly addresses the country's reading crisis by establishing strict guidelines for both student performance as well as school and state accountability. The overriding goal of the law is for every child to be reading at a proficient level by the 2013-2014 school year (United States Department of Education, 2003b). To this end, each state is required to demonstrate Adequate Yearly Progress (AYP) in educating its students, though states are allowed some freedom in the adoption of procedures used to monitor progress. States are required to set annual goals for adequate progress for their schools and measure progress toward the attainment of those goals in a systematic way through standardized tests; funding will be secured on the basis of states' success in meeting their established goals. In addition, schools that do not meet those academic standards will be identified as needing improvement and the parents of children enrolled in such schools can, with the aid of district funds, choose to send their children to a higher performing public school or to secure additional tutoring (United States Department of Education, 2003c).

Monitoring of academic progress is not a new concept in this country; however, the No Child Left Behind Act also has identified and addressed some issues with past methods of accountability. For example, previously, schools as well as states could report aggregated data across the total population of students; high scores could balance out very low scores, giving the impression that the school was effectively teaching all of its students. However, many schools were effectively reaching only a certain subset of students; minorities, poor students and children with disabilities or limited English proficiency (LEP) were consistently being "left behind." In response to this problem,

NCLB no longer allows data to be aggregated across students, schools, districts or states; test scores must be disaggregated and reported by subgroups based on economic background, race/ethnicity, English proficiency, and disability status. It is hoped that with these procedures in place, the achievement gaps that exist between groups of students will be diminished (United States Department of Education, 2003a).

In addition to lofty achievement goals and stringent accountability plans, the NCLB allows for additional federal funds to be distributed to states for the purposes of implementing scientifically proven, empirically based reading programs. The National Reading Panel (2000) identified five basic skills essential for early reading success, phonemic awareness, phonics, fluency, vocabulary and comprehension. States can apply for funds through the federal “Reading First” program with funding based on the number of children aged 5-17 years who are considered low-income (United States Department of Education, 2003d). The funds are to be used to teach the five basic skills in a systematic and evidence-based manner, and the students must be monitored closely (yearly state assessments) to ensure that they are moving toward success (United States Department of Education, 2003d).

Since each state can establish and enforce its own standards and design its own tests, it is helpful to look more specifically at Florida’s response to the No Child Left Behind Act. The State of Florida is committed to compliance with the new federal law and has set intermediate goals to this end. The state has established goals stipulating that 31% of students are to be reading proficiently by the 2003-04 school year, 48% by 2006-07, 65% by 2009-10, 82% by 2012-13, and 100% by 2013-14 (Florida Department of

Education, 2003a). The State has applied for and received \$52 million in annual funding (over \$300 million across six years) in Reading First funds and plans to use the money in schools around the state with large numbers of at-risk children (Florida Department of Education, 2003b). These schools must demonstrate through a competitive application process that they are committed to providing not only the required research-based reading programs, but an additional block of time devoted to the use of these techniques as well. Children in these schools are to receive nearly twice the amount of reading instruction as children in typical schools. The hope is that intensive reading intervention will have a preventative effect.

Prevention programs such as these have received a tremendous amount of attention and are argued to be a key in addressing the reading crisis in Florida. In addition to prevention, the State of Florida has implemented what some would consider to be rather drastic intervention strategies to address other current educational issues. New retention policies went into effect on January 7, 2003, the consequences of which continue to sweep the state (The Florida Senate, 2003). These policies include mandatory retention at the end of grade 3 for students whose reading deficiencies are not successfully remediated, as measured by performance on the Florida Comprehensive Assessment Test (FCAT). This policy is designed to align Florida policies with federal law to ensure that the educational system in this state will not let any child fail to learn to read; it will instead continue its efforts in assisting all children to achieve academic success. Inherent in the retention policy is an emphasis on prevention and/or remediation of reading deficiencies, which will be addressed later, as well as strong ties to the

standardized test used to quantify those deficiencies, the FCAT.

In looking at Florida's standardized test, the standards that it purports to measure cannot be ignored. The Sunshine State Standards, developed by the Florida Department of Education and adopted in 1996, identify the academic skills that the State of Florida wants its students to have attained at each grade level (Florida Department of Education, 2003c). For each grade, a series of benchmarks have been identified which, when met, move a child toward meeting the standards for that grade. Within the State of Florida, the FCAT represents the foundation of the federally mandated and state created accountability system; it purports to directly measure students' progress in reading, writing, and mathematics as outlined in the Sunshine State Standards.

The reading portion of this test is administered every year to students in grades 3-10. Students receive a scaled score for each subject, and from that scaled score they are categorized as performing at one of five levels, with Level 1 being the lowest and Level 5 being the highest. A great deal of attention is given to the lowest two levels; at Level 1, students are considered to be having "little success with the challenging content of the Sunshine State Standards." Students who achieve Level 2 are assumed to be having "limited success" with the same content. Statewide policy mandates that unless a "good cause exemption" can be demonstrated, all students achieving at Level 1 will be retained at the third grade level (The Florida Senate, 2003). These students are assumed to be struggling with learning to read such that they have not mastered the most basic skills at the most basic levels. Without these basic skills, chances of success in subsequent grades are severely limited, as the approach to teaching reading shifts at this critical juncture.

Prior to and through the third grade, students are learning to read; beginning in fourth grade, students are instead expected to utilize reading skills in all areas of their schooling. In essence, they would be “reading to learn” (Just Read, Florida, 2003).

According to the scores reported in 2003, almost one out of four (23%) of Florida’s third grade students were reading at Level 1, indicating potential retention for the following school year of approximately 47,000 students (Florida Department of Education, 2003d). Results the following year showed only slight improvements, with 22% of students achieving Level 1 on the FCAT-Reading (Florida Department of Education, 2004b). Six guidelines, or “good cause exemptions” indicate those circumstances under which a student scoring Level 1 on the FCAT can still be promoted to the fourth grade. These include “[a]limited English proficient (LEP) students with less than two years in an English for Speakers of Other Languages (ESOL) program, [b]students with disabilities whose individual educational plan (IEP) indicated that participation in the FCAT was not appropriate, [c]demonstration of an acceptable level of performance on an alternate assessment...or scoring at the 51st percentile or higher on the norm referenced test portion of the FCAT, [d]demonstration of proficiency in Sunshine State Standards through a student portfolio, [e]students with disabilities who participate in the FCAT...(but)...still demonstrate a deficiency in reading after more than two years of intensive remediation and were previously retained in kindergarten, first, second, or third grade, and [f]students who still demonstrate a deficiency in reading after two or more years of intensive remediation and were previously retained in kindergarten, first, second, or third grade for a total of two years” (Florida Department of Education,

2003e, p. 1).

### *Recent Trends in Grade Retention*

Historically, grade retention has been viewed in education as the exception, rather than the norm, yet has been a widespread practice despite persisting empirical evidence that it is not typically a successful intervention strategy (e.g., Jimerson, 2001; Phelps, Dowdell, Rizzo, Ehrlich, & Wilczenski, 1992). Despite the promotion of 12,403 (6.5%) students due to “good cause”, a total of 28,028 (14.6%) third grade students were retained after the 2002-2003 school year due to scoring at a Level 1 on the FCAT (Florida Department of Education, 2003d). After the 2003-2004 school year, similarly large numbers of third grade students were retained (23,283; 11.0%) due to poor FCAT performance (Florida Department of Education, 2004a). By the end of the 2006 academic year, the percentage of third grade students retained due to scoring at Level 1 on the FCAT-Reading had decreased to 6.8% (Florida Department of Education, 2007a). While a downward trend seems to be emerging, there still have been tremendous numbers of affected children each year. In light of this, it is important to re-examine the grade retention literature, looking past the longstanding assumptions with regards to the intents, functions, and long-term outcomes of retention.

The body of retention literature is extensive and spans several decades; however, much of the earliest research is saddled with substantial methodological flaws, including a lack of comparison group (Simmons & Blyth, 1987), and an insufficient sample in terms of size and diversity (Jackson, 1975). In addition, much of the earlier work in this area was unpublished. In an early review of 63 retention studies, only 20 had gone

through the rigorous review of a refereed publication process (Holmes, 1989). Thus, the scientific quality of many early studies could not be monitored. Due to these issues, this review will not attempt to be exhaustive; rather, it will focus on the most recent and relevant research to date and will provide a synthesis of the findings. This review is organized by the following categories: (a) characteristics of retained children, (b) evidence of positive outcomes of retention, (c) evidence of negative outcomes of retention, (d) factors that determine/contribute to retention outcomes, (e) effects of retention policy, (f) methodological issues to consider, and (g) future directions in retention research.

### *Characteristics of Retained Children*

While low academic performance has been assumed to be the primary reason to retain a child in grade, significant differences in achievement between retained students and low-achieving-but-promoted students have not been found, suggesting that among low-achieving students, social and/or behavioral factors may influence decisions to retain (Jimerson, Carlson, Rotert, Egeland, & Sroufe, 1997). Over the years, certain factors have been identified that consistently characterize retained students, providing an opportunity to predict who is more likely to be retained in the future. These factors include being male, an ethnic minority, and socially immature. Additionally, having poor emotional health, maladaptive or aggressive behaviors, and reading or math problems have been found to put a child at higher risk for retention (Jimerson, 1999; Jimerson et al., 1997; McCoy & Reynolds, 1999; Reynolds, 1992). Children also are more likely to be retained if they live in poverty, change schools frequently, or have parents who are not

involved in school, do not value education or have completed limited education themselves (McCoy & Reynolds, 1999; Reynolds, 1992). It is unclear whether these factors, particularly the demographic characteristics, are themselves decreasing the odds of academic success, or if they are serving as mediators on an unidentified variable to put those children at a higher risk of being retained.

As retained children generally lag behind their peers academically, it is assumed that one goal of retention is remediation. Therefore, children who are retained should show an improvement in academic performance following additional time in the same grade; without this improvement, it would be hard to justify such a costly method of intervention, both financially, as well as emotionally. Though outcome research has yielded mixed results in this area, with findings generally indicating negative effects, several studies indicate that early retention can benefit the future academic careers of low achieving students.

#### *Evidence of Positive Outcomes of Retention*

A recent study by Mantzicopoulos (1997) serves as an example of positive outcomes resulting from retention. The 40 children included in this study were selected in the spring of their kindergarten year, based on high inattention status (as measured by teacher ratings using the Revised Problem Behaviors Checklist) from a matched group of 62 children who had participated in a previous study (Mantzicopoulos & Morrison, 1992). This larger group had been matched on demographic and academic variables including sex, age, at-risk status and reading/math achievement. Of the 40 children included in the study, 25 had been retained in kindergarten and 15 had been promoted;

the progress of all children was followed through the second grade. Measures included the SEARCH instrument, which is composed of 10 subscales designed to assess spatial and temporal orientation skills deemed necessary for beginning reading, the Revised Problem Behaviors Checklist (RPBC) to measure the severity of problem behaviors, and the Total Reading and Total Math subtests from one of two standardized achievement tests (Stanford Achievement Test or California Test of Basic Skills). Children were given the SEARCH screening during their kindergarten year and were rated with the RPBC each year. Academic achievement tests were administered near the end of each school year.

Repeated measures multivariate analyses of covariance were performed for both same-age and same-grade comparisons. The results for the same-grade comparisons indicated a significant main effect for math achievement, favoring the retained group ( $F = 5.63, p < .05$ ). In addition, the adjusted mean differences showed that retained children performed above the national mean in math while promoted children remained below the national mean in first and second grades ( $M_{Adj} = -0.25$  and  $-0.18$ ), respectively. No significant differences were found in reading performance.

Same-age comparisons yielded similar results; retained children performed significantly better than promoted children in math, while the reading performance of both groups showed a slight decline over time, with comparable levels at the end of second grade. The author concluded that retention was not a beneficial educational intervention due to the lack of consistent findings in reading; however, for the purposes of this review, the clear increase in math performance cannot be overlooked.

Alexander, Entwisle, and Dauber (1994) offer more conclusive evidence of retention as a positive intervention tool. Though never achieving at the same level as never-retained students, the retainees in their study showed marked improvements over time. The authors utilized a stratified random sample of 800 Baltimore children entering first grade and monitored their school progress for eight subsequent years. The design was longitudinal, leading to subject groups that were too complex to be separated into distinct retained versus promoted groups based on one point of time. For instance, 127 first graders were retained in year 1 of the study. Over the next eight years, because of erratic grade-level progressions such as double retentions (49), triple retentions, (3), mid-year retentions (1), mid-year promotions (17), and double promotions (12), only 9 of the original 127 students who had been retained in year 1 were on grade level after eight years. In a similar vein, by year eight of the study, 142 children were in the seventh grade, one year behind the progression schedule that would be expected if there had been no deviations; however, only 22 were actually seventh grade repeaters; the vast majority were a year behind because of retention(s) earlier in their school careers.

This design resulted in a large amount of long-term data, garnered mainly from school records (including test scores and grade reports), one-on-one interviews with the children and their parents, and self-administered questionnaires by teachers. This study was unique in comparison to other recent retention research in that it did not employ any type of matched control group. As will be discussed later, an appropriately matched comparison group is very difficult to construct and define consistently across studies, leaving the possibility that an undetermined variable could skew the results; these authors

chose to circumvent these issues by using statistical methods to adjust for prior performance, as well as demographic risk factors such as low socioeconomic status, low parent education, minority status, and low school readiness. The statistical adjustments allowed them to make comparisons between retainees, those who performed similarly but were promoted, and all remaining never-retained children. In addition, the researchers compared students' post-retention performance to their own pre-retention performance in order to examine a change in trajectory.

To measure achievement, the researchers administered two subtests from the California Achievement Test battery (reading and math) twice annually. The reading portion consists of 20-40 items (20 items in first and second grade versions, 27 items in third grade version, and 40 items for all subsequent versions), while the math subtest consists of 36-45 items (36 items in first grade, 40 items in second grade, and 45 thereafter). The authors report scaled scores that have been calibrated to account for all versions of the test, allowing for meaningful interpretations across grade levels. With regard to the CAT-R (reading), children who were retained at the end of first grade ( $n = 127$ ) lagged significantly behind their same-age, never-retained peers at the beginning as well as at the end of the first grade (33.5 points lower in the fall; 59.5 points lower in the spring;  $p < .01$ ), but had made up seven points in reading and four in math when compared to those same peers after the retained year ( $p < .01$ ). Gains were larger when comparing the children retained in first grade to younger, same-grade peers 17 points higher in math ( $p < .01$ ). Unfortunately, these gains were not maintained past the first follow-up (i.e., after one year of retention in first grade); these children's scores in later years remained similar

to what they had been in the pre-retention year.

The authors noted that retention in the first grade seemed qualitatively different from retention in later grades; they were uncertain as to whether this was due to the first grade retention itself or because first grade retainees' academic difficulties were so severe. They found that for children retained in the second and third grades, results were more positive. Second grade children were farther behind their peers at the end of the year than the first grade retainees had been; however, after retention for one year, they had improved their standing relative to same-age peers. At the time of retention, they scored an average of 52.0 points lower than their peers; after one year of retention in second grade, they scored only 34.7 points lower ( $p < .01$ ). In addition, it appears that those gains were long-term, as by the seventh year of the study they were only 24.8 points lower than never-retained peers ( $p < .01$ ). This pattern of improvement was even stronger for children retained in third grade, who lagged 28.4 points behind their same age peers at the point of retention, but only 9.7 points behind by year 7 of the study ( $p < .01$ ).

The authors point out that these retainees were still lagging far behind their never-retained peers, seen in the adjusted scores, but especially in the unadjusted numbers. They stated that it is crucial, however, to remember the retained children's pre-retention problems. Whatever the problem causing poor achievement, it was present before the retained year, leading these authors to conclude that the retention year actually improved a previously charted course. They hypothesized, based on early test scores, that without retention, these children would have continued to fail, and that the gap between their

scores and those of never-retained children would have continued to widen. While this explanation makes conceptual sense, it lacks an empirical basis, as this study lacked the data to pinpoint specific variables involved in school failure, as well as detailed information regarding which aspects of retention made a positive impact on children's academic careers.

