

4-4-2006

An Exploratory Analysis Of The Effects Of A Statewide Mandatory Grade Retention Policy And Student Academic Achievement

Larry J. Porter Jr.
University of South Florida

Follow this and additional works at: <https://digitalcommons.usf.edu/etd>

 Part of the [American Studies Commons](#), and the [Education Commons](#)

Scholar Commons Citation

Porter, Larry J. Jr., "An Exploratory Analysis Of The Effects Of A Statewide Mandatory Grade Retention Policy And Student Academic Achievement" (2006). *Graduate Theses and Dissertations*.
<https://digitalcommons.usf.edu/etd/3762>

This Ed. Specialist is brought to you for free and open access by the Graduate School at Digital Commons @ University of South Florida. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Digital Commons @ University of South Florida. For more information, please contact scholarcommons@usf.edu.

An Exploratory Analysis Of The Effects Of A Statewide Mandatory Grade Retention
Policy And Student Academic Achievement

by

Larry J. Porter, Jr.

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

Major Professor: George Batsche, Ed.D.
Michael Curtis, Ph.D.
Constance Hines, Ph.D.

Date of Approval:
April 4, 2006

Keywords: retention, promotion, high stakes testing, student progression, remedial
instruction

© Copyright 2006, Larry J. Porter, Jr.

Table of Contents

List of Tables	iii
List of Figures	iv
Abstract	v
Chapter I: Introduction	1
Purpose	8
Research Questions	9
Hypotheses	10
Definition of Terms	11
Chapter II: Review of Selected Literature	13
Introduction	13
Characteristics of Retained Students	15
Negative Outcomes of Student Retention	16
Positive Outcomes of Student Retention	26
The State of Florida's Retention Model	29
Supplemental Instruction	30
Early Intervention	31
Conclusion	32
Chapter III: Method	35
Introduction	35
Sample	35
Instruments	37
Procedure	39
Data Analysis	40
Chapter IV: Results	48
Introduction	48
Data Screening	48
Characteristics of Final Samples	49
Performance of the Samples on the 2003 3 rd -grade FCAT Reading Test	50
Performance of Higher Achieving Retained Students	52
Performance of Promoted Low Achieving Students	61
Chapter V: Discussion	69

Introduction	69
Student Characteristics	69
Retained Student Outcomes	71
Promoted Student Outcomes	74
Student Outcomes by Gender	75
Student Outcomes by Race/Ethnicity	76
Limitations	77
Delimitations	79
Implications for Florida’s Student Progression Policy	80
Implications for Future Research	83
Conclusion	84
References	86

List of Tables

Table 1	Demographic Characteristics of Retained and Promoted Students	50
Table 2	Means and Standard Deviations of 2003 3 rd grade FCAT-Reading Scaled Scores for Retained and Promoted Students	51
Table 3	Number and Percent of Retained Students on the 2004 3 rd grade FCAT Reading Test by Performance Level	52
Table 4	Number and Percent of Retained Students by Performance Level and Gender on the 3 rd grade FCAT-Reading in 2004	55
Table 5	Number and Percent of Retained Students by Performance Level and Race/Ethnicity on the 3 rd grade FCAT-Reading in 2004	56
Table 6	Number and Percent of Retained Students by Performance Level and Size of District on the 3 rd grade FCAT-Reading in 2004	57
Table 7	Logistic Regression Analysis of Retained Students Meeting State Standards on the 2004 3 rd grade FCAT-Reading	59
Table 8	Observed and Predicted Frequencies for Attainment of State Standards in Reading by Logistic Regression with the Cutoff of .87	60
Table 9	Number and Percent of Promoted Students on the 2004 4 th grade FCAT-Reading by Performance Level	61
Table 10	Number and Percent of Promoted Students by Performance Level and Gender on the 4 th grade FCAT-Reading in 2004	62
Table 11	Number and Percent of Promoted Students by Performance Level and Race/Ethnicity on the 4 th grade FCAT-Reading in 2004	63
Table 12	Number and Percent of Promoted Students by Performance Level and Size of District on the 4 th grade FCAT-Reading in 2004	64
Table 13	Logistic Regression Analysis of Promoted Students Meeting State Standards on the 2004 4 th grade FCAT-Reading	66