To summarize, these studies and a small number of others (Gottfredson, Fink, & Graham, 1994; Peterson, DeGracie, & Ayabe, 1987) have reported positive results to varying degrees. Unfortunately, results such as these appear to be in the minority, as past research has consistently shown retention to be an overwhelmingly negative experience for children. Small academic gains may be seen in certain retainees (Mantzicopoulos, 1997), especially when comparing them with younger, same-grade peers (Alexander et al., 1994); however, these gains are apparently temporary and as these and other studies have shown, within two to three years, their achievement is equal to or lower than that of both same-grade and same-age regularly promoted students (Jimerson, 1999; Jimerson et al., 1997; Mantzicopoulos, 1997; Mantzicopoulos & Morrison, 1992; McCoy & Reynolds, 1999). Interestingly, studies that have found modest, temporary increases in performance most often find them in math; reading remains a consistent and distinct deficit in retained children (Jimerson et al., 1997; Mantzicopoulos, 1997). This finding is of particular relevance considering that the retention policy in the State of Florida targets reading remediation as its primary goal. It should also be noted that according to the existing literature, socio-emotional and behavior deficits such as inattention, immature or aggressive behavior, low self-esteem, and poor peer relationships tend to be exacerbated

rather than improved by retention (Jimerson, 1999; Jimerson et al., 1997; Mantzicopoulos, 1997). More specifically, early grade retention has been found to increase physical aggression in boys (Nagin, Pagani, Tremblay & Vitaro, 2003). The direct effects of socio-emotional/behavioral health on academic performance are not the focus of this review; however, issues such as these are likely to have a negative overall effect on a child's school experience.

#### *Evidence of Negative Outcomes of Retention*

Although the present study will target only academic outcomes of retention, numerous studies have examined both behavioral and academic factors in retention (Ferguson et al., 2001; Gottfredson et al., 1994; Mantzicopoulos, 1997; Mantzicopoulos & Morrison, 1992). Pagani, Tremblay, Vitaro, Boulerice, and McDuff (2001) maintained this dual focus in their research efforts. Their 1,830 subjects were selected from a larger sample (6,397) of kindergarten students who had been randomly selected for a separate study and followed through age 12 (Vitaro, Desmariais-Gervais, Tremblay, & Gagnon, 1992). The independent variable was grade retention (one year maximum) and the dependent variables were teacher ratings of children's global academic performance as well as scores on the Social Behavior Questionnaire, which is based on teacher ratings and is comprised of items from the Preschool Behavior Questionnaire and the Prosocial Behavior Questionnaire. In order to control for differences in naturally occurring trajectories expected from children of differing achievement levels, the researchers used a statistic relatively unique to retention research, a basic autoregressive model. Testing separately for boys and girls, results indicated that grade retention negatively affected

multiple areas of children's development, regardless of prior characteristics. With regard to academics, for example, 10-year-old boys experienced negative deviations from the expected trajectory of academic growth after being retained between the ages of 6 and 8 years (path =  $-.12$ ,  $p < .01$ ). This negative effect was still evident at age 12. In girls as well, early grade retention had a profound and lasting negative effect on later academic performance (measured at ages 10 and 12), as their trajectories had diverted negatively from the expected path at age 10 (path =  $-.07$ ,  $p < .01$ ) and remained off course at age 12 (path =  $-.07$ ,  $p < .05$ ). The authors also report negative behavioral outcomes, particularly in boys, for whom a persistent and worsening of disruptive behavior was reported. They hypothesized that these behavioral outcomes and the emotional component that accompanied them could have interfered with the learning process and could, therefore, have been acting upon the academic outcomes that were observed. While their hypothesis is worth further examination, the most salient information involves the authors' evidence-based conclusions that retention leads to long-lasting negative effects.

McCoy and Reynolds (1999) provided another illustration of negative retention outcomes in findings from their Chicago Longitudinal Study. This study examined school achievement of children who were graduates of a government funded kindergarten program and who were at risk for underachievement due to poverty. Participants were seventh and eighth grade students (depending on retention status) at the time of enrollment, thus eliminating the opportunity to match subjects based on pre-existing achievement or behavior characteristics known to predict retention. Of 1,164 children, 315 had been retained in grade at least once. Students were assessed using the reading

comprehension and mathematics subsections of the Iowa Test of Basic Skills (ITBS, Level 13 or 14;  $\bar{X}=145.06$ ,  $SD=22.38$ ) and hierarchical multiple regression was used to analyze the results. When comparing same-age retained versus non-retained children, grade retention was found to consistently predict reading and mathematics performance ( $R^2 = .47$ ,  $R^2 = .57$ , respectively) as measured by this standardized test. Retention was significantly and negatively correlated with reading achievement scores (10.6 standard score points lower,  $p<.01$ ), even after the inclusion of socio-demographic factors and early adjustment indicators in the analysis. The same finding was true of mathematics achievement (9.7 standard score points lower,  $p<.01$ ), although fewer socio-demographic characteristics were related. The authors also conducted analyses using same-grade comparisons; again, retention was found to predict negative outcomes. Retained seventh grade students scored significantly lower than promoted seventh graders in reading achievement ( $\beta = -4.6$  standard score points,  $p<.001$ ); they also scored lower in mathematics ( $\beta = -1.3$  standard score points), but the difference was not significant. The achievement gap between retained and promoted (same grade) students was found to have widened four years later, despite the stated goal of retention as decreasing the gap. The authors concluded, therefore, that retention was ineffective as an intervention and even harmful to students' future academic achievement.

Such unequivocal conclusions that retention outcomes have a negative impact on children are not unique; still, caution needs to be exercised when interpreting results, as methodological issues abound in retention research. For example, although findings from the McCoy and Reynolds (1999) provide important preliminary information, several

limitations in the design of their study indicate the necessity for further corroboration. The authors used a correlational design with a low-income sample, limiting the generalizability; however, it is the retrospective nature of the study that is most problematic. Enrolling subjects after the retention occurred disallowed an examination of predisposing factors, as well as examination of the trajectories of experimental and control groups. Partly as a response to these and other design issues, Jimerson et al. (1997) attempted to examine retention effects by utilizing a prospective longitudinal design. Since matching is not typically feasible when looking at students only after retention has occurred, Jimerson et al. (1997) utilized subjects whose mothers had been participating in another longitudinal study prior to the birth of their children. These children were identified prenatally as being at-risk for social and emotional development problems due to maternal characteristics. They were followed through age 16 and placed into one of three research groups: retained once in grades K-3 ( $n = 29$ ), low-achieving promoted ( $n = 50$ ) and control ( $n = 100$ ). The children in the low-achieving promoted group served as a matched comparison group, as they did not differ from the retained children in terms of academic achievement or intellectual functioning as measured by standardized assessments (Peabody Individual Achievement Test, Wechsler Preschool and Primary Scale of Intelligence, Wechsler Intelligence Scale for Children-Revised, Woodcock-Johnson Achievement Test-Revised). Interestingly, they did display fewer maladjusted behaviors in the classroom ( $p < .05$ ) and higher levels of emotional health ( $p < .01$ ) than did retained students, based on teacher interview, the Child Behavior Checklist-Teacher Form, and emotional health/self-esteem measures developed by the

authors. Short-term effects were examined one year following retention; retained students did not appear to display an advantage over the comparison students in academic functioning or school adjustment. When prior achievement levels were covaried, the retained group showed a slight advantage in math achievement ( $p < .01$ ). Students were assessed in the sixth grade as well as at 16 years of age to examine long-term effects of retention. Results favored the comparison group on measures of emotional health ( $p < .05$ ), but the two groups did not differ in academic achievement. Essentially, despite an extra year of instruction, the retained students were comparable to low-achieving promoted students in the short term as well as years later. Though this is not as negative a conclusion as the outcomes found by other studies, neither is it supportive of retention as a positive, successful intervention for underachieving students.

In an effort to extend these results, Jimerson (1999) followed these same students into their 21<sup>st</sup> year of life. The students were assessed in 11<sup>th</sup> grade and again at age 19, in an effort to answer questions regarding academic adjustment (as measured by grade point average, credits earned, behavioral problems and attendance), dropout status, receipt of a certificate of high school completion and postsecondary education. In reference to academic adjustment in 11<sup>th</sup> grade, the retained group of students achieved significantly lower levels as compared to both the comparison ( $F = 6.59, p < .01$ ) and the control ( $F = 13.95, p < .001$ ) groups. The latter two groups were not significantly different from each other. By age 19, a larger percentage of retained children had dropped out of high school as compared to both the comparison ( $\bar{X} = 3.57, p < .05$ ) and control groups ( $\bar{X} = 13.79, p < .001$ ), with the comparison group having a significantly greater number of dropouts

than the control group ( $F = 3.77, p < .05$ ). By age 20, a significantly lower percentage of the retained students had received a certificate of high school completion as compared to both the comparison ( $\bar{X} = 5.44, p < .01$ ) and control ( $\bar{X} = 23.66, p < .001$ ) groups. As expected, the number of retained students who eventually enrolled in postsecondary education was significantly lower than the number of control students ( $\bar{X} = 8.74, p < .01$ ). Incidentally, as adults, retained students were less likely than were comparison students to have a full-time job, be a full-time student, or be involved in a combination of work and school ( $F = 4.62, p < .05$ ). They also earned lower wages ( $F = 4.09, p < .05$ ) and were rated as less competent in the workplace ( $F = 4.39, p < .05$ ). Low-achieving-but-promoted students were more comparable to the control group than to the retained group in regards to all previously described employment outcomes.

Fine and Davis (2003) found similar results in their study examining the post-secondary enrollment (PSE) rates of high school graduates who had been retained at some point in their school careers. These students were labeled “persisters”, as they graduated from high school despite a previous retention. The sample consisted of 11,637 high school graduates who completed surveys in 1988 (base year), 1992 (graduation year), and 1994 (follow-up). The surveys included questions regarding retention status, grade of retention, race/ethnicity, gender, and enrollment in post-secondary education. Parent surveys included a series of questions used to determine socioeconomic status (SES; e.g., educational attainment, occupation). Holding constant demographic variables (i.e., gender, race, SES, and previous achievement), researchers conducted a logistic regression analysis to calculate the odds of a persister enrolling in PSE. Results

demonstrated that grade retention greatly diminished the odds of those students enrolling in a 4-year institution (.47,  $p < .001$ ) or any PSE (.62,  $p < .001$ ) when compared to promoted graduates. Furthermore, post-secondary education enrollment odds tended to drop as the grade of retention increased. For example, following a retention in kindergarten through second grade, the odds of enrolling at a post-secondary institution was .67 ( $p < .002$ ) when compared to promoted graduates; following a retention during sixth through eighth grades, the odds dropped to that of promoted peers, .32 ( $p < .001$ ). Like Jimerson (1999), these authors concluded that grade retention was associated with negative long-term outcomes; even those who manage to graduate from high school were less likely to access the economic and social benefits of a post-secondary education. The authors also recognize that the age of the data (first collection occurred in 1988, 15 years prior to publication) as well as the inclusion criteria [i.e., retained students who earned a general equivalency diploma (GED) were not included in the study, although 25% of them went on to enroll in post-secondary education] constituted substantial limitations to the interpretation of the results. However, it should be noted that despite these limitations, the findings corroborated the existing body of literature examining the long-term, negative outcomes of early grade retention.

#### *Factors that Influence Retention Outcomes*

It is reasonable to ask which factors, if any, can be used to first determine why some children experience success after a retention year, in the hopes that the information can be utilized to manipulate outcomes for all children. Ferguson, Jimerson and Dalton (2001) examined within group variables in an attempt to answer this question. These

authors looked at the performance of retained (only once in grades K-2;  $n = 58$ ) and promoted ( $n = 48$ ) students on several school readiness (Brigance Kindergarten and First Grade Screen, Gesell School Readiness Test, Metropolitan Readiness Tests) and standardized achievement instruments (Science Research Associates' Survey of Basic Skills, 2<sup>nd</sup> and 5<sup>th</sup> grades; Stanford Achievement Test 8<sup>th</sup> Edition, 8<sup>th</sup> and 11<sup>th</sup> grades) as well as on academic performance indicators such as grade point average (GPA) and teacher ratings, which were administered at various points from 2<sup>nd</sup> through 11<sup>th</sup> grades. They defined "successful" retained students as those having a mean seventh grade GPA of at least 3.2 and a mean eighth grade SAT (Stanford Achievement Test) score at the 53<sup>rd</sup> percentile or higher. These students represented 25% of all retained students, had scored higher on early school readiness measures ( $p < .05$ ), were significantly younger ( $p < .05$ ), and were less aggressive ( $p < .01$ ) than unsuccessful retained peers. In addition, they were from families with a higher level of maternal education; their parents also placed a higher value on education for their child.

Additional literature also supports the impact of student variables; for example, males are more likely to be retained and to experience negative outcomes after the retention year (e.g., McCoy & Reynolds, 1999). Students living in poverty and those who belong to a racial or ethnic minority group have long been known to be more likely to experience academic difficulties, in terms of general achievement (e.g., McCoach, O'Connell, & Reis, 2006) and specifically relating to retention outcomes (Ferguson et al., 2001; Jimerson et al., 1997). Such findings suggest that the outcomes of retention, positive or negative, are dependent on variables beyond just retention itself. Perhaps

criticism that focuses solely on retention should be replaced with investigation of mediating factors that may influence outcomes.

When considering factors that may mediate academic outcomes, it is helpful to broaden the discussion beyond individual student characteristics and examine ecological factors that might influence school performance. Rumberger & Palardy (2005) conducted a study investigating the potential impact of racial and socioeconomic segregation on high school achievement levels. They examined survey results for 14,217 students attending 913 high schools across the country. Data were collected at three time points, when students were in grades 8, 10, and 12. High-SES schools (n=151) were defined as those with an average student SES at least one standard deviation above the overall average SES; middle-SES (n=641) schools had an average student SES within one standard deviation of the mean SES, and low-SES schools (n=121) had an average SES at least one standard deviation below the mean SES.

To account for the multi-leveled data, a hierarchical linear model (HLM) was used, with test scores in reading, mathematics, science, social science, and a composite score serving as dependent measures. A comprehensive set of independent variables measuring individual, family, and school characteristics was used, including data from students, parents, and school personnel. Overall, results demonstrated that school-wide socioeconomic status, not racial composition of the student body, impacted achievement, a finding consistent with the body of literature on this issue (e.g., Coleman, 1990; Jencks & Mayer, 1990, Ma & Willms, 2004).

Findings also indicated that while student variables accounted for nearly three-

fourths (.872) of the variability in initial achievement as measured by the composite test score, they accounted for only approximately one-third (.317) of the variability in achievement growth over time, indicating that school characteristics likely played a role in explaining the remaining variability in achievement growth. Notably, however, after controlling for school policy and practice, the effects of school SES on student learning were not significant, a finding that is particularly relevant in light of current changes to educational policy across the country.

#### *Effects of Retention Policy*

Familiarity with the impact of school-level characteristics on student achievement is especially critical in the face of unique contextual variables created by new state, district, or local retention policies. Given the impact of recent federal legislation on educational practices, it is not surprising that after nearly a century of research examining retention as a student-level variable, studies are beginning to emerge that examine retention within the larger context of policy (i.e., school, state; Hong & Raudenbush, 2005; Powell, 2005). Hong & Raudenbush expanded investigation of retention to include the policy level by considering not only the effect of retention on retained students, but also the average effect of a retention policy on all students, particularly on those who would be promoted under a school-wide retention policy (i.e., high achievers). The researchers utilized data from the Early Childhood Longitudinal Study Kindergarten cohort (ECLS-K) which was released by the US National Center for Education Statistics (NCES). The data were composed of repeated measures of nationally representative kindergarten and first-grade students and their families, teachers and schools (collected

between 1998 and 2000) and included 11,843 students (471 kindergarten retainees, 10,255 promoted) in 1,080 schools with a retention policy (i.e., retention schools) and 1,117 students (all promoted) in 141 schools without a retention policy (i.e., non-retention schools). Outcome variables that were examined included reading and math achievement, measured through four repeated assessments over two years and calibrated according to item-response theory (IRT). Since student outcomes across schools are likely related to the school that one attends, which would include school, classroom, teacher, and peer variables, the authors chose to combine multilevel modeling and propensity score stratification to adjust for this selection bias. Results indicated that students attending retention schools were not better off in terms of reading and math when compared with students in non-retention schools; within stratum, policy effects ranged from -4.00 to 4.96 in reading and from -3.83 to 3.78 in math. Given this potential bias, a two-level hierarchical linear model was constructed in which adjustments for selection bias were made. Based on the results of this analysis, the authors concluded that overall effects of a retention policy were very small (mean differences = -.24 in reading, -.14 in math). In additional analyses, policy effects on the subset of children who were not at risk of being retained (i.e., high-achievers) were examined and no effects on these children's learning were found. To estimate the average effects of kindergarten retention on the at-risk students (i.e., low achievers) attending retention schools, another two-level hierarchical linear model was constructed. Results demonstrated that if an at-risk child who was promoted was instead retained, his or her reading achievement would be expected to be 9.01 ( $p=.013$ ) points lower at the end of the treatment year. For math, this

expected score would be 5.89 ( $p=.005$ ) points lower. Another way to consider these results comes in the form of growth trajectories, which reveal that the achievement gap between the retained and promoted students widened over the treatment year from about 10 points to almost 21 points. Counterfactual growth trajectories hypothesized under this model predict that had the retained children instead been promoted to first grade, they would have demonstrated a growth rate similar to that of their peers and this achievement gap would have been much smaller than was actually observed. Thus, the combined results of this study established very small or null average school-wide effects of the policy, no benefits for high-achieving students, and negative effects for retained students in both reading and math. The authors concluded that kindergarten retention left most retainees even farther behind, and that at-risk children who were promoted to the first grade had a better chance of increased growth in learning. Despite these results, the extent to which they can be generalized is questionable, as transition from kindergarten to first grade may be inherently different, conceptually speaking, from the transition between other grades. In addition, the basis for retention decisions in retention schools is unclear, as is the context of these school-based retention policies. Given the national sample, it is possible that the schools were spread out geographically, and the impact of a policy in a single school or district may have affected student outcomes differently than a statewide policy. Nevertheless, this study indicated that retention research is beginning to focus on the impact of systems variables on student achievement.

Another example of a systems focus is found in a preliminary study examining the academic outcomes of retention within a state-mandated remediation program

(Powell, 2005). Using data obtained from the Florida Department of Education, this study attempted to determine the academic achievement of low-achieving students (i.e., scoring at a Level 1 on the FCAT-Reading) who were either retained or promoted through a good cause exemption. Data were analyzed for 29,317 students who scored at a Level 1 on the third grade FCAT-Reading in 2003; 20,618 students were promoted and 8,699 were promoted through a good cause exemption. The following year, the retained students took the third grade FCAT-Reading for a second time and results showed that 12,806 retained students (62%) scored at a Level 2 or higher (i.e., passing), while 7,812 (32%) retained students scored at a Level 1 again. When considering students promoted due to good cause, it is important to distinguish between the two exemptions that require an alternative demonstration of proficiency (e.g., alternative test, student portfolio) and the four exemptions (n=5,301; 61%) that allow students to be promoted despite a lack of proficiency (e.g., previous retentions, Limited English Proficiency). Of the 3,398 (39%) students promoted through one of the exemptions demonstrating proficiency, 2,220 (65%) scored at a Level 2 or higher on the 4<sup>th</sup> grade FCAT-Reading, while only 1,181 (35%) of those promoted through one of the “nonproficient” exemptions performed at the same level. A logistic regression analysis was conducted to investigate the impact of students’ gender, race, socioeconomic status, and previous achievement on future “success” (i.e., Levels 2 or higher on 2004 FCAT-Reading level). Results revealed that this four-factor model more reliably predicted student success than did a constant-only model,  $\chi^2(6, 20001) = 4170.81$  ( $p < .0001$ ). More specifically, the model correctly predicted 88.3% of the students achieving success, 44.6% of students not achieving

success, and 71.7% of students overall. Results also indicated that even when controlling for previous performance (i.e., 2003 FCAT-Reading score), success was significantly associated with gender, race, and SES. Specifically, African American students were less likely ( $B = -.262$ ) than their Caucasian peers to achieve success on the FCAT-Reading in 2004 ( $p < .0001$ ); their odds of scoring at Level 2 or higher were only .77 times that of the odds for a Caucasian student. In addition, students of low SES were less likely ( $B = -.266$ ) to achieve success on the FCAT-Reading, and had decreased odds (.77) of scoring at Level 2 or higher when compared to peers classified as Not low SES. And finally, with regard to female students ( $B = .122$ ), results showed that they were more likely to achieve success than male students; the odds of a female scoring at Level 2 or higher in 2004 were 1.13 times that of the odds for a male student ( $p < .01$ ).

In a related study that utilized the same database, Porter (2005) examined the cluster of students scoring within 10 points in either direction of the state-defined cut-off score between Levels 1 and 2. The FCAT-Reading is scored on a continuum between 100 and 500 points; scores of 100-258 are classified as Level 1, with 259 being the lowest score possible in Level 2. In addition to defining levels (1-5) of performance, this cut-off is also significant in that students scoring 258 or below face a state-mandated grade retention, while students scoring 259 or above are not subject to that outcome. Porter included two sub-samples of students in his study, those who received a scaled score of 248-258 (i.e., higher achieving retained students) and those who received a scaled score of 259-269 (i.e., low achieving promoted students). Test scores one year later indicated that 32% of the low achieving promoted students scored at a level 1 on the 4<sup>th</sup> grade

version of the FCAT-Reading, while 13% of the higher achieving retained students scored at a level 1 on the 3<sup>rd</sup> grade FCAT-Reading. Low achieving promoted students who were most likely to succeed the following year were female, Hispanic, Asian, and Caucasian. Higher achieving retained students most likely to succeed the following year were Hispanic and Caucasian; gender was not significantly associated with outcomes for this group. Logistic regression analyses were conducted for each group of students, with gender, race, SES and previous score on the FCAT-Reading entered into the model as predictor variables. Results showed that for the low achieving promoted group, the model with predictors was significantly more effective in predicting students' future success on the FCAT-Reading than was a constant-only model ( $\chi^2(6,9007) = 167.02, p < .01$ ). At a .67 probability level, the model correctly predicted 59% of the students achieving success, 53% of students not achieving success, and 57% of students overall. In addition, when previous achievement was held constant, female students were significantly more likely than male students to achieve state standards ( $B = .135, p < .01$ ); their odds of scoring at a Level 2 or higher on the FCAT-Reading in 2004 was 1.15 times that of male counterparts. African-American students were less likely ( $B = -.266, p < .01$ ) to achieve state standards than their Caucasian peers, with their odds of success being .77 times that of a Caucasian student. Hispanic students were more likely to achieve state reading standards than their Caucasian peers ( $B = .165, p < .01$ ), with 1.18 times the odds of achieving success when compared with Caucasian peers. Finally, students of low SES were less likely ( $B = -.273, p < .01$ ) to meet state reading standards and had decreased odds (.76) of scoring at Level 2 or higher when compared to peers classified as Not low

SES.

Results of a logistic regression analysis conducted with the higher achieving retained group suggested that the five-factor model was a good fit ( $\chi^2(df) = 51.33, p < .01$ ), indicating that the model with five factors was significantly more effective in predicting students' achievement of state standards than a constant only model. More specifically, and at a .87 probability level, the model correctly predicted 63% of the students achieving success, 52.1% of students not achieving success, and 61.6% of students overall). When controlling for previous achievement, only race was significantly associated with the performance level obtained on the 2004 FCAT Reading-Test. Specifically, African-American students were less likely than their Caucasian peers to achieve success ( $B = -.594, p < .01$ ) in 2004; their odds of scoring at Level 2 or higher were only .55 times that of the odds for a Caucasian student.

Though these studies (Porter, 2005; Powell, 2005) are among the first to directly examine retention outcomes within the specific context that has been created in the State of Florida, they are not without their limitations. First, they utilize only cross-sectional data, when previous research has shown that longitudinal outcomes of retaining students in grade are often more negative than early results tend to suggest. These studies fall prey to many of the flaws inherent in other retention studies, including retrospective sampling, lack of a control group, and largely correlational analyses. In addition, the analysis that was used (i.e., logistic regression) in these studies, as well as in others (e.g., Jimerson, 1999; McCoy & Reynolds, 1999) assumes independent data, which is violated in these studies due to the nesting of students within schools. As pointed out by Hong &

Raudenbush (2005), it is likely that school-level variables, including teachers, peers, curriculum, and demographics play a role in student outcomes and it is important that future studies consider such factors.

#### *Methodological Issues to Consider*

The potential impact of moderating variables on retention effects is only one of many reasons to reexamine longstanding assumptions that retention is harmful to students. For example, the retention literature has been characterized by various and longstanding methodological issues, which underlie the difficulties in making accurate interpretations of results. Design deficiencies include, but are not limited to, the absence of comparison groups (McCoy & Reynolds, 1999; Meisels & Liaw, 1993), a lack of consideration of socio-economic variables (Mantzicopoulos, 1997; Mantzicopoulos & Morrison, 1992), and a lack of longitudinal studies (Buntaine & Costenbader, 1997; Dennebaum & Kulberg, 1994; Mantzicopoulos & Morrison, 1992). Studies conducted over the last decade have been markedly improved over their predecessors with regard to methodology; however, there is still considerable variability in methodological quality. For example, in the studies reviewed, there is evidence of a lack of consistency in the composition of comparison groups, with some studies making same grade comparisons, but others making same age comparisons (Mantzicopoulos, 1997; McCoy & Reynolds, 1999; Reynolds, 1992). There is also a lack of adequate matching in control groups (Mantzicopoulos & Morrison, 1992), and a lack of student information prior to retention (McCoy & Reynolds, 1999).

The policy studies included in this review have highlighted another important

methodological issue that needs attention. When examining student outcomes and including students from multiple systems (e.g., classrooms, schools, districts), it is essential that varying system-level effects be considered and accounted for in chosen analyses. Hong and Raudenbush (2005) handled this issue appropriately by using a two-level hierarchical linear model (HLM), which allows investigators to simultaneously examine student-level and school-level variables. Though it would have been desirable to include classroom-level variables in their study, limited data made the addition of this level impossible. Most of the studies reviewed (e.g., Jimerson, 1999; McCoy & Reynolds, 1999), and more specifically, the exploratory studies conducted by Powell (2005) and Porter (2005), did not address school-level variables, which represents a methodological weakness. The present study, designed to function as a follow-up study to the preliminary results provided by Powell (2005), will not only address longitudinal outcomes with a comparison group, but also will account for school-level variables (e.g., existence of a Reading First program) using the statistical method of hierarchical linear modeling (HLM).

#### *Future Directions in Retention Research*

Finally, it should be noted that *all* published retention studies to date have examined retention itself as intervention (Armistead, Kempton, Lynch, & Forhand, 1992; Gottfredson et al., 1994; Jimerson, 2001; Phelps et al., 1992). There is no mention of remediation efforts in conjunction with, or in lieu of, retention. An exception can be found in unpublished exploratory studies by Powell (2005) and Porter (2005); regardless, the literature published thus far has not only failed to describe remediation efforts, but

also has neglected to examine individual components of retention in an attempt to determine exactly which variables are associated with the negative effects. These are perhaps the most significant characteristics of the available literature and the overriding presence of such methodological limitations render the application of published findings to the present student progression plan in the State of Florida inappropriate.

The State of Florida has been clear in its emphasis on remediation. That retention is part of its remediation plan has caused alarm; however, systematic efforts to address academic difficulties are mandated throughout the plan. Academic Improvement Plans, or AIPs, (Florida Department of Education, 2002) were a set of formalized instructional modifications and related goals, both of which were mandated by state law and designed to address specific skill deficits in any child who was not meeting academic benchmarks in a timely manner. The purpose of an AIP was to delineate the specific instructional and environmental variations that, when implemented as designed, were intended to encourage remediation of a student's academic deficiencies. Examples of AIP interventions include 1) time spent with a peer tutor, 2) one-to-one instruction with a reading coach, and 3) extra time (30 minutes per day) devoted to reading instruction. Important features of AIPs were their provision of a direct link to previous assessment as well as their inclusion of objective, measurable goals; both considered critical for effective progress monitoring, which was required for every AIP. Children were to be reevaluated at specific intervals to determine if they were making progress toward specified goals; these goals could also be re-evaluated and revised if necessary. AIPs were available to any struggling student, regardless of age or retention status, as the goal

was to remediate problems before retention becomes necessary. If a child's reading performance was determined to be at Level 1 on the FCAT at the end of third grade and no cause for exemption was found, he was then retained; still, remediation efforts continued and likely intensified (Florida Department of Education, 2002). While state law was quite clear regarding remediation requirements, specific data verifying treatment integrity of AIPs was lacking. Nevertheless, most of the students in this study were required to have AIPs in place following their FCAT-Reading performance in 2003, thus holding negative beliefs about the efficacy of the current student progression plan in Florida is premature as research evaluating the outcomes of the plan remains limited. It should be noted that in July 2006, legislation replaced the AIP with a mandated Progress Monitoring Plan (PMP) that places greater emphasis on the use of data to monitor student performance issues relating to the integrity on PMP; however, similar implementation and integrity issues need to be addressed (Florida Department of Education, 2006).

Regardless, conclusions about the virtues of Florida's retention policy based on past research alone may not be valid in the current educational atmosphere of reform and accountability. The State has chosen a relatively radical response to dramatic changes in federal policy. This is a time of transition, but it affords the opportunity to add to the body of literature in a meaningful way. Retention has been long considered an inadequate intervention strategy for low-achieving students; the emergence of new policies and practices involving retention calls for a new generation of literature investigating this model of "retention plus remediation".

## CHAPTER THREE

### Method

#### *Purpose of the Study*

The purpose of this study was to evaluate the longitudinal effects of third-grade retention practices in the State of Florida as measured by scores on the state mandated standardized test, the Florida Comprehensive Assessment Test - Reading (FCAT-Reading). Retention as practiced in the State of Florida (i.e., within the context of a state mandated remediation program) is largely unstudied. Though exploratory data on the short-term academic outcomes of third grade retention for one cohort exist (Powell, 2005), results were somewhat mixed (i.e., 62% of retained students scored at a Level 2 or higher after the retention year, yet a full 38% failed for the second year in a row) and suggested that a longitudinal examination of data is warranted. This study extended those preliminary findings by examining retained students' performance in fifth grade (i.e., three years post-retention) and comparing their performance to the fifth grade performance of low-achieving third grade students who were promoted to the fourth grade in 2003 through good cause exemptions. This study also examined the relationship between future performance (i.e., three years following retention) on the FCAT-Reading and various student variables (i.e., gender, race/ethnicity, SES, and retention status) and school characteristics (i.e., Reading First status, school size, school socioeconomic

status).

### *Sample*

The initial sample for this study included all students whose reading proficiency was classified as Level 1 (a scaled score between 100 and 258) on the 2003 FCAT-Reading and who, in 2003, were either retained in the third grade or promoted to the fourth grade due to a good cause exemption (N = 34,671). These students were all nested within their respective schools, thus the schools themselves were also included in the sample (N = 1,910). From this sample, student- and school-level exclusion criteria were created. Students who attended a lab school, attended school at home, or had been retained multiple times were excluded. Students who met the inclusion criteria for this study but who had missing data for any of the variables of interest (e.g., gender, race, SES) were excluded. Schools with less than 20 students in the sample (i.e., less than 20 students scoring at Level 1) were excluded.

The retained students included in this study were those who failed to demonstrate reading achievement thought to be necessary in order for them to experience success in the fourth grade. Consequently, those students completed the third grade curriculum for a second time during the 2003-2004 school year, this time according to an individualized intervention plan, at that time referred to as an Academic Improvement Plan (AIP). The students who were promoted due to a good cause exemption, despite scoring at Level 1 on the 2003 FCAT Reading, may have experienced a regular fourth-grade curriculum, depending on their eligibility for an AIP based on lowered performance, during the 2003-2004 school year.

The final sample included 17,087 students in 796 schools, including 12,685 students in 793 schools who were retained and 4,402 students in 710 schools who were promoted through good cause exemptions. The school sample (n = 796) included schools in all 67 Florida public school districts, and of these, 189 (n = 3,624) were classified as Reading First schools and 607 (n = 13,453) were not classified as Reading First schools.

To examine the impact of retention more closely, each student was coded on several characteristics, including race, gender and socioeconomic status. Race distinctions included Caucasian, African American, Asian-Pacific Islander, Hispanic and Alaskan Native/Native American. Socioeconomic status (SES) was categorized by each student's eligibility for Free and Reduced Lunch (FRL), as determined by the income parameters set by the State of Florida. In addition, schools were coded on size (i.e., student population) and socioeconomic status (i.e., proportion of students eligible for FRL), as well as Reading First status, which refers to whether a school received federal funds to assist in the implementation of empirically-based reading programs.

### *Measures*

*Florida Comprehensive Assessment Test.* As it is a dominant factor in the State of Florida's policy relating to the evaluation of the academic progression of its students, this study utilized scores from the Florida Comprehensive Assessment Test – Reading (FCAT-Reading) in order to examine the effectiveness of retention on low achieving third-grade students. According to the Florida Department of Education (2003), the FCAT is a standardized, criterion-referenced test developed by a panel of curriculum specialists. Its intent is to assess student achievement of the higher-order cognitive skills

represented in the Sunshine State Standards (SSS). The FCAT-Reading test reports scores in four areas including: (1) main idea, plot and purpose; (2) words and phrases in context; (3) comparisons of cause/effect; and (4) reference and research. Included in the FCAT are literary passages, and informational passages. Another portion of the FCAT reading section is used normatively, comparing the students of Florida with the rest of the nation. The FCAT contains both multiple-choice and performance questions; the multiple-choice questions are scored by computer, while the performance tasks are hand scored. During the months of February and March of each year, the FCAT is administered to over 1.5 million students across the state. The completed tests are then sealed and sent to the Florida Department of Education.