Table 14 Observed and Predicted Frequencies for Success by Logistic Regression with the Cutoff of .67

67

List of Figures

- Figure 1 Percentages of retained students attaining Levels 1, 2, and 3 through 5 on the 2004 3rd grade FCAT reading test and promoted, low achieving students attaining Levels 1, 2, and 3 through 5 on the 2004 4th grade FCAT reading test 54

An Exploratory Analysis Of The Effects Of A Statewide Mandatory Grade Retention Policy And Student Academic Achievement

Larry J. Porter, Jr.

ABSTRACT

The literacy skills of students have become a significant concern among legislators and educators. The federal government has responded to this by enacting legislation that increases state accountability to provide evidence-based interventions to struggling readers. In response, the State of Florida has mandated mandatory retention for third-grade students who are at risk for reading failure. Third-grade students who do not pass the Florida Comprehensive Assessment Test-Reading (FCAT) are retained. Students who score at Level 1 are retained, and students who scored at Levels 2 through 5 are promoted.

Research has indicated that retention has been an ineffective intervention to improve academic performance. However, it is difficult to compare research findings with Florida's current retention plan. Previous research has not delineated the intervention strategies that were utilized during the retention year. Florida requires that all students are provided evidence-based reading remediation.

The purpose of this study was to explore the association of Florida's model of student progression and academic achievement. More specifically, the study investigated

the academic outcomes of third-grade students who scored within 10 scaled score points below the student progression achievement cut-off, attained a Level 1 designation in 2003 and were retained, and students who scored within 10 scaled score points above the student progression achievement cut-off, attained a Level 2 designation in 2003 and were promoted to fourth grade.

Results indicated that 87% of the higher performing retained students subsequently scored at Level 2 or higher in 2004 while 67% of the promoted, low achieving student scored at Level 2 or higher in 2004. Furthermore, gender, SES and race were significantly associated with the reading outcomes of higher achieving retained and promoted, low achieving students.

This study contributes to the literature by examining the outcomes of a retention model within a framework of academic remediation. In addition, the utility of high stakes testing and retention decisions were also examined. Future implications for research include direct comparisons of retained and promoted students, a longitudinal research design to examine the long-term effects of retention, and the identification of more effective services and intervention strategies to target at-risk students.

Chapter One

Introduction

Since the introduction of the Elementary and Secondary Education Act (ESEA) in 1965, the federal government has funded \$320 billion for education. However, data suggest that students are not achieving desired levels of proficiency in the academic areas of reading, mathematics, science, and writing (United States Department of Education, 2002). In 2000, only 32% of fourth-grade students in the United States were considered proficient in reading (Donahue, Finnegan, Lutkus, Allen, & Campbell, 2001). The disparity between funding in education and academic achievement resulted in the Congress of the United States requiring state-specific standards for student achievement and the use of evidence-based instruction and interventions. In the past, education has been the sole responsibility of the states (U.S. Constitution). While this is still the case, the federal government has used funding to states as leverage for policy change. In order to obtain federal financial assistance for education funds, states must ensure that their policies and practices are consistent with the ESEA.

Elementary and Secondary Education Act

Previous laws such as the ESEA (1965) focused on ensuring that states complied with the provisions in the law. States followed strict regulations in order to obtain federal funds. The No Child Left Behind Act (NCLB, 2002), the 2002 version of ESEA, represented a shift in focus from compliance with regulations to outcome-based services

(Linn, Baker, & Betebenner, 2002). States are required to demonstrate that their schools use disaggregated student data to demonstrate outcomes, that their curriculum is evidence-based, and that all students are held accountable to a single standard. In order to ensure that all students are proficient in reading, each state must establish benchmarks, known as adequate yearly progress (AYP), that lead to full proficiency by 2012 (Linn et al., 2002).