Scoring of the FCAT is based on item response theory (IRT, Lord & Novick, 1968). IRT theory assumes that student responses to individual questions are directly related to underlying achievement in a given content area. Cronbach's reliability coefficient estimates of scores from the fourth-grade reading portion of the FCAT were reported by total score ( $\alpha = .88$ ), literary text ( $\alpha = .79$ ) and informational text ( $\alpha = .79$ ) (Florida Department of Education, 2002b).

Five levels have been defined for the FCAT reading achievement test, based on scaled scores ranging from 100 to 500. Level 1 represents the lowest level of reading achievement; each subsequent level represents an increasing level of reading proficiency. In 2003, the five levels were defined according to the following scaled scores: Level 1: 100-258, Level 2: 259-283, Level 3: 284-331, Level 4: 332-393, and Level 5: 394-500. In reading, it is assumed that a student who achieves a scaled score categorized as Level 1 is

experiencing little success with the content of the Sunshine State Standards, while students achieving a scaled score in the Level 2 category are assumed to be experiencing limited success with the same content. In comparison, students achieving scores in the range of Levels 3, 4, or 5 are considered proficient readers, to varying degrees (Florida Department of Education, 2003c).

#### *Procedures and Data Preparation*

Upon the approval of the Institutional Review Board, the principal investigator obtained the following data from the Florida Department of Education on students who scored Level 1 on the 2003 FCAT-Reading: retention status, reading scores, reading levels, race, gender, socioeconomic status, good cause promotion status and type, and school attended at the time of examination. This data collection procedure was repeated with the data resulting from the 2005 and 2006 FCAT-Reading administrations to examine the FCAT scores of the students who were either retained in third grade in 2003 or promoted due to a good cause exemption. A second database consisting of school-level data (i.e., school SES and school size) from the intervention year (2003-2004) was collected using the published school indicators from the Florida Department of Education website (Florida Department of Education, 2007b). A third database consisting of schools classified as Reading First schools for the 2003-2004 school year was also obtained from the same source. The resultant three databases (i.e., student-level variables, school size and school SES, and Reading First status) were merged using a data merging procedure in the SAS 9.1 software. The data were then screened and students who did not meet established criteria, who were retained two or more times, or who had missing data on

variables of interest, were excluded. In addition, students in schools having less than 20 students in the sample were excluded.

### *Data Analysis*

Descriptive analyses of the data were computed on all the variables of interest and the results describe the participant pool in terms of gender, race, and SES. To answer each research question, data were subjected to statistical analyses, as described below.

*Research Question 1: For Florida students who scored at Level 1 on the 3<sup>rd</sup> grade FCAT-Reading in 2003 and were retained in third grade during the 2003-2004 school year, what proportion scored at Level 2 or higher on the 5<sup>th</sup> grade FCAT-Reading in 2006?* For students who scored Level 1 and were retained in 2003, the number and percent who then scored Level 1, Level 2, and Level 3 or higher on the fifth grade FCAT-Reading in 2006 were computed.

*Research Question 2: For Florida students who scored at Level 1 on the 3<sup>rd</sup> grade FCAT-Reading in 2003, but were promoted to the fourth grade for the 2003-2004 school year through one of six good cause exemptions, what proportion for each exemption scored at Level 2 or higher on the fifth grade FCAT-Reading in 2005, and the sixth grade FCAT-Reading in 2006?* For students who scored Level 1 in 2003 but were promoted due to good cause exemptions, the number and percent of students who then scored Level 1, Level 2, and Level 3 or higher on the sixth grade FCAT in 2006 were computed for each of the six exemptions.

*Research Question 3: To what extent do differences in performance on the 5<sup>th</sup> grade FCAT-Reading exist between students who scored at Level 1 in 2003 and were*

*retained in third grade during 2003-2004 and those who scored at Level 1 in 2003 but were promoted through one of the two good cause exemptions requiring a demonstration of proficiency?*

*Research Question 4: To what extent are observed differences on the 5<sup>th</sup> grade FCAT-Reading between students who were retained in 2003 and students who were promoted in 2003 through a good cause exemption indicating proficiency moderated by the following student characteristics: gender, race/ethnicity, and socioeconomic status?*

*Research Question 5: To what extent are observed differences in performance on the 5<sup>th</sup> grade FCAT-Reading between students who were retained in 2003 and students who were promoted in 2003 through a good cause exemption indicating proficiency moderated by the following school characteristics: Reading First status, school size, schoolwide student socioeconomic status?* To answer research questions 3, 4, & 5, data for students who scored Level 1 in 2003 were subjected to hierarchical linear modeling (HLM) in order to test for school level effects on the student level predictors of 5<sup>th</sup> grade achievement. HLM is a specialized regression technique that was developed to address issues specific to nested or multi-level datasets (Raudenbush & Bryk, 2002). This analysis allowed comparisons to be made between the performance of retained and promoted students, based on individual FCAT-Reading scores, while accounting for the fact that students are nested within schools. Because students are grouped within schools, it is likely that student performance is systematically related to that of other students in the same school. Given the non-independent nature of the data, utilizing a traditional regression approach could lead to biased regression coefficients and smaller than

expected estimates of standard error. This would lead to a higher risk making a Type I error, i.e., identifying a statistically significant relationship between variables (e.g., student race/ethnicity and performance on the FCAT-Reading) when none exists. The procedures underlying HLM correct for this possibility by conducting separate regressions for each school using the students as the unit of analysis. The resulting coefficients are then employed as dependent measures for the subsequent analyses in which the school is the unit of analysis. This process allows for consideration of the potential influence of level 2 variables (i.e., school size, school SES, Reading First status) on level 1 variables (i.e., student gender, race/ethnicity, SES), as well as on the outcome variable of interest (i.e., 5<sup>th</sup> grade FCAT-Reading score).

HLM Data Preparation. Subsequent to the descriptive analyses outlined earlier, a subset of the final sample was created for analysis using hierarchical linear modeling. The dependent variable, scaled scores on the 5<sup>th</sup> grade FCAT-Reading, was created using scores from the appropriate administration (i.e., 2006 administration for retained students, 2005 administration for students promoted through good cause). Only those students promoted through one of two good cause exemptions indicating proficiency were included in these analyses; students promoted through the remaining four good cause exemptions not indicating proficiency were excluded. Finally, data were screened again using additional variables of interest (i.e., school SES, school size, Reading First status, 5<sup>th</sup> grade FCAT-Reading score) to exclude students with missing data.

As was mentioned earlier, the dependent variable was scaled scores on the 5<sup>th</sup> grade version of the FCAT-Reading, which was administered to retained and promoted

students in different years (i.e., retained students in 2006, promoted students in 2005).

Student-level predictor variables included retention status, gender, race, and SES, while school-level predictors included school size, school SES, and Reading First status. Both Level 1 predictor variables (i.e., retention status, gender, race, and SES) and one Level 2 predictor variable (i.e., Reading First) were dummy coded, as is shown in Table 1.

Table 1

*Dummy Coding for Student- and School-Variables*

Variable	Definition	Dummy code
Retention status (ret)	Whether a student was retained or promoted through good cause in 2003	0 = retained 1 = promoted through good cause
Gender (gen)	Student gender	0 = Female 1 = Male
Race (race)	Student race	0 = Caucasian 1 = Not Caucasian
SES (ses)	Student eligibility for Free or Reduced Lunch (FRL)	0 = Not low SES 1 = Low SES
Reading First status (RF)	Whether a school received funds to implement empirically-based reading programs in 2003-2004	0 = Not Reading First 1 = Reading First
		Centering procedure
School SES (sses)	Continuous variable indicating the proportion of students within a school qualifying for FRL	Grand mean centered
School size (ssize)	Continuous variable indicating school enrollment in 2003-2004	Grand mean centered

Gender and student SES are dichotomous variables, which were dummy coded (see Table 1). Gender was coded 0 = female and 1 = male for gender. The proxy used for SES was eligibility for Free/Reduced Lunch, coded 0 = no and 1 = yes. The variable of race had five categories (i.e., Caucasian, African-American, Hispanic, Asian, Other), and

was dummy coded such that 0 = Caucasian and 1 = all other races.

The remaining Level 2 predictor variables (i.e., school size and school SES) were continuous variables. The means of the overall sample for each variable, or the grand means, were centered at zero. Centering the predictor variables of school SES and school size is an important procedure to ensure meaningful later interpretation of the model, as the intercept is interpreted as the estimated FCAT-Reading score when all coefficients are set to zero. The meaning of zero is clear in the case of variables that are dummy-coded (e.g., female, retained, Caucasian); however, without centering, a school size of zero is not possible. Rescaling the continuous variables of school SES and school size such that their actual grand means are equal to zero enables a more meaningful explanation of the results.

A total of 14,593 students and 792 schools were included in the final subset used in the HLM analyses. The mean student FCAT-Reading score was 261.08 ( $SD = 49.60$ ). Mean values of retention status ( $M = .15$ ;  $SD = .36$ ), as well as student gender ( $M = .57$ ;  $SD = .49$ ), race ( $M = .78$ ;  $SD = .42$ ), and SES ( $M = .85$ ;  $SD = .36$ ) were computed. Analyses were conducted using the intervention year (i.e., 2003-2004), or the academic year directly following the retention decision. School size was defined as the number of enrolled students, school SES was defined as the proportion of students within a school eligible for Free or Reduced Lunch, and Reading First status was determined using the published list of 322 schools that received this identification for the 2003-2004 academic year. Overall, 792 schools in 67 Florida school districts were included in the final subset. Of these, 189 schools were classified as Reading First schools, while 603 were

considered non-Reading First. The number of students in each type of school was 3,180 and 11,413, respectively. The mean school size, prior to grand mean centering was 808.78 (SD = 272.22), with enrollment ranging from 20 to 2,328 students. The mean school SES (i.e., average schoolwide proportion of low SES students across all schools) prior to grand mean centering was .72 low SES (SD = .2028).

Though it is often recommended to rely on a statistical power analysis to determine an appropriate sample size for the number of units at each level (Raudenbush & Liu, 2000), because of the extremely large sample size created by including all Florida students scoring at Level 1 on the FCAT-Reading, such an analysis was deemed unnecessary in this study.

HLM Models. For the HLM analysis, five distinct models, beginning with a model were developed, beginning with the unconditional model, which included no predictors.

Model 1: The equation for the unconditional model is

$$Y_{ij} = \gamma_{00} + \mu_{0j} + r_{ij}$$

where

$Y_{ij}$  is the student-level outcome, or 5<sup>th</sup> grade FCAT-Reading score,

$\gamma_{00}$  is the fixed effect, or the overall intercept,

$\mu_{0j}$  is the random effect of the school and

$r_{ij}$  is the random effect of within school variability.

This model enabled the calculation of the intraclass correlation coefficient (ICC), which indicates the proportion of variance in the dependent variable that is attributable to

between-school factors. Because the ICC was significant, it was determined that the extent of the nesting was such that an HLM was warranted.

Model 2: This model was built to include all student-level variables (i.e., Level-1 predictors), which were added to the unconditional model. The resultant equation is

$$Y_{ij} = \gamma_{00} + \gamma_{01} * \text{retstat} + \gamma_{02} * \text{gender} + \gamma_{03} * \text{race} + \gamma_{04} * \text{SES} + \mu_{0j} + \mu_{1j} * \text{gender} + \mu_{2j} * \text{race} + r_{ij}$$

where

$Y_{ij}$  is the student-level outcome, or 5<sup>th</sup> grade FCAT-Reading score,

$\gamma_{00}$  is the fixed effect, or the overall intercept,

$\gamma_{01} * \text{retstat}$  is the fixed effect of retention status,

$\gamma_{02} * \text{gender}$  is the fixed effect of gender,

$\gamma_{03} * \text{race}$  is the fixed effect of race,

$\gamma_{04} * \text{SES}$  is the fixed effect of SES,

$\mu_{0j}$  is the random effect of the school,

$\mu_{1j} * \text{gender}$  is the random effect of gender,

$\mu_{2j} * \text{race}$  is the random effect of race, and

$r_{ij}$  is the random effect of within school variability.

Each student-level predictor (student retention status, gender, race, and SES ) was entered as a fixed effect. In addition, based on theory, student variables were allowed to covary in this model, meaning that the effects of student retention status, gender, race, and SES may vary across schools. The purpose of the HLM Level-1 model was to examine the effect of student-related factors on the dependent (outcome) measure of 5<sup>th</sup>-grade reading

performance as measured by the FCAT.

Model 3: Next, both Level-1 predictors (i.e., student-level variables) and Level-2 predictors (school-level variables) were added to the model to yield Model 3. The equation is expressed as follows:

$$Y_{ij} = \gamma_{00} + \gamma_{01} * \text{retstat} + \gamma_{02} * \text{gender} + \gamma_{03} * \text{race} + \gamma_{04} * \text{SES} + \gamma_{05} * \text{schoolses} + \gamma_{06} * \text{schoolpop} + \gamma_{07} * \text{readingfirst} + \mu_{0j} + \mu_{1j} * \text{gender} + \mu_{2j} * \text{race} + r_i$$

where

$Y_{ij}$  is the student-level outcome, or 5<sup>th</sup> grade FCAT-Reading score,

$\gamma_{00}$  is the fixed effect, or the overall intercept,

$\gamma_{01} * \text{retstat}$  is the fixed effect of retention status,

$\gamma_{02} * \text{gender}$  is the fixed effect of gender,

$\gamma_{03} * \text{race}$  is the fixed effect of race,

$\gamma_{04} * \text{SES}$  is the fixed effect of SES,

$\gamma_{05} * \text{schoolses}$  is the fixed effect of school SES,

$\gamma_{06} * \text{schoolpop}$  is the fixed effect of school size,

$\gamma_{07} * \text{readingfirst}$  is the fixed effect of Reading First status,

$\mu_{0j}$  is the random effect of the school,

$\mu_{1j} * \text{gender}$  is the random effect of gender,

$\mu_{2j} * \text{race}$  is the random effect of race, and

$r_{ij}$  is the random effect of within school variability.

All student- and school-level predictors were entered as fixed effects, and student-level predictors were allowed to covary. The purpose of the Model 3 was to examine the effect

of the student- and school-level variables (as fixed effects only) on 5<sup>th</sup>-grade reading performance.

Model 4: Next, a full model that included the Level-1 and Level-2 predictor variables as fixed effects as well as all possible interactions between Level-1 and Level-2 predictor variables was created. The purpose of this expanded HLM Level-2 model was to examine the effects of the student- and school-level variables as well as possible interactions of these variables on the outcome measure. Model fit was assessed using the AIC fit index.

Model 5: This final model was built based on examination of the expanded Model 4 and elimination, as appropriate, of the interaction effects that were non-significant.

In order to assess the extent to which each of these models were accounting for explained variability between students and between schools, pseudo  $R^2$  values were calculated for each model relative to the variability between students (within school variance) and between schools in the unconditional model (Model 1). The following equations were used to compute the within school variance

$$\text{Psuedo } R_1^2 = \frac{\mathbf{s}_{Unconditional}^2 - \mathbf{s}_{Model}^2}{\mathbf{s}_{Unconditional}^2}$$

and the between school variance

$$\text{Psuedo } R_2^2 = \frac{\mathbf{t}_{00Unconditional} - \mathbf{t}_{00Model}}{\mathbf{t}_{00Unconditional}}$$

The pseudo  $R^2$  indicates the proportion of explainable variance that a particular model explains, and uses the unconditional model as a comparison. Model fit was assessed for

each model using Akaike's Information Criterion (AIC, Akaike, 1974). Estimation procedures utilizing Empirical Bayes estimates was employed at each level of the model to ensure the data were normally distributed with sufficiently equal variance. An outlier was identified; however, removal of the outlier did not substantially change the parameter estimates of the predictors, so was not permanently removed from the dataset. The alpha significance level for all analyses was established at .05.

## CHAPTER FOUR

### Results

As was stated previously, this study was intended to examine the reading performance of students, considering the effects of retention in third grade as well as student demographic characteristics (gender, SES, race/ethnicity) and school variables (school size, Reading First status, and school-wide SES) . In addition, the future performances of low-performing students who were promoted to the fourth grade through good cause exemptions were examined. This chapter begins with a description of the sample and results of descriptive analyses for the first two research questions. Then, the sub-sample used for research questions 3-5 which employed hierarchical linear modeling (HLM) to examine student- and school-level factors associated with student performance in reading subsequent to the retention year is described, and the results of the analyses relating to those questions are reported. The data were analyzed using SAS Version 9.1. An alpha level of .05 was employed for all statistical tests.

#### *Description of Sample*

The final sample for this study consisted of 17,087 students in 796 schools who scored at Level 1 on the 2003 third-grade Florida Comprehensive Assessment Test-Reading (FCAT-Reading) and were either retained in the third grade or promoted to fourth grade through a good cause exemption for the 2003-2004 academic year. Race and

socioeconomic status were defined in this study according to the data made available by the Florida Department of Education (FLDOE). The FLDOE defines race using six categories (Caucasian, African-American, Hispanic, Asian, Native American, and Mixed Race); however, the categories of Asian, Native American and Mixed Race were collapsed in this study due to small sample sizes and are included in the category labeled “Other”. Florida defines SES by a student’s eligibility status for receiving lunch at free or reduced prices; therefore, in this study, students who were eligible to receive free or reduced lunch (FRL) were coded as “Low SES” and those who were not eligible were coded as “Not low SES”.

A breakdown of the final sample in terms of retention/promotion status at the end of third grade in 2003 is reported in Table 2.

Table 2

*Students Scoring Level 1 on the 2003 FCAT-Reading and Retention/Promotion Status by Gender, Race/Ethnicity, and SES*

	2003 FCAT- Reading <u>Level 1</u>		<u>Retained in Third Grade</u>		<u>Promoted due to Good Cause Exemption</u>	
	n	%	n	%	n	%
<i>Race/Ethnicity</i>						
Caucasian	4002	23	2767	22	1235	28
African American	6628	39	5268	42	1360	31
Hispanic	6008	35	4339	34	1669	38
Other	449	3	311	2	138	3
<i>Gender</i>						
Male	9909	58	7353	58	2556	58
Female	7178	42	5332	42	1846	42
<i>SES</i>						
Low	14487	85	10877	86	3610	82
Not low	2600	15	1808	14	792	18
Total	17087		12685	74	4402	26

### *Research Questions*

*Question 1. For Florida students who scored at Level 1 on the 3<sup>rd</sup> grade FCAT-Reading in 2003 and were retained in third grade during the 2003-2004 school year, what proportion scored at Level 2 or higher on the 5<sup>th</sup> grade FCAT-Reading in 2006?*

The number and percent of 2003 third grade retainees performing at Levels 1-5 on the 5<sup>th</sup> grade FCAT-Reading in 2006 were computed. After examination of the data, Levels 3, 4, and 5 were collapsed due to small cell sizes. It should be noted that students performing at Levels 3 through 5 are considered to be proficient readers by the FLDOE,

while students performing at Levels 1 and 2 are experiencing “little” and “limited” success with the Sunshine State Standards, respectively. The resultant performance levels for retained students are reported in Table 3. As is shown, 60% (n = 7,687) of the third grade retainees who took the 5<sup>th</sup> grade FCAT-Reading in 2006 scored at Level 2 or higher. More specifically, 28% (n = 3,571) of the retained students scored at Level 2, while 32% (n = 4,116) scored in the range of proficiency, or Levels 3-5. Forty percent (n = 4,995) of the retained students scored at Level 1.

Table 3

*Performance of Retained Students on the 5<sup>th</sup> grade FCAT-Reading in 2006 by Level*

Reading Level	Number	Percent
1	4995	40
2	3571	28
3-5	4116	32
Total	12685	100

*Question 2. For Florida students who scored at Level 1 on the 3<sup>rd</sup> grade FCAT-Reading in 2003, but were promoted to the fourth grade for the 2003-2004 school year through one of six good cause exemptions, what proportion for each exemption scored at Level 2 or higher on the fifth grade FCAT-Reading in 2005, and the sixth grade FCAT-Reading in 2006?*

Altogether, 26% (n = 4,402) of the third-grade students who scored at Level 1 on

the 2003 FCAT-Reading were promoted to the fourth grade due to a good cause exemption. Table 4 displays a breakdown of this group of students by type of exemption.

Table 4

*Number and Percent of Students Promoted to 4<sup>th</sup> Grade due to Good Cause Exemption by Type of Exemption*

Exemptions	N	%
1. Proficiency demonstrated through alternative standardized reading assessment	1157	26
2. Proficiency demonstrated through student portfolio	652	15
3. Limited English Proficient (LEP)	1059	24
4. Did not participate due to Individualized Educational Plan (IEP)	46	1
5. Previous retention (1 year) + IEP	1268	29
6. Previous retentions (2 years)	220	5
Total	4402	100

*Performance on the 5<sup>th</sup> grade FCAT-Reading.* Student performance on the 5<sup>th</sup> grade FCAT-Reading test in 2005 for third grade students who scored at Level 1 in 2003 but who were promoted to fourth grade for 2003-04 through a good cause exemption were examined by computing the number and percent of students who scored at each

reading level. After examination of the data, Levels 3, 4, and 5 were collapsed due to small cell sizes. The resultant performance levels are reported in Table 5. As is shown, of the 4,402 third grade students promoted through a good cause exemption in 2003 who went on to take the 5<sup>th</sup> grade FCAT-Reading in 2005, 41% (n = 1,805) scored at Level 2 or higher. More specifically, 22% (n = 958) of the promoted students scored at Level 2, while 19% (n = 847) scored at Levels 3 – 5. Fifty-nine percent (n = 2,697) of the promoted students scored at Level 1. Thus, nearly two-thirds of students continued to struggle with grade level content after being promoted through good cause exemptions after third grade. A more detailed look at the performance of students promoted through good cause by type of exemption is provided in Table 6.

Table 5

*Performance of Students Promoted through Good Cause on the 5<sup>th</sup> grade FCAT-Reading in 2005 by Level*

Reading Level	Number	Percent
1	2697	59
2	958	22
3-5	847	19
Total	4402	100

Table 6

*Level 1 Promoted Students' Performance on the 2005 5<sup>th</sup> grade FCAT-Reading by Good Cause Exemption*

Type of Exemption	<u>Level 1</u>		<u>Level 2</u>		<u>Levels 3-5</u>	
	n	%	n	%	n	%
1. Proficiency demonstrated through alternative standardized reading assessment	403	35	376	32	378	33
2. Proficiency demonstrated through student portfolio	296	45	206	32	150	23
3. Limited English Proficient (LEP)	623	59	207	20	229	21
4. Did not participate due to Individual Education Plan (IEP)	40	87	4	9	2	4
5. Previous retention (1 year) + IEP	1063	84	130	10	75	6
6. Previous retentions (2 years)	172	78	35	16	13	6
<b>Total</b>	<b>2597</b>	<b>59</b>	<b>958</b>	<b>22</b>	<b>847</b>	<b>19</b>

Looking more closely at only the group of students (n = 1,809) who demonstrated proficiency through an alternative standardized reading assessment (Exemption 1) or through student portfolio (Exemption 2), the data reveal a different pattern of performance than that of students who did not take the FCAT-Reading due to an

Individual Education Plan (IEP), or who were promoted due to previous retentions (Exemptions 4-6). Specifically, the percentages scoring at Level 2 or Levels 3 - 5 on the 5<sup>th</sup> grade FCAT-Reading in 2005 were quite high for those who were promoted due to Exemption 1 (65%, n = 774) and Exemption 2 (55%, n = 356) when examining this smaller subset of students. In contrast, fewer students (4-16%) promoted due to Exemptions 4-6 scored at Level 2 or Levels 3-5 in 2005.

In addition, 2005 data for students promoted due to Limited English Proficiency with less than two years of English as a Second Language (ESOL) instruction (Exemption 3) indicate a trend similar to that for students who demonstrated proficiency during the 2002-2003 school year. Specifically, almost half, or 41% (n = 436) of LEP students taking the 5<sup>th</sup> grade FCAT-Reading in 2005 scored at Level 2 or higher.

*Performance on the 6<sup>th</sup> grade FCAT-Reading.* Student performance on the 6<sup>th</sup> grade FCAT-Reading test in 2006 for third grade students who scored at Level 1 in 2003 but who were promoted to fourth grade for 2003-04 through a good cause exemption were examined by computing the number and percent of students who scored at each reading level. After examination of the data, Levels 3, 4, and 5 were collapsed due to small cell sizes. The resultant performance levels are reported in Table 7. As is shown, of the 4,402 third grade students promoted through a good cause exemption who went on to take the 6<sup>th</sup> grade FCAT-Reading in 2006, 44% (n = 1,942) scored at Level 2 or higher. More specifically, 23% (n = 1,028) of the promoted students scored at Level 2, while 21% (n = 914) scored at Levels 3-5. Fifty-six percent (n = 2,460) of the promoted students scored at Level 1. Table 8 displays performance by type of exemption.

Table 7

*Performance of Students Promoted through Good Cause on the 6<sup>th</sup> grade FCAT-Reading in 2006 by Level*

Reading Level	Number	Percent
1	2460	56
2	1028	23
3-5	914	21
Total	4402	100

Table 8

*Level 1 Promoted Students' Performance on the 2006 6<sup>th</sup> grade FCAT-Reading by Good Cause Exemption*

Type of Exemption	<u>Level 1</u>		<u>Level 2</u>		<u>Levels 3-5</u>	
	n	%	n	%	n	%
1. Proficiency demonstrated through alternative standardized reading assessment	377	33	408	35	372	32
2. Proficiency demonstrated through student portfolio	277	42	221	34	154	24
3. Limited English Proficient (LEP)	536	51	222	21	301	28
4. Did not participate due to Individual Education Plan (IEP)	40	87	4	9	2	4
5. Previous retention (1 year) + IEP	1052	83	148	12	68	5
6. Previous retentions (2 years)	178	81	25	11	17	8
<b>Total</b>	<b>2460</b>	<b>56</b>	<b>1028</b>	<b>23</b>	<b>914</b>	<b>21</b>

Looking more closely at only the group of students (n = 1,809) who demonstrated proficiency in 2003 through an alternative standardized reading assessment (Exemption 1) or through student portfolio (Exemption 2), the data reveal a different pattern of performance than that for students who did not take the FCAT-Reading due to an

Individual Education Plan (IEP), or who were promoted due to previous retentions (Exemptions 4-6). Specifically, the percentages of students scoring at Level 2 or Levels 3-5 on the 6<sup>th</sup> grade FCAT-Reading in 2006 were quite high for those who were promoted due to Exemption 1 (67%, n = 780) and Exemption 2 (58%, n = 375) when examining this smaller subset of students. In contrast, fewer students (4%-12%) promoted due to Exemptions 4-6 scored at Level 2 or Levels 3-5 in 2006.

In addition, 2006 data for students promoted due to Limited English Proficiency with less than two years of English as a Second Language (ESOL) instruction (Exemption 3) indicate a trend similar to that for students who demonstrated proficiency during the 2002-2003 school year. Specifically, almost half, or 49% (n = 523) of LEP students taking the 6<sup>th</sup> grade FCAT-Reading in 2006 scored at Level 2 or higher.

*Question 3. To what extent do differences in performance on the 5<sup>th</sup> grade FCAT-Reading exist between students who scored at Level 1 in 2003 and were retained in third grade during 2003-2004 and those who scored at Level 1 in 2003 but were promoted through one of the two good cause exemptions requiring a demonstration of proficiency?*

*Question 4. To what extent are observed differences on the 5<sup>th</sup> grade FCAT-Reading between students who were retained in 2003 and students who were promoted in 2003 through a good cause exemption indicating proficiency moderated by the following student characteristics: gender, race/ethnicity, and socioeconomic status?*

*Question 5. To what extent are observed differences in performance on the 5<sup>th</sup> grade FCAT-Reading between students who were retained in 2003 and students who were*

*promoted in 2003 through a good cause exemption indicating proficiency moderated by the following school characteristics: Reading First status, school size, schoolwide student socioeconomic status?*

To answer these questions, data were analyzed at both the student and school levels using hierarchical linear modeling (HLM). Data collection and screening procedures were described earlier. The subset included Level 1 students who were retained in third grade in 2003, as well as Level 1 students who were promoted through a good cause exemption indicating proficiency (N = 14,593). Schools with at least 20 students in the final subset were included, resulting in a total of 792 schools in all 67 Florida school districts. Of these, 189 schools with a total of 3,180 students were classified as Reading First, while 603 schools with 11,413 students were classified as non-Reading First. The mean school size was 808.78 (SD = 272.22, range = 20-2,328), and the mean school SES (i.e., average schoolwide proportion of low SES students across all schools) was .72 low SES (SD = .2028).

Prior to subjecting the data to a hierarchical linear model (HLM), data were screened to test for violations of assumptions (i.e., normality, collinearity, homoscedacity). Each student level variable was examined separately, and results indicated that the variables of retention status and SES were both skewed, or non-normal, likely due to the sample of low-achieving students. These variable distributions were also leptokurtic, with positive kurtosis values (1.78 and 1.74, respectively). The variables of gender and race were less skewed, meaning that the distributions were closer to normal. The school-level variables of Reading First and SES were also examined individually for

normality, and both appeared to approach normality. School size was extremely leptokurtic, with a kurtosis value of 4.66. Descriptive statistics of student and school level variables included in the HLM (prior to grand mean centering), including means, standard deviations, skewness, and kurtosis, appear in Table 9.

Table 9

*Means and Standard Deviations for Student- and School-Level Variables*

Variable	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
Student level ( <i>N</i> = 14,593)				
FCAT-Reading	261.08	49.60	-0.91	1.43
Retention status	.15	.36	1.94	1.78
Gender	.57	.49	-0.30	-1.91
Race	.78	.42	-1.34	-0.19
Student SES	.85	.36	-1.93	1.74
School level ( <i>N</i> = 792)				
FCAT-Reading	261.01	16.34	-0.91	1.43
School SES 2003-2004 (proportion of students qualifying for FRL)	.7199	.2028	-0.60	-0.40
School size 2003-2004	808.78	272.22	1.33	4.66
Reading First status	.24	.41	1.37	-0.13

*Note:* Retention status coded 0 for retained and 1 for good cause promoted; gender is coded 0 for female and 1 for male; race is coded 0 for Caucasian and 1 for all other races; student SES is coded 0 if not eligible for free or reduced lunch and 1 if eligible; School SES is the grand mean centered proportion of students in the study at a school that are eligible for free or reduced lunch; schoolpop is the grand mean centered variable for school population; Reading First is coded 0 for a control school and 1 for a program school; level-1 sample size equals 14,593; level-2 sample size equals 792.

Student- and school-level variables as well as the outcome variable, FCAT-Reading scores, were examined through separate series of bivariate correlations to determine the extent of any existing multicollinearity. Although most relationships among student level variables were significant, they were also relatively weak, with the relationship between race and SES demonstrating the strongest relationship ( $r = .34, p < .0001$ ). With regard to the school level variables, all relationships were significant, with only the relationship between school SES and school size demonstrating a moderate level of strength ( $r = .40, p < .0001$ ). Because of the relative weakness of the relationships between both the student- and school-level variables the decision was made that each of these variables was an appropriate predictor variable for these models. The results of all bivariate correlations are displayed in Table 10.

Table 10

*Pearson Product Moment Correlations for Student- and School-Level Variables*

Level 1 (N = 14,593)					
	FCAT-Reading	Retention status	Gender	Race	SES
FCAT-Reading	1.00				
Retention status	0.03**	1.00			
Gender	-0.11***	0.03**	1.00		
Race	-0.02*	-0.04**	-0.02**	1.00	
SES	-0.05***	-0.09**	-0.03**	0.34**	1.00
Level 2 (J = 792)					
	School SES	School size	Reading First status		
School SES	1.00				
School size	-0.40***	1.00			
Reading First status	0.15***	-0.17***	1.00		

\*  $p < .001$ , \*\*  $p < .001$ , \*\*\*  $p < .0001$ .

*Unconditional model.* To first estimate the proportion of variation in 5<sup>th</sup> grade FCAT-Reading scores within and between schools, data were subjected to an unconditional means model

$$Y_{ij} = \mu_{00} + \mu_{0j} + r_{ij}$$

The unconditional model in HLM can be viewed as a one-way ANOVA with random effects, in that no predictor or moderator variables are included in the analysis. Rather, the model has one fixed effect (i.e., the overall mean of FCAT-Reading scores ( $\mu_{00}$ ) and two variance components [i.e., variation between school means ( $\mu_{0j}$ ) and variation among students within schools ( $r_{ij}$ )]. This model served to partition the variance in FCAT-Reading scores into its within- and between-school components. In other words, it provided an estimation of the amount of variance that is attributed to student- and school-

level variables, or a level of dependence of the data on group effects. Results from this model showed that

$$\text{FCAT-Reading (5}^{\text{th}} \text{ grade)} = 261.09 + 102.52 * \text{between-school variance} + (2357.11 * \text{within-school variance}).$$

The average school-level FCAT-Reading score was 261.09 [ $t = 470.80, p < .0001$ ] in this sample of schools. The intra-class correlation coefficient (ICC;  $\hat{r} = .04$ ) was calculated and indicated that while the total proportion of variance in FCAT-Reading scores that was between schools (about 4%) was very small relative to within-school variance, differences among schools in their mean outcomes were significant [ $Z = 8.64, p < .0001$ ]. As fit indices were used later to guide selection among alternative models, it is of note that the Akaike's Information Criterion (AIC; Akaike, 1974) for the unconditional model was 155181.3.