No Child Left Behind (NCLB) Act

The impact of the No Child Left Behind (NCLB) Act on district, and local school educational practices has been significant. In order to secure funding under NCLB, states must provide documentation of student performance relative to state goals. Also, states must submit their goals to the United States Department of Education for approval (Mann & Shakeshaft, 2003). These goals are then evaluated by the U.S. Department of Education, and funds are dispersed on the approval of those state goals. In addition, NCLB provides federal funding for after-school programs and requires that every classroom be staffed by highly qualified teachers to teach in the curriculum content area (Canales, Frey, Walker, Walker, Weiss, & West, 2002).

A central tenet of the Improving the Academic Achievement of the Disadvantaged Title I in NCLB (2002), is that reading proficiency is paramount for positive student achievement outcomes. Children must learn to read in order to read to learn (Donnelly, 2000). In the grades Kindergarten through third grade, students are instructed in basic reading skills such as decoding, oral fluency, phonics, and phonemic awareness (Sindelar, Lane, Pullen, & Hudson, 2002). However, in the fourth grade, a curriculum shift requires students to read in order to obtain knowledge. The primary

focus of reading instruction shifts from reading mechanics to understanding what is read and to using that information appropriately. Therefore, reading becomes a vehicle for subsequent learning. In order to ensure that students were “reading to learn” by fourth grade, the federal government allocated a significant amount of funding, through the Reading First and Early Reading First initiatives, to states for evidence-based reading instruction in kindergarten through grade three. The Reading First initiative makes \$900 million (in addition to \$275 million for early reading first) available through grants to states to support reading instruction in the early grades (NCLB, 2002). The goal of Reading First is to have every student proficient in reading by the end of third grade (United States Department of Education, 2003). An integral component of Reading First is the application of evidence-based reading research to instructional techniques employed by schools. In order to receive a Reading First grant, states must submit a proposal to the U.S. Department of Education that delineates the specific conditions under which the Reading First initiative will be implemented in that state. This proposal is then reviewed by a panel which then makes recommendations based upon individual state needs.

Reading First and NCLB specify that reading interventions should focus on grades kindergarten through third grade (NCLB, 2002). The fundamental reason for this narrow focus is that (a) reading difficulties are more easily prevented than remediated (Coyne, Kame’enui, & Simmons, 2001), and (b) the remediation of reading difficulties is most successful if interventions are employed early in the development of the problem (Haager & Windmueller, 2001, Jenkins & O’conner, 2002; Phillips, Norris, Osmond, &

Maynard, 2002). By the time a student reaches the secondary grades, it may be too late to implement basic reading interventions successfully (Coyne et al., 2001).

Reading Interventions

Many reading strategies are used to advance the reading proficiency of individual students. Interventions such as previewing and repeated readings are utilized to increase oral reading fluency in students (Sindelar, et al, 2002). Each strategy is contingent on the amount of time students are exposed to text. Text comprehension interventions include vocabulary instruction, guided oral reading, and increased teacher preparation.

Vocabulary instruction involves the direct instruction of vocabulary words while teacher preparation focuses on increasing teachers' instructional competence (National Reading Panel, 2000). Other strategies have been employed to increase reading achievement in multiple students including class-wide peer tutoring programs and self-monitoring interventions (Greenwood, Maheady, & Delquadri, 2002; Shapiro, Durnan, Post, & Levinson, 2002).

Student Retention

One of the interventions traditionally used for students performing below grade level in reading has been grade retention (Jimerson, 2001; National Association of School Psychologists, 2003a). According to the National Association of School Psychologists (2003a) 15% to 20% of students are retained in the United States each year and 30% to 50% of students are retained before ninth grade. Retention occurs when a student is required to repeat a particular grade year as a result of delayed academic progress (Rafoth, 1991). The assumptions underlying the use of retention are that students who do not possess basic academic skills will not be successful in successive grade levels. It is

also assumed that students have not developed these skills because they have not had sufficient practice time and opportunities to learn in order for these skills to develop. Therefore, retained students may benefit and respond to extra instructional time and become more proficient in reading, writing, and mathematics (Graue & DiPerna, 2000).