Given the significant variation of FCAT-Reading scores between schools, the HLM procedure, which accounts for these between-school differences, was warranted. Thus, student variables (viz., retention status, gender, race and SES) and school level variables (viz., school SES, school size, Reading First status) were entered into the model as predictor variables, first separately (i.e., Level 1 model, Level 2 model), then together (i.e., Full model, Full model with interactions) (Raudenbush & Bryk, 2002). Results of the models in terms of fixed and random effects, variance components and pseudo  $R^2$  are reported in Table 11.

Table 11

*HLM Means as Outcomes Models: Selected School Characteristics as Predictors of Variation Between Schools in Performance on 5<sup>th</sup> Grade FCAT-Reading*

	Model 1 Unconditional Model			Model 2 Level 1 Model			Model 3 Full Model			Model 5 Full Model + 1 interaction		
<i>Fixed effect</i>	PE	SE	<i>t</i>	PE	SE	<i>t</i>	PE	SE	<i>t</i>	PE	SE	<i>t</i>
Intercept	261.09	.55	470.80 ***	273.28	1.34	203.28 ***	272.99	1.44	190.06 ***	273.87	1.44	189.30 ***
ret				3.43	1.16	2.95 **	3.12	1.16	2.69 **	-2.62	1.68	-1.56
gender				-11.39	.85	-13.39 ***	-11.33	.85	-13.27 ***	-12.99	0.92	-14.06 ***
race				-1.60	1.10	-1.45	-1.80	1.13	-1.60	-1.78	1.13	-1.58
ses				-5.86	1.23	-4.78 ***	-4.78	1.25	-3.82 **	-4.72	1.25	-3.77 **
sses							-0.02	0.03	-0.65	-0.02	0.03	-0.67
ssize							0.01	0.00	6.10 ***	0.01	0.00	6.12 ***
RF							-0.98	1.28	-0.77	-0.95	1.27	-0.74
<i>Interactions</i>												
ret*gen										10.69	2.27	4.71 ***
<i>Random Effects</i>												
				PE	SE	Z	PE	SE	Z	PE	SE	Z
gen				37.27	17.68	2.11 *	41.35	17.60	2.35	42.65	17.76	2.40 **
race				5.51	19.09	0.29	5.23	18.08	0.29	4.12	18.01	0.23

Tau00	102.52	11.87	8.64***	81.11	18.38	4.41***	65.36	17.35	3.77***	65.02	17.36	3.75***
Sigma <sup>2</sup>	2357.11	28.31	83.25***	2312.35	28.09	82.33***	2312.77	28.08	82.36***	2309.56	28.04	82.37***
Pseudo R <sup>2</sup>				R <sup>2</sup> within	R <sup>2</sup> between		R <sup>2</sup> within	R <sup>2</sup> between		R <sup>2</sup> within	R <sup>2</sup> between	
				.02	.21		.02	.36		.02	.37	

\* p < .05; \*\* p < .01; \*\*\* p < .0001

*Note:* PE = parameter estimate; SE = standard error; “ret” is retention status coded 0 for retained and 1 for good cause promoted; “gen” is gender is coded 0 for female and 1 for male; “race” is coded 0 for Caucasian and 1 for all other races; student “ses” is coded 0 if not eligible for free or reduced lunch and 1 if eligible; “sses” is the grand mean centered proportion of students in the study at a school that are eligible for free or reduced lunch; “ssize” is the grand mean centered variable for school population; “RF” is Reading First coded 0 for a control school and 1 for a program school; level-1 sample size equals 14593; level-2 sample size equals 792; and the intra-class correlation (ICC) derived from the unconditional model equals .04.

*Model 2 – Level 1 (student) variables.* This model included the Level 1 student variables of retention status, gender, race, and SES. The equation is given as

$$Y_{ij} = \gamma_{00} + \gamma_{01} * \text{retstat} + \gamma_{02} * \text{gender} + \gamma_{03} * \text{race} + \gamma_{04} * \text{SES} + \mu_{0j} + r_{ij}$$

Because student-level predictors were entered as both fixed and random effects, not only did the model stipulate that a student's FCAT-Reading score is related to these student-level variables, but also that these relationships can vary across schools. The results of this model demonstrated that the covariance parameters for retention status and SES were estimated to be zero, suggesting that the relationship between 5<sup>th</sup> grade FCAT-Reading and these variables did not vary by school in this sample. Thus, the decision was made to eliminate retention status and SES as covariates in this and all subsequent models. Within the Level 1 model, fixed effects of retention status (3.43;  $t = 2.95$ ,  $p < .01$ ), gender (-11.39;  $t = -13.39$ ,  $p < .0001$ ) and SES (-5.86 ( $t = -4.78$ ,  $p < .0001$ ), were demonstrated to be significant predictors of FCAT-Reading performance. Gender (37.27;  $Z = 2.11$ ,  $p < .05$ ) was a significant covariate, indicating that whether a student was male or female had varying effects on FCAT-Reading performance, depending on which school he or she attended. Model fit was assessed and the AIC (154936.40) indicated that this model fit the data slightly better than the unconditional model. However, even after including these specific student variables, this model explained only 2% of the explainable variance within schools ( $\sigma = 2312.35$ ,  $p < .0001$ ); considerable additional explainable variation in FCAT-Reading scores within schools remains.

*Model 3 - Full model.* Because the goal of this study was to examine the impact of student-level variables on students' predicted FCAT-Reading performance while

accounting for key differences in the school that they attended, the next step involved entering school-level variables into the model.

$$Y_{ij} = \beta_{00} + \beta_{01} * \text{retstat} + \beta_{02} * \text{gender} + \beta_{03} * \text{race} + \beta_{04} * \text{SES} + \beta_{05} * \text{schoolses} + \beta_{06} * \text{schoolpop} + \beta_{07} * \text{readingfirst} + \mu_0 + r_i$$

The student variables of gender and race were again allowed to covary by school in this model. The fit results indicated four significant main effects, including retention status (3.11;  $t = 2.69$ ,  $p < .01$ ), gender (-11.33;  $t = -13.27$ ,  $p < .0001$ ), SES (-4.78;  $t = -3.82$ ,  $p < .001$ ), and school size (-0.01;  $t = 6.10$ ,  $p < .0001$ ). Gender was again a significant covariate (41.35;  $Z = 2.35$ ,  $p < .01$ ). The fit of this model was slightly improved over the previous model (AIC – 154903.30).

*Model 4 - Full model with interactions.* The full model included all student- and school-level variables as fixed effects, and took into account random effects of the student variables gender and race. In addition, a total of 21 interaction effects between all variables were examined, with only one interaction (i.e., retention status x gender) demonstrating statistical significance (10.69,  $t = 4.68$ ,  $p < .0001$ ). For full detailed results of this model, see Appendix A. The AIC for Model 4 was larger than that for Model 3 (154906.60), suggesting that the inclusion of all possible interactions was not a good fit for the data, and was possibly masking the significance of some main effects. With this information, the fifth, and final, model was built using all student and school predictors as fixed effects, the student variables of gender and race as random effects, and the interaction between retention status and gender.

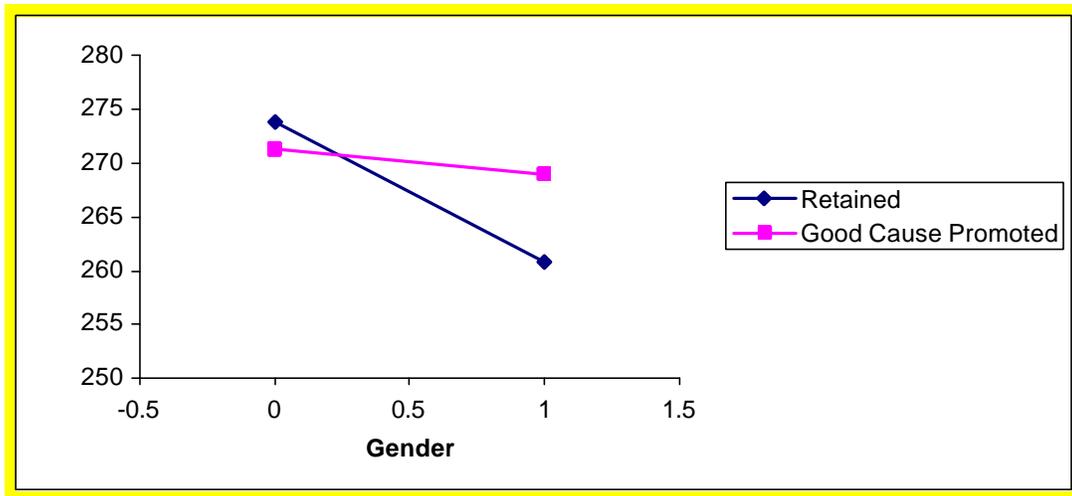
*Model 5 – Full Model with one significant interaction.* Results from this model showed that

$$\text{FCAT-Reading (5}^{\text{th}} \text{ grade)} = 273.87 + (-12.99*\text{female}) + (-4.72*\text{not low SES}) + 0.01*\text{schoolpop} + 10.69*\text{retained*female} + 42.65*\text{gender variance} + 65.02*\text{between-schools variance} + 2309.56*\text{within-schools variance}.$$

This model indicated a significant interaction effect between retention status and gender (10.69;  $t = 4.71, p < .0001$ ). Thus, the slopes for these variables differ depending on whether a student was retained or promoted. A graph of the interaction effect is shown in Figure 1. Specifically, being male and promoted through a good cause exemption resulted in an average increase of 8.04 points on the FCAT-Reading when compared to being male and retained. Alternatively, being retained or promoted through an exemption resulted in more similar outcomes among female students (273.87 and 271.243, respectively). As an alternative perspective on this interaction effect, it appears that retention status affected female and male students much differently, with female retained students demonstrating higher estimated scores on the FCAT-Reading (273.87), and male retained students not faring as well (260.881). Alternatively, being promoted through a good cause exemption indicating proficiency seemed to result in similar outcomes for both males (268.942) and females (271.243), that is, there was a smaller difference in mean reading scores between males and females who were promoted through a good cause exemption indicating proficiency.

Figure 1

*Retention Status X Gender Effects on 5<sup>th</sup> grade FCAT-Reading Score*



There was a significant main effect for student SES, indicating that when all other variables are average or zero values, student SES is associated with FCAT-Reading, such that the average score for students who do not qualify for Free or Reduced Lunch (273.87) is significantly higher than that of students who qualify for Free or Reduced Lunch (269.15). Another main effect demonstrating significance was school size, indicating that students in larger schools scored higher on the FCAT-Reading than students in smaller schools (.01;  $t = 6.12, p < .0001$ ).

The AIC value is the lowest observed (154877.70), indicating that of the models compared, this model represents the best fit for the data. Importantly, the remaining explainable between-school variance not accounted for in this model (65.02) is significant ( $Z = 3.75, p < .0001$ ); however, it is considerably smaller than that in the unconditional model (102.52). This decrease in between-school variance indicates that the predictors and interaction effects in the full model explain 37% of the explainable

variation in FCAT-Reading scores between schools. Because the initial amount of between-school variation was small (see unconditional model) compared to the amount of within-school variation, this 37% must be interpreted accordingly. With regard to within-school variance, the predictors and interaction effects in this model explained only 2% of the explainable variation in FCAT-Reading scores within schools. The remaining explainable within-school variance remains high (2290.13), suggesting that there are additional student-level factors that might explain the variation in school means.

## CHAPTER FIVE

### Discussion

The pupil progression plan for the State of Florida mandates retention in the third grade for those students who fail to demonstrate adequate reading skills as measured by the statewide reading achievement test, the Florida Comprehensive Assessment Test - Reading (FCAT-Reading). Students who do not meet this criterion, but who qualify for a good cause exemption [e.g., demonstration of proficiency through a portfolio or another other norm-referenced achievement test, previous grade retention(s)] can be promoted to the fourth grade nevertheless. The present study investigated the longitudinal academic outcomes of the practices of retention and promotion of low-achieving third grade students within the context of statewide performance assessment and state-mandated remediation efforts.

The educational context and climate in Florida continues to be in a state of flux, as new policies and procedures have been perceived by many educators in the state as being established almost continuously since the enactment of the No Child Left Behind (NCLB) federal legislation and subsequent state legislation. The implementation of policy changes has created an educational climate that is experienced by many educators, as well as students and families, to be characterized by pressure and uncertainty at every level. Legislators and policy-makers are being asked to apply a major piece of federal

legislation to the educational system in the State of Florida. As a result of state subsequent state legislation tying the FCAT to high-stakes decisions such as retention and grades for schools, administrators are dealing with very large numbers of third-grade students at a time when a recently passed constitutional amendment limits the size of classroom, meaning that extra classes have had to be created to accommodate these extra students. In addition, they are also faced with the possibility of losing access to funding on which they have counted in the past. Teachers are impacted by the controversy sparked by pressure that exists to “teach to the test.” Finally, students as young as kindergarteners are aware of the FCAT-Reading and the ramifications they will face if they “fail” the test; many students experience considerable anxiety about being left behind in the third grade. The anxiety that has impacted the entire state has resulted in large part from the recency of the legislative changes, lack of familiarity with or uncertainty about the procedures, and the rapidity with which educators have been compelled to comply. This climate of stress may affect students’ scores in these early years; however, it is possible that time will ease the impact of some or all of these factors, and these changes in Florida’s educational climate may impact the academic outcomes of students who are retained due to inadequate performance on the FCAT-Reading. In fact, the percentages of third grade students being retained have declined in more recent years (i.e., from 14.6% in 2003 to 6.8% in 2006). Nevertheless, this is a situation that should continue to be monitored closely.

### *Summary of Findings*

Findings of the present study indicated that, of the students retained in third grade in 2003, 60% scored at Levels 2 or higher on the 5<sup>th</sup> grade FCAT-Reading (i.e., “passed

the FCAT-Reading”) three years later in 2006. This finding is encouraging, as it suggests that retention was associated with positive academic outcomes; after repeating the third grade curriculum and, at least in terms of mandate by policy, receiving intensive academic supports, almost two-thirds of the retained students were successful in achieving a passing FCAT-Reading score as defined by state standards.

Upon closer examination, however, it is noted that only about one-third (32%) of all retained students were considered to be proficient readers according to the FLDOE (i.e., they scored at Levels 3-5) at the end of fifth grade. This percentage is quite small, less than one half the rate when compared to *all* fifth-grade students taking the 5<sup>th</sup> grade FCAT-Reading in 2006 (i.e., 84% scored Levels 2-5, 68% scored Levels 3-5; Florida Department of Education, 2007c); on the other hand, it also represents considerable improvement within this retained group of students when compared to their earlier performance (i.e., all had scored at Level 1 in 2003).

These findings also illustrate the lack of consistency in Florida's definitions of "adequate reading skills", which makes the interpretation of such seemingly contradictory results difficult. Students who score at Level 2 on the FCAT-Reading are demonstrating only "limited success" with the curriculum; however, according to the state pupil progression plan, these same students are eligible for promotion to the fourth grade. As such, they are considered to be "adequate" to be promoted, but "inadequate" as readers. According to state legislation and for the purposes of this study, Levels 2-5 define success; however, it is important to keep in mind that many of the students who have been promoted actually have limited reading skills. It should also be noted that federal and state legislation have targeted different goals with regards to this issue, with the

federal government being primarily concerned that students in Florida schools are making Adequate Yearly Progress (AYP), or increasing the number of proficient readers (defined by the State of Florida as students scoring Levels 3-5 on the FCAT).

Alternatively, local and state policies focus on decreasing the number of students who score at Level 1 on the FCAT and face mandatory retention. Thus, decreasing retention rates (the state goal) and increasing the proportion of proficient readers (the federal goal) are two separate issues, providing a possible explanation for the difficulty that exists in categorizing students scoring at Level 2.

The findings are less equivocal, but also less positive, for students who scored at Level 1 in 2003, but were promoted due to a good cause exemption. Results show that 59% of these students failed the fifth grade version of the FCAT-Reading (i.e., they scored at Level 1) in 2005. Taken alone, this is an alarmingly high number; however, when compared with the 40% of retained students who failed the fifth grade FCAT-Reading, results suggest that students who were promoted through good cause fared much worse in later years than those who were retained.

With regard to the 2006 performance of students who were promoted due to a good cause exemption in 2003, 56% failed the sixth grade version of the FCAT-Reading, a proportion very similar to the failure rate in the fifth grade. Since these students took a different version of the FCAT-Reading than did the retained students (i.e., they took the sixth grade test, rather than the fifth grade test), the results are not directly comparable; however, these findings again indicate that within their respective curricula, the retained Level 1 students were more competent in reading proficiency relative to the promoted Level 1 students.

In addition, while 41% (2005) and 44% (2006) of Level 1 students promoted through a good cause exemption “passed” the 6<sup>th</sup> grade FCAT-Reading at a Level 2 or higher, only 19% (2005) and 21 % (2006) were considered proficient readers (i.e., scoring at Level 3 or higher). It appears that the majority of students promoted to fourth grade through a good cause exemption in 2003 did not meet with success with either the fifth- or sixth-grade curriculum, a situation predicted by their scores on the third-grade FCAT-Reading.

It is important to consider the results for each type of exemption, however, as students who were promoted due to an exemption that required a demonstration of proficiency in reading (i.e., passing an alternate test, student portfolio) fared quite well in both fifth and sixth grades. Of students promoted for one of the two “proficiency” exemptions, 61% “passed” the FCAT-Reading in fifth grade in 2005, and 64% “passed” the FCAT-Reading in sixth grade in 2006, (i.e., scoring at a Level 2 or higher). In addition to these encouraging results with regards to students promoted through alternate demonstrations of reading proficiency, another group that fared similarly well consisted of students who were promoted due to having Limited English Proficiency (LEP) in 2003 with less than two years of English as a Second Language. Although they lacked adequate reading skills in 2003, and large numbers of these students were still significantly behind in 2004 (69% scored at Level 1 in 4<sup>th</sup> grade in 2004; Powell, 2005), a significant portion of this group improved their reading skills by the time they took the FCAT-Reading in the sixth grade. Results demonstrate that almost half of these students (49%) scored at Level 2 or higher in 2006, presumably aided by an improved mastery of the English language. On the other hand, the majority (51%) still scored at Level 1.

Taken altogether, these findings indicate that there is a distinct difference between students who were promoted to grade four as a result of demonstrating proficiency in reading and those who were promoted for a reason unrelated to adequate academic skills (e.g., IEP indicating inappropriateness of FCAT-Reading, previous retentions, etc.). Thus, it would be unwise to draw conclusions about the appropriateness of promoting students through good cause exemptions without more closely examining the implications of each type of exemption. For example, allowing students to display their skills through alternative mechanisms appears to be an asset to the pupil progression plan; such exemptions provide an additional tool with which students who are likely to be successful with later curricula can be identified. For students who experience test anxiety, or a diminished performance on the FCAT-Reading for any reason other than inadequate skills, these exemptions seem appropriate. These descriptive data suggest that the majority of these students possess the skills necessary to be successful in the next grade and beyond, maintaining their relative skill level into the sixth grade.

In addition, students who qualified for exemption through Limited English Proficiency (LEP) due to less than two years of instruction in an English for Speakers of Other Languages program improved over time; by grade six, their performance on the FCAT-Reading at that time was similar to that of the group promoted through a demonstration of proficiency. This could indicate a need for extra time to master certain aspects of the basic reading curriculum (e.g., phonemic awareness, phonics) before they can be expected to progress along with their same-age peers for whom English is their primary language. However, results from this study suggest that, despite lacking basic reading skills in third grade, these students are not falling farther behind as a result of

being promoted to fourth grade. Rather, they seem to have made up ground, with almost half of them passing the FCAT-Reading in the sixth grade (i.e., Levels 2-5).

Students qualifying for one of the other three good cause exemptions were also promoted and experienced the fourth grade curriculum without adequate basic reading skills; these students appear to have spent the subsequent few years falling even farther behind. Indeed, their performance three years after promotion to fourth grade shows no evidence of an increase in skills relative to their low-achieving peers. Though they are classified as being “promoted due to good cause,” this group of students may have very different needs from those who were able to demonstrate reading competency in an alternate manner. It may be appropriate for policy-makers to re-examine the purpose of promoting these students and to identify where and how their academic needs would best be served.

This study also conducted a direct comparison of two groups of students scoring at Level 1 in third grade: those who were retained in third grade and those who had demonstrated proficiency through another method and were promoted through one of two good cause exemptions. Students in both groups completed the 5<sup>th</sup> grade FCAT-Reading, albeit in different years (i.e., 2005 and 2006, depending on retention status) at the time of this study. Preliminary analyses demonstrated that average student performance varied across schools, indicating the need to statistically control for this between-school variance. In the best-fitting model that accounted for the fact that student data are nested within schools, analyses revealed that whether a student was retained or promoted through a good cause exemption was indeed associated with FCAT-Reading scores, but the magnitude of this effect varied across student gender. Specifically, being male and

promoted through a good cause exemption resulted in an average increase of 12.99 points on the FCAT-Reading when compared to being male and retained. Alternatively, being retained or promoted through an exemption resulted in similar outcomes among female students. In addition to this significant interaction effect, student SES was significantly associated with a school's estimated mean FCAT-Reading score. Specifically, students who were low SES tended to be associated with lower school average FCAT-Reading scores by an estimated 4.72 points.

The main effects and interaction effects found in this study involving the student characteristics of gender and SES are not particularly surprising. Previous research has established that gender (Jimerson, 1999; Pagani et al., 2001), race (Jimerson, 1999; McCoy & Reynolds, 1999), and SES status (Jimerson, 1999; McCoy & Reynolds, 1999; Ferguson et al., 2001) may act as predictors of retention during the elementary grades. Specifically, these studies report that being male, a member of an ethnic minority group, and having a low SES status put children at a higher risk for grade retention. The findings from the current study indicate that gender and student SES continue to impact academic performance during and after the retention year.

With regard to retention status, it appears that being retained in third grade after scoring at Level 1 on the FCAT-Reading is particularly detrimental to male students, as measured by FCAT-Reading scores in the 5<sup>th</sup> grade. With the FCAT-Reading scoring scale set such that Level 2 is defined by a range of only 24 points (i.e., scaled score = 259-283), the almost 13-point difference between the retained and promoted groups among males in this study seems qualitatively important. This result seems to cast doubt on the remedial effect of retention for these students and also raises the question of

gender differences in remediation responsiveness, as well as in acquiring reading skills in general.

It is somewhat curious, given the efforts and resources invested in the development of Reading First schools, that a clear benefit was not found across this low-achieving sample. One possible explanation for this finding is that this study examined the outcomes for already struggling readers. It could be that the real strength of the Reading First program will be seen more clearly in its preventative efforts, that is, a reduction in the percentage of students attending Reading First schools since kindergarten who score at Level 1 in third grade. Another possible explanation for the observed outcome in this study is that in the State of Florida, Reading First funds are distributed according to, among other things, a demonstration that a significant portion of students in a district/school receiving support are eligible for Free or Reduced Lunch. Given this low-income student sample, the known association between low student SES and higher academic need, and a previously demonstrated lack of reading skills in 2003, it seems reasonable to expect that even with the additional interventions associated with Reading First schools, students may need more time before identifiable improvements are seen. Finally, because the interventions associated with the Reading First program target students in grades K-3, only the retained students in this study could have had the opportunity to be exposed to this program, and for only one year. It could be that the limited number of students who were exposed to the program for a limited amount of time contributed to the lack of significant intervention effects.

### *Limitations*

One identified threat to validity in this study is related to the central measure of reading skills used both in this study, as well as in Florida's student progression plan, the FCAT-Reading. There are many issues, too many to delineate here, concerning the use of standardized tests for high stakes decisions such as grade retention. A major concern is the possibility that FCAT-Reading scores may not be a true representation of a student's actual reading skills. Certainly the results of the present study have found that for some students, particularly those who qualified for promotion through an alternate test or portfolio, the FCAT-Reading did not provide the best estimate of their reading skills as the majority managed to be quite successful in sixth grade regardless of their poor performance on the test in the third grade. Although other methods of evaluating reading skills may need to be explored, the State of Florida is currently relying heavily on the FCAT to determine reading skill level; thus, FCAT-Reading scores were used in this study as well.

Another threat to validity in this study involves treatment integrity with regard to state-mandated intervention through Academic Improvement Plans (AIPs) for retained students. The Florida Department of Education delineated several components of an AIP in reading, including: 1) a description of specific deficiencies in phonemic awareness, phonics, fluency, comprehension, and vocabulary, 2) concrete goals in each area, and 3) the specific instructional and/or support services provided (Florida Department of Education, 2003g). For students who were retained in third grade due to receiving a Level 1 score on the FCAT-Reading, it would have been important to review their AIPs at the beginning of the retention year and again as soon as it was determined that the student

might fail the FCAT-Reading again (Florida Department of Education, 2003h). The AIP implementation plan attempted to control for the quality of each student's reading remediation; however, the plan was deficient in at least two areas that may have impacted the conclusions drawn in the present study. Although every student who received a Level 1 score on the FCAT-Reading, regardless of retention/promotion status, should have had an AIP, there was no method in place for evaluating the consistency of AIPs with state requirements or to ensure that each student was actually receiving the intervention as intended. It is highly possible that students' AIPs varied widely both in quality of construction and in implementation, making statistical results difficult to interpret. It is only through detailed documentation regarding the extent to which each student received interventions as mandated that an evaluation of the effectiveness of the interventions, as well as the retention year itself, would be possible.

Another issue that could have impacted the construction and implementation of AIPs was the variability of the remedial activities, which were defined by the state as "effective instructional practices" and "scientifically-based reading instruction." No further details were provided regarding specific activities and interventions, making it incumbent upon practitioners such as school psychologists and teachers to make those decisions according to the specific needs of each student (Florida Department of Education, 2003h). While this afforded great individualization, it also implies that actual interventions might have varied widely, disallowing generalizations about their effectiveness. For the reasons mentioned above, examination of the quality of AIPs was considered to be beyond the scope of the present study and was not included. It must be pointed out that legislation has now replaced the AIP with a mandated Progress

Monitoring Plan (PMP) that places greater emphasis on the use of data to monitor student performance. Nevertheless, issues relating to the integrity on PMP implementation need to be addressed.

In addition to the uncertainties surrounding the implementation of AIPs, there are similar questions regarding the consistency with which Reading First schools were able to implement specified reading interventions as intended. It is important to keep in mind that the “intervention year”, or the year directly following the retention decision for this sample of students, was the first year that Reading First programs were instituted within designated schools. Given the recency of the new federal policies and state legislation at that time, it is probable that an adjustment or transition period existed for this program as well. As with any program, increased time and familiarity with the Reading First program would likely lead to improved implementation, and improved outcomes. The monitoring of adjustment issues, and later modifications to the Reading First program is also beyond the scope of the current study, but these issues remain an important consideration when interpreting the results of this study.

This study was not able to control for inconsistencies in data coding or entry, factors that also threaten the validity of the findings. Data were entered in thousands of schools across the state over several years by numerous coders, using a coding system that was initially unknown and untested. In addition, the study variable of socioeconomic status was defined using students’ free or reduced lunch qualification status. These data were collected by the state and thus this category could not be defined more clearly. While the qualification status for free or reduced lunch as a definition of high or low SES is perhaps overly broad and may not offer a clear picture of students’ actual

socioeconomic status, it is a widely accepted method of categorization in both research and practice.

Additionally, the present study did not include a control group, as the retained and promoted groups of students were not matched. Although the study investigated long-term outcomes of retention and grade promotion, the data were cross-sectional in nature, which limited the nature of the conclusions that could be drawn. Inclusion of a true experimental control group is not ethically possible in this type of research; however, a longitudinal design in which individual student scores were examined across time (i.e., growth curve analysis) would strengthen the study by allowing conclusions to be drawn about the long-term outcomes of early grade retention at the student level.

#### *Delimitations*

The results may be generalized to students who are retained in the third grade in the State of Florida, or in a similar student progression program which requires a structured plan of remediation for every retained student. In addition to retained students, the study includes participants who performed poorly (Level 1) on the FCAT but were promoted due to one of six good cause exemptions. Results can be generalized to other third grade students who meet these requirements. Results cannot be generalized to older or younger populations, or to students with special needs not included in the predefined good cause exemptions.

#### *Implications for Practice and Future Research*

Most educators and researchers concur that neither repeating the same curriculum by repeating a grade nor merely being promoted to the next grade will provide sufficient opportunities for students who are experiencing academic difficulties to experience

success; the results of the present study support this position. Instead, struggling students will need consistent, intensive, and empirically-supported remediation efforts to make sufficient academic gains, and maintain them over time. Florida has instituted AIPs and now PMPs to address this need; however, student outcomes suggest that gaps existed in one or more areas that may have impeded the effectiveness of AIPs, for example, in the development and monitoring of the plans, as well as continuing professional development for personnel relative to the use of the AIPs. Findings from the present study also suggest that certain subgroups of students (e.g., those promoted through good cause exemptions not related to reading proficiency, males, students with a low SES) are at-risk for lowered performance on later assessments of reading proficiency (i.e., 5<sup>th</sup> grade FCAT-Reading); the AIP system and resulting interventions did not prove as effective for these subgroups. This indicates a need to make appropriate modifications that will target these at-risk students and schools, including the development of new interventions or more frequent monitoring and appropriate modifications.

The data from this study also raise questions in terms of policy decisions that promote students scoring at Level 1 through good cause exemptions. The purpose of student promotion through good cause is arguably to help these students achieve future academic success; however, findings of the current study indicate that only one exemption provides a viable predictor of long-term success. Students promoted through Exemption 1 (demonstration of proficiency through an alternative assessment) have a 67% chance of “passing” the FCAT-Reading at Level 2 or higher in the sixth grade. Students promoted through Exemption 2 (demonstration of proficiency through portfolio) have a 58% chance of passing, while students promoted through Exemption 3

(demonstration of limited proficiency in English) have roughly a 50% chance of passing. Students promoted through the other three categories have only a 13%-19% chance of passing. The State of Florida has not defined what percentage of students promoted through a good cause exemption must pass the FCAT-Reading in the fourth grade in order for the exemption to be considered a valid reason for promotion.

This study provides further support for the practice of promoting students who demonstrate proficiency through alternate assessment tests or through portfolios, as there was not a demonstrated difference between scores of retained and promoted female students. There was an observed difference between scores of retained and promoted male students; however, the male students promoted through good cause exemptions indicating proficiency scored significantly *higher* on the 5<sup>th</sup> grade FCAT-Reading. In addition to providing added support for the practice of promoting students based on alternate demonstrations of reading proficiency, the findings in this study raise questions with regard to the appropriateness of the other exemptions as reasons to promote students to the fourth grade.

The good cause exemption issue, it seems, relates to the overall purpose of the policy with regard to students who “fail” the FCAT-Reading. If the purpose is to promote students who demonstrate reasonable evidence of being able to “make it” with successive curricula, the rationale for Exemptions 1 and 2 may be supported. The purpose of the policy could also be better defined with regard to the exemptions indicating non-proficiency. If policy-makers wish to promote only those students with a reasonable chance for success, they would be wise to reexamine students who are qualifying for promotion through Exemptions 4-6. What is contributing to their high rates of long-term

academic failure (i.e., scoring Level 1 on the FCAT-Reading in sixth grade) and their apparent resistance to remediation efforts, and what can be done to reduce those rates? If the purpose of promoting these students lies in something other than academic achievement (e.g., social or emotional health), further research is needed to determine whether these outcomes can be empirically supported and in turn, whether promoting students through these exemptions can be considered a valid practice.

Future research should also expand on the findings presented in this study. More information is needed regarding the student (e.g., demographic characteristics, family support, etc.) or ecological variables (e.g., teacher variables, rural locations, specific interventions used, etc.) that may have contributed to the observed variability in outcomes. For example, school size was demonstrated to have a significant impact on student performance, with students from larger schools scoring higher than students from smaller schools. Based upon what is known about the relationship between school size and location in the state of Florida, one cannot reliably conclude that size alone accounts for the difference in student performance. Specifically, within the state of Florida, small schools are typically located in rural, less affluent areas of the state with higher populations of transitory families. Future studies could address the types variables that may be at work behind a more general variable like school size.

The findings reported in the present study also raise important questions regarding the effects of treatment integrity on outcomes (Do higher levels of treatment integrity during the retention year result in better academic outcomes?), the use of cutoff scores in determining student progression, and which students will most benefit from which types of remediation efforts (What is the treatment utility of using a cutoff score to determine

student progression, and what are the most valid selection criteria?), and variability in performance according to retention status across gender, as well as SES (What interventions may be more beneficial for retained male and/or poor students?). Given the descriptive findings in this study indicating that academic gains demonstrated by the majority of retained students continue through subsequent grades, an even more detailed examination of these students over time (i.e., growth curve analysis) is warranted. Finally, a more detailed comparison of Level 1 students who were retained with those who were promoted through exemptions indicating proficiency should include a sample that is matched on various student characteristics (e.g., 2003 FCAT-Reading scaled score, gender, race, SES).