Each school district in the state of Florida is required to develop a district student progression plan (K-20 Education Code, 2003). Progression plans delineate the criteria necessary for a student to be moved from one grade to the next, and ultimately, graduate from secondary school. Some states have implemented grade retention in order to prepare children more adequately for the increasing demands of the curriculum as grade level increases. For example, Florida has instituted a policy which mandates that students pass the reading section of the Florida Comprehensive Assessment Test (FCAT) in third-grade in order to be promoted to the fourth grade (Florida Department of Education, 2002a). Thus, students who do not possess basic reading skills by the end of the third grade are retained (Florida Department of Education, 2002a). Exceptions are written into this law that permit for the exclusion of students with disabilities, limited English proficient students and students who have been retained previously. In addition, exemptions are made for those students who perform poorly on standardized tests such as the FCAT, but who can otherwise demonstrate proficiency through portfolios or other alternate assessments such as the Stanford Achievement Test – Tenth Edition (SAT-10; Florida Department of Education, 2002a).

Despite the widespread adoption of retention by states and districts, a review of the retention literature has suggested that retention has been an ineffective intervention to improve academic performance (Denton, 2001; Jimerson, 2001; Jimerson & Kaufman,

2003). Negative side effects have also been identified, including, (a) increased drop out rate (Jimerson, Anderson, & Whipple, 2002), (b) lower performance on standardized academic achievement tests (Graue & Diperna, 1999; Jimerson 1999; Reynolds, 1992), (c) increased negative feelings towards learning (Fergusen, Jimerson, & Dalton, 2001), and (d) increased behavior problems as measured by ratings on behavior scales (National Association of School Psychologists, 2003a). Research has suggested that students who are retained are at risk for adverse social adjustment in school, and may suffer from lower self-esteem than students who are promoted (National Association of School Psychologists, 2003a). Moreover, research has indicated that males, minority, and low socioeconomic students are retained at disproportional rates relative to their peers. Research seems to project a grim picture of what happens when districts embrace a policy of retention.

The methods that have been used to explore the effects of retention typically compare aggregated outcome measures of students who were retained and those who were not. However, a comprehensive examination of the research reveals that there may be serious limitations in the methods used to evaluate the effectiveness of retention. The general format of the research on retention has been to compare groups of retained students to those not retained or to conduct a longitudinal study demonstrating the long-term outcomes for students who were retained. There has been no research found evaluating the different methods of retention, comparing different activities that occur during the retention year, or evaluating the effects of retention across diverse student demographic characteristics (e.g., age, grade, gender, race). No study was found that has differentiated between various types of retention practices (e.g., retention in one subject

area only, retention with intense remediation), and no study was found that compared students who simply repeated a grade to those for whom systematic, evidence-based interventions were used during the retention year.

Florida's Model of Retention

The state of Florida has adopted a model of retention that is conceptually different than previous models of retention (Florida Department of Education, 2002a). Retention in the State of Florida calls for retention with remediation of academic skills. Policies in the State of Florida dictate that the needs of retained students are addressed systematically during the retention year. More specifically, the State of Florida's retention policy requires that students do not merely repeat the same curriculum and experiences. Rather, interventions are developed that are student specific and are designed to address individual skill deficits. The policy calls for Academic Intervention Plans (AIPs) to be developed for every student who is retained. These plans include instructional modifications that are linked to individual skill deficiencies in students by setting clear and measurable academic goals. These AIPs are then evaluated frequently to determine if retained students are making progress to attain their academic goals (Florida Department of Education, 2002a). Although AIPs are required by the State of Florida, the implementation and integrity of AIPs vary by location and it was not possible to account for these differences. Examples of modifications supported by AIPs include, pull-out services, one-on-one tutor instruction, peer tutors, and the employment of reading coaches.

In order to determine which third-grade students have not attained the reading proficiencies necessary for promotion to fourth grade, Florida uses a high stakes testing

procedure. Third-grade students must pass the Florida Comprehensive Assessment Test - Reading in order to be promoted to fourth grade. Third-grade students who do not meet state standards on the reading portion of the FCAT are required to be retained (Florida Department of Education, 2002a). The scoring on the FCAT consists of scaled scores of which are broken into five achievement levels of reading proficiency. The decision to retain third-grade students in the State of Florida is contingent, among other factors, on FCAT reading achievement levels. The achievement Level 1 represents scaled scores of 258 and lower and does not meet reading state standards. Third-grade students who attain a scaled score of 258 or lower and subsequently a Level 1 designation on the FCAT reading test are retained. The achievement Levels 2 through 5 includes scaled scores of 259 and higher. These achievement levels are considered to meet state standards for reading and third-grade students who obtain the reading achievement Levels 2 through 5 are not required to be retained.

Purpose

Given the lack of research that has delineated explicit or implied policies for the type of services provided to students during the retention year, the effects of retention on diverse student populations, and the usefulness of using cutoff scores for identifying students for mandated retention, the purpose of the present study was to explore the association of Florida's model of student progression and academic achievement. More specifically, the study investigated the academic outcomes, as measured by FCAT reading levels, of third-grade students who scored within 10 scaled score points below the student progression achievement cut-off (248-258), attained a Level 1 designation in 2003 and were retained, and students who scored within 10 scaled score points above the

student progression achievement cut-off (259-269), attained a Level 2 designation in 2003 and were promoted to fourth grade. Student population characteristics (e.g., ethnicity, social economic status) and size of district were also explored to extend previous research on retention and to determine if Florida's Retention Policy is equitable for diverse populations of students.

Research Questions

In order to explore the effectiveness of Florida's retention policy, the present study examined the relationship between student retention and reading outcomes measured by FCAT levels attained by third-grade higher achieving retained students and fourth-grade promoted, low achieving students. Specifically, the present research questions included:

1. What proportion of higher performing retained third-grade students who had reading scaled scores of 248-258 on the 2003 FCAT reading test subsequently scored at Level 2 or higher on the 2004 3rd-grade FCAT reading test?
2. What is the relationship between a) gender, and b) race/ethnicity and performance on the 2004 3rd-grade FCAT reading test for higher performing third-grade retained students who had scaled scores of 248-258 on the 2003 FCAT reading test?
3. Is there a relationship between the size of school district attended and performance on the 2004 third-grade FCAT reading test for higher performing third-grade retained students who had scaled scores of 248-258 on the 2003 FCAT reading test?

4. What is the relationship among gender, race/ethnicity, prior performance on the FCAT reading test and attaining state reading standards on the 2004 FCAT reading test for higher performing students retained in third grade?
5. What proportion of promoted third-grade students who had reading scaled scores of 259-269 on the 2003 FCAT reading test subsequently scored at Level Two or higher on the 2004 fourth-grade FCAT reading test?
6. What is the relationship between a) gender, and b) race/ethnicity and performance on the 2004 fourth-grade FCAT reading test for third-grade promoted students who had scaled scores of 259-269 on the 2003 FCAT reading test?
7. Is there a relationship between size of school district attended and performance on the 2004 fourth-grade FCAT reading test for low achieving students promoted to 4th grade who had scaled scores of 259-269 on the 2003 FCAT reading test?
8. What is the relationship among gender, race/ethnicity, and prior performance on the 2003 3rd-grade FCAT reading test and attainment of state standards on the 2004 fourth-grade FCAT reading test for low achieving students who were promoted to fourth grade?

Hypotheses

Previous research has suggested that retention is not an effective academic intervention for students (Jimerson, 1999; Jimerson, 2001). Therefore, it was hypothesized that:

1. There is a significant difference in the obtained 2004 FCAT reading levels of higher performing retained students by a) gender, and by b) race/ethnicity.
2. There is a significant difference in the obtained post-retention 3rd-grade FCAT reading levels of higher performing retained students as a function of district size.