Another important issue beyond the scope of this study is the social and emotional ramifications of being retained. Research has consistently reported the negative social and emotional effects of grade retention, such as poorer school attendance, social adjustment, and more problem behaviors (Holmes, 1989), as well as peer difficulties (Shepard & Smith, 1990). Grade retention has historically been a stigmatizing experience for many students such that even the large numbers seen in the State of Florida would not necessarily improve the emotional outcomes of being ostracized from one's peer group. Despite this concern, however, no studies have examined these constructs in the context of large-scale retention as part of a state-wide pupil progression plan.

### *Conclusions*

Recognition of the importance of early literacy skills has resulted in a national educational climate that emphasizes reading development in students in grades K through 3. Education policies in the State of Florida have reflected this climate; its pupil

progression plan includes the use of high-stakes testing and mandatory retention for third-grade students who do not display adequate progress toward the acquisition of reading skills. In addition, third-grade students who are retained in Florida are required through legislation and policy to receive intensive academic interventions and their progress must be monitored through the collection and analysis of data. The literature is replete with studies indicating the potentially harmful effects of grade retention on students' future academic progress; however, previous research has not included an examination of the effects of remediation efforts – such as those that exist in Florida – occurring during the retention year. The present study found that the majority of students who were retained in third grade scored at Levels 2 or higher on the FCAT-Reading several years following the retention decision (60%). Proportionately fewer students who failed the FCAT-Reading in 2003 but were promoted through a good cause exemption requiring a demonstration of proficiency were successful in the sixth grade (40%); those promoted through a good cause exemption not requiring demonstration of reading proficiency were found to be still struggling in sixth grade despite presumed remedial efforts; more than 80% of students in these groups failed the 6<sup>th</sup> grade FCAT-Reading in 2006. However, those promoted through one of the good cause exemptions requiring a demonstration of proficiency fared quite well (i.e., 67% and 58% scored at Level 2 or higher, for the use of an alternative standardized test or of a student portfolio, respectively). When comparing the fifth-grade performance of this smaller group of good cause promoted students with the fifth-grade performance of retained students, no difference in average in FCAT-Reading performance was seen. Taken together, these results suggest support for both retention and good cause promotion practices based on demonstration of reading proficiency, as

well as the appropriateness of examining subgroups of students when creating policy. In addition, several risk factors for lowered performance in subsequent years after the decision to retain or promote were identified, including being male, and being of low SES. However, the cross-sectional design and lack of experimental control in this study preclude making definitive conclusions. Future efforts should focus on learning more about these specific populations, as current interventions appear to be less effective with certain groups of students.

## References

- Akaike, H. (1974). A new look at the statistical model of identification. *IEEE Transaction on Automatic Control*, *19*, 716-723.
- Alexander, K. L., Entwisle, D. R., & Dauber, S. L. (1994). *On the success of failure: A reassessment of the effects of retention in the primary grades*. New York, NY: Cambridge University Press.
- Armistead, L., Kempton, T., Lynch, S., & Forhand, R. (1992). Early retention: Are there long-term beneficial effects? *Psychology in the Schools*, *29*(4), 342-347.
- Buntaine, R. L., & Costenbader, V. K. (1997). The effectiveness of a transitional prekindergarten program on later academic achievement. *Psychology in the Schools*, *43*, 41-50.
- Coleman, J. S. (1990). *Equality and achievement in education*. San Francisco: Westview Press.
- Dennebaum, J. M., & Kulberg, J. M. (1994). Kindergarten retention and transition classrooms: Their relationship to achievement. *Psychology in the Schools*, *31*, 5-12.
- Ferguson, P., Jimerson, S. R. & Dalton, M. J. (2001). Sorting out successful failures: Exploratory analyses of factors associated with academic and behavioral outcomes. *Psychology in the Schools*, *38*(4), 327-341.
- Fine, J. G., & Davis, J. M. (2003). Grade retention and post-secondary enrollment. *Journal of School Psychology*, *41*, 401-411.
- Florida Department of Education. (2002). *Student Progression*. Retrieved March 7, 2006 from <http://www.firn.edu/doe/commhome/progress/proghome.htm>

Florida Department of Education (2002b). Technical report: For operational test Administrations of the 2000 Florida Comprehensive Assessment Test. Retrieved March 7, 2006 from <http://www.firn.edu/doe/sas/fcat/pdf/fc00tech.pdf>

Florida Department of Education. (2003a). No Child Left Behind: Adequate yearly progress fact sheet. Retrieved March 7, 2006 from <http://www.fldoe.org/NCLB/>

Florida Department of Education. (2003b). Just Read, Florida!: Reading First. Retrieved March 7, 2006 from [http://www.justreadflorida.com/reading\\_first.asp](http://www.justreadflorida.com/reading_first.asp)

Florida Department of Education. (2003c). FCAT reading SSS. *The Florida Comprehensive Assessment Test*. Retrieved March 7, 2006 from <http://www.firn.edu/doe/sas/fcat.htm>

Florida Department of Education. (2003d). Third grade reading – 2003 profiles. Retrieved March 7, 2006 from <http://www.firn.edu/doe/bin00014/progress/progress/file.htm>

Florida Department of Education. (2003e). Good Cause Exemptions. Retrieved March 7, 2006 from <http://www.fldoe.org/ese/pdf/y2004-17.pdf>

Florida Department of Education. (2003f). Third Grade Progression – Requirements for students who were retained. Retrieved March 7, 2006 from <http://www.firn.edu/doe/commhome/progress/proghome.htm>

Florida Department of Education. (2003g). Public School Student Progression. Retrieved March 7, 2006 from [http://info.fldoe.org/docushare/dsweb/Get/Document-433/DPS\\_Memo\\_03-024-attach2.pdf](http://info.fldoe.org/docushare/dsweb/Get/Document-433/DPS_Memo_03-024-attach2.pdf)

Florida Department of Education. (2003h). Technical Assistance Paper Related to Third Grade Student Progression. Retrieved March 7, 2006 from

<http://www.firn.edu/doe/commhome/progress/proghome.htm>

Florida Department of Education (2004a). Third Grade Reading 2004 Profile – State. Retrieved March 7, 2006 from

<http://www.firn.edu/doe/commhome/pdf/3statesum.pdf>

Florida Department of Education (2004b). Florida Comprehensive Assessment Test 2004: State Report of District Results. Retrieved March 7, 2006 from

<http://fcats.fldoe.org/2004/default.asp>

Florida Department of Education (2005). Reading and Mathematics Scores – GRADE 3 Statewide Comparison for 2001 to 2005. Retrieved March 7, 2006 from

<http://fcats.fldoe.org/>

Florida Department of Education (2006). Report on the Recommendations of the Paper Reduction Task Force. Retrieved June 24, 2007 from

[http://www.fldoe.org/PaperReduction/pdfs/PRTF\\_Report.pdf](http://www.fldoe.org/PaperReduction/pdfs/PRTF_Report.pdf)

Florida Department of Education (2007a). Non-Promotions in Florida's Public Schools, 2005-2006. Retrieved June 3, 2007 from

<http://www.firn.edu/doe/eias/eiaspubs/pdf/nonpromotions.pdf>

Florida Department of Education (2007b). Florida School Indicators Report. Retrieved May 29, 2007 from <http://data.fldoe.org/fsir/>

Florida Department of Education (2007c). FCAT 2006 Reading and Mathematics Scores. Retrieved June 7, 2007 from <http://fcats.fldoe.org/fcainfopg.asp>

- Goddard, R. D. (2001). Collective efficacy: A neglected construct in the study of schools and student achievement. *Journal of Educational Psychology, 93*(3), 467-476.
- Gottfredson, D. C., Fink, C. M., & Graham, N. (1994). Grade retention and problem behavior. *American Educational Research Journal, 31*(4), 761-784.
- Holmes, C. T. (1989). Grade level retention effects: A meta-analysis of research studies. In L. A. Shepard & M. L. Smith (Eds.), *Flunking grades: Research and policies on retention* (pp. 16-33). London: Falmer Press.
- Hong, G., & Gaudenbush, S. W. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis, 27*(3), 205-224.
- Jackson, G. B. (1975). The research evidence on the effects of grade retention. *Review of Educational Research, 45*, 613-635.
- Jencks, C., & Mayer, S. E. (1990). The social consequences of growing up in a poor neighborhood. In L. Lynn Jr. & M. G. H. McGeary (Eds.), *Inner-city poverty in the United States* (pp. 111-186). Washington, DC: National Academy Press.
- Jimerson, S. (1999). On the failure of failure: Examining the association between early grade retention and education and employment outcomes during late adolescence. *Journal of School Psychology, 37*(3), 243-272.
- Jimerson, S. (2001). Meta-analysis of grade retention research: Implications for practice in the 21<sup>st</sup> century. *School Psychology Review, 30*(3), 420-437.
- Jimerson, S., Carlson, E., Rotert, M., Egeland, B., & Sroufe, A. (1997). A prospective, longitudinal study of the correlates and consequences of early grade retention. *Journal of School Psychology, 35*(1), 3-25.

- Just Read, Florida. (2003). Read to Learn: *Intervention*. Retrieved March 7, 2006 from <http://www.read-to-learn.org/intervention.html>
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental scores*. Reading, MA: Addison-Wesley.
- Ma, X., & Willms, J. D. (2004). School disciplinary climate: Characteristics and effects on eighth grade achievement. *Alberta Journal of Educational Research, 50*(2), 169-188.
- Mantzicopoulos, P. (1997). Do certain groups of children profit from early retention? A follow-up study of kindergartners with attention problems. *Psychology in the Schools, 34*, 115-127.
- Mantzicopoulos, P., & Morrison, D. (1992). Kindergarten retention: Academic and behavioral outcomes through the end of second grade. *American Educational Research Journal, 29*, 182-198.
- McCoach, D. B., O'Connell, A. A., & Reis, S. M. (2006). Growing readers: A hierarchical linear model of children's reading growth during the first 2 years of school. *Journal of Educational Psychology, 98*(1), 14-28.
- McCoy, A. R., & Reynolds, A. J. (1999). Grade retention and school performance: An extended investigation. *Journal of School Psychology, 37*(3), 273-298.
- Meisels, S. J., & Liaw, F. (1993). Failure in grade: Do retained students catch up? *Journal of Educational Research, 87*(2), 69-77.
- Nagin, D. S., Pagani, L., Tremblay, R. E., & Vitaro, F. (2003). Life course turning points: The effect of grade retention on physical aggression. *Development and Psychopathology, 15*, 343-361.

- National Reading Panel. (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Retrieved March 7, 2006 from <http://www.nichd.nih.gov/publications/nrp/smallbook.htm>
- No Child Left Behind. (2002). Retrieved March 7, 2006 from <http://www.ed.gov/nclb/landing.jhtml?src=pb>
- Pagani, L., Tremblay, R., Vitaro, F., Boulerice, B., & McDuff, P. (2001). Effects of grade retention on academic performance and behavioral development. *Development & Psychopathology, 13*(2), 297-315.
- Peterson, S. E., DeGracie, J. S., & Ayabe, C. R. (1987). A longitudinal study of the effects of retention/promotion on academic achievement. *American Educational Research Journal, 27*, 107-118.
- Phelps, L., Dowdell, N., Rizzo, F. G., Ehrlich, P., & Wilczenski, F. (1992). Five to ten years after placement: The long-term efficacy of retention and pre-grade transition. *Journal of Psychoeducational Assessment, 10*, 116-123.
- Powell, H. (2005). Third grade retention and the Florida Comprehensive Assessment Test: An exploratory study. Unpublished thesis
- Porter, L. (2005). The effects of grade retention on the performance of students on high stakes testing. Unpublished thesis.
- Pressley, M., Wharton-McDonald, R., Allington, R., Block, C. C., Morrow, L., Tracey, D., Baker, K., Grookes, G., Cronin, J., Nelson, E., & Woo, D. (2001). A study of effective first-grade literacy instruction. *Scientific Studies of Reading, 5*(1), 35-58.

- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2<sup>nd</sup> ed.). Newbury Park: Sage Publications.
- Reynolds, A. J. (1992). Grade retention and school adjustment: An explanatory analysis. *Educational Evaluation & Policy Analysis, 14*(2), 101-121.
- Rumberger, R. W., & Palardy, G. J. (2005). Does segregation still matter? The impact of student composition on academic achievement in high school. *Teachers College Record, 107*(9), 1999-2045.
- Simmons, R. G., & Blyth, D. A. (1987). *Moving into adolescence: The impact of pubertal change and school context*. Hawthorne, NY: Aldine de Gruyter.
- The Florida Senate. (2003). The 2003 Florida Statutes: Title XLVIII; Chapter 1008.  
Retrieved March 7, 2006 from  
[http://www.flsenate.gov/Statutes/index.cfm?App\\_mode=Display\\_Statute&Search\\_String=&URL=Ch1008/SEC25.HTM&Title=-](http://www.flsenate.gov/Statutes/index.cfm?App_mode=Display_Statute&Search_String=&URL=Ch1008/SEC25.HTM&Title=-)
- United States Department of Education, (2002). *Overview: Why NCLB is Important to America*. Retrieved March 7, 2006 from  
<http://www.ed.gov/nclb/overview/importance/edlite-index.html>
- United States Department of Education. (2003a). Stronger accountability: Accountability for schools. Retrieved March 7, 2006 from  
<http://www.ed.gov/nclb/accountability/schools/accountability.html>
- United States Department of Education. (2003b). Introduction: No Child Left Behind  
Retrieved May 3, 2005 from <http://www.ed.gov/nclb/overview/importance/edlite-index.html>

United States Department of Education. (2003c). Stronger accountability: The facts about making gains each year. Retrieved March 7, 2006 from

<http://www.ed.gov/nclb/accountability/ayp/yearly.html>

United States Department of Education. (2003d). Proven methods: Reading. Retrieved

March 7, 2006 from <http://www.ed.gov/nclb/methods/reading/reading.html>

Vitaro, F., Desmarais-Gervais, L., Tremblay, R. E., & Gagnon, C. (1992). Social adjustment at the beginning of elementary school and preschool day care arrangements. *Revue des Sciences de l'Education*, 18, 1-15.

## Appendices

Appendix A. Full Model + All Interactions

Model 4			
Full Model + All Interactions			
<i>Fixed effect</i>	PE	SE	t
Intercept	273.20	2.85	95.93 ***
ret	-3.61	3.47	-1.04
gender	-11.15	2.64	-4.21 ***
race	0.37	2.97	0.12
ses	-4.56	2.90	-1.57
sses	-0.03	0.08	-0.31
ssize	0.02	0.00	3.09 **
RF	-3.70	3.52	-1.05
<i>Interactions</i>			
ret*gen	10.69	2.29	4.68 ***
ret*race	-0.85	2.95	-0.29
ret*ses	3.20	3.16	1.01
ret*sses	0.02	0.07	0.30
ret*ssize	-0.00	0.00	-0.07
ret*RF	-4.05	2.92	-1.39
gen*race	-1.07	2.21	-0.48
gen*ses	-1.88	2.55	-0.74
gen*sses	0.03	0.05	0.63
gen*ssize	0.00	0.00	0.02
gen*RF	2.73	2.11	1.29
rac*ses	-0.88	2.65	-0.33
race*sses	-0.04	0.06	-0.63
race*ssize	-0.01	0.01	-1.53
race*RF	-4.32	2.64	-1.63
ses*sses	0.02	0.07	0.25
ses*ssize	0.00	0.01	0.38
ses*RF	3.91	3.28	1.19
sses*ssize	0.00	0.00	1.25
sses*RF	0.03	0.09	0.29
ssize*RF	-0.02	0.01	-1.93
<i>Random Effects</i>			
gen	42.57	17.82	2.39 **
race	2.82	17.88	0.16
Tau00	65.28	17.27	3.78 ***
Sigma	2310.59	28.07	82.31 ***
	R <sup>2</sup> within	R <sup>2</sup> between	
	.02	.36	

\* p < .05; \*\* p < .01; \*\*\* p < .0001

### About the Author

Heather A. Powell received her Master of Arts degree from American University in 2000, and will receive her Doctor of Philosophy degree in School Psychology from the University of South Florida in 2007. She was the recipient of a graduate training award from USF in 2003. She completed her predoctoral training at the Kennedy Krieger Institute/Johns Hopkins School of Medicine in pediatric and school psychology, and will begin a postdoctoral fellowship in the Department of Child and Adolescent Psychiatry at the Johns Hopkins Hospital following graduation, providing psychological services to children and adolescents with a variety of medical, emotional, behavioral, and educational issues.