3. For higher performing students retained in the third grade, there is a significant relationship among gender, race/ethnicity, prior performance on the FCAT reading test and attaining state reading standards on the 2004 FCAT reading test.
4. There is a difference in the obtained 2004 FCAT reading levels of promoted, low achieving students by a) gender, and by b) race/ethnicity.
5. There is a difference in the obtained post-retention FCAT reading levels of promoted, low achieving students as a function of district size.
6. For students promoted to fourth grade, there is a relationship among gender, race/ethnicity, prior performance on the FCAT reading test and attaining state reading standards on the 2004 FCAT reading test.

Definition of Terms

Retention: Repeat a grade the subsequent year because of inadequate academic progress.

High Performing Retained Students: Third-grade students in the State of Florida who were retained at the end of 2003 academic year, attained a scaled score of 248-258 on the 2003 third grade FCAT reading test, and were designated as scoring at Level One on the 2003 FCAT reading test.

Promoted, Low Achieving Students: Third-grade students in the State of Florida who were promoted to fourth grade at the end of 2003 academic year, attained a scaled score of 259-269 on the 2003 third-grade FCAT reading test, and were designated as scoring at Level Two on the 2003 FCAT reading test

Low Socio- Economic Status: Students who received free and reduced lunch in Florida during the 2003-2004 school year.

Meeting State Standards: Students who attained a scaled score of 259 or higher on the FCAT reading test.

Chapter Two

Review of Selected Literature

Introduction

Chapter II contains a review of related literature. An introduction to the effectiveness of retention on academic achievement is considered along with the characteristics of students who are retained. Also, the factors that are associated with retention are reviewed. This review of literature is not intended to be exhaustive, rather it is intended to provide the most relevant and current research regarding retention practices.

Student retention refers to a practice in which a student is required to repeat a particular grade year as a result of delayed academic progress (Rafoth, 1991). In accordance with this concept, students who need support services are provided extra time and opportunities to learn basic academic skills. Retention has been used as an intervention for students with academic difficulties for many years (Jimerson, 2001). However, due to increasing numbers of students in need of extra instructional time, more students are being retained (Jimerson & Kaufman, 2003). According to the National Association of School Psychologists (2003a) 15% to 20% of students are retained each year and 30% to 50% of students are retained before ninth grade. Jimerson (2001) offered more conservative estimates of 5% to 10% students retained each year. Other studies

have indicated a wide range in the rate of student retention (Fine & Davis, 2003; McCoy & Reynolds, 1999).

Interest in the effects of retention on academic and behavioral outcomes has increased among researchers in the past ten years due, at least in part, to many politicians and educators supporting the end of “social promotion” (Jimerson, 2001). Social promotion is regarded as promoting students in spite of delayed academic performance. In 1999, President Clinton called for an end to social promotion in his State of the Union address. This theme was continued in 2002 with the No Child Left Behind (NCLB) Act. NCLB set educational guidelines and required states to be accountable in order to obtain federal funding. NCLB particularly focused on student reading proficiency. Therefore, the goal set in NCLB was for every student to be proficient in reading by the 2013-2014 school year. Many states have responded to NCLB by incorporating evidence-based reading interventions for students who are not obtaining adequate yearly progress. Student retention is one of the interventions that has been utilized.

Retention is defined as a practice that requires a student to repeat a grade designation for one subsequent year (Jimerson, 2001). Retention is an expensive intervention, adding the cost of one extra year for each student retained (Pagani, Tremblay, Vitaro, Boulerice, & McDuff, 2001). According to Dawson (1998) retention costs 14 billion dollars a year. This may be the reason, at least in part, to the insufficient funding for extra support services for students who are retained. The question for many researchers is if the cost of retaining students is worth the academic outcomes. Do retained students demonstrate significantly higher levels of academic achievement meriting the cost? The methods that researchers have utilized to examine the impact of

grade-retention on academic performance has traditionally compared aggregated data of outcome measures of students who were retained to students who were not retained.

Researchers have utilized prospective longitudinal designs, meta-analyses, and qualitative methods to investigate grade retention.

Research regarding the effect of retention on academic and behavioral outcomes has been inconclusive. According to Tomchin and Impara (1992), teachers view retention as a useful strategy to increase mastery of academic tasks. A handful of studies have found that retention is linked to limited improvement in self-esteem and mathematics scores (Alexander, Entwisle, & Dauber, 1994; Gottfredson, Fink, & Graham, 1994; Mantzicopoulos & Morrison 1992). Pagani et al. (2001) indicate that the positive effects of retention tend to fade out over time. A larger number of studies have linked negative academic and behavioral outcomes to retention (Jimerson, 2001; National Association of School Psychologists, 2003a).

Characteristics of Retained Students

As would be expected, low academic achievement is common among students who are retained (Jimerson, 2001). Additionally, research indicates that a student is more likely to be retained if the student is male, an ethnic minority, and of low social economic status (Abidin, Golladay & Howerton, 1971; Niklason, 1984). Retained students are typically considered to be younger than their grade-level peers, however, studies addressing the age of students who are retained have been inconclusive (Shepard & Smith, 1987).

Jimerson, Carlson, Rotert, Egeland, and Sroufe (1997) identified many additional variables that are associated with retained students. Their study included 179 participants

(80% Caucasian, 14% African American, and 6% Hispanic) in three groups: a retained group (n = 29), a low achieving group (n = 50), and a control group (n =100). The results of the study indicated that males are more likely to be retained than females ($M = 6.76$, $p < .05$). The participants in the retained group were comprised of 74% males, while the participants in the low achieving but promoted group were comprised of 56% males. The results of the study also indicated that ethnic minority students were more likely to be retained than Caucasian students and that low social economic status (SES) was a risk factor for grade retention. Specific internal student characteristics were also identified as being associated with retention. These included lower cognitive ability, poor peer relations, and high rates of absenteeism.

A study conducted by Graue & Diperna (2000) corroborated the findings of Jimerson et al. (1997). Results indicated that males were retained more often than females, and that minorities were retained more often than Caucasians. In addition, students whose birthdays were close to the school entrance cutoff were more likely to be retained.

Negative Outcomes of Student Retention

As previously stated, retention is used for students with delayed academic performance. However, a great number of studies indicate that it is not an effective strategy. A study conducted by Pagani, et al. (2001) focused on the academic and behavioral outcomes of retained students. The study included 1,830 students who were randomly selected from a larger pool of 6,397 participants. Each of the participants selected were followed until the age of 12. The independent variables in the study included retention and gender. The study included teachers' ratings of students overall

academic performance, and the teachers completion of the Social Behavior Questionnaire. The questions on the Social Behavior Questionnaire are derived from the Preschool Behavior Questionnaire and the Prosocial Behavior Questionnaire. The data were analyzed using a basic autoregressive model. This model allows for the control of changes that would be expected from students with different achievement levels. Therefore, achievement level trajectories are held constant.

Analysis of the data suggested that retention had a negative affect on the academic trajectory of both girls and boys. Specifically, after being retained between the ages of 6 and 8, boys showed signs of negative academic trajectories at the times of follow up (e.g., ages 10 and 12) relative to students who were not retained (path = $-.12$, $p < .01$). Girls displayed similar negative academic effects at age 10 (path = $-.07$, $p < .01$) and at age 12 (path = $-.07$, $p < .05$). Retention also seemed to have a negative effect on the behavioral trajectories of the boys who participated in the study.

The authors suggested that the negative behavioral effects may have caused the negative academic effects of retention and it was difficult to separate the two variables. The study relied on overall student performance ratings from teachers to assess academic achievement. Teachers' ratings may not have been reliable, and the ratings may not be an accurate representation of actual student academic performance. Finally, while the study's trajectories were found to be statistically significant, the sample size was large. Therefore, the statistical significance may not indicate a large effect size designating real world significance.

Another longitudinal study conducted by McCoy and Reynolds (1999) indicated similar results. Data were analyzed from the Chicago Longitudinal Study to determine

the academic outcomes of students who were retained. The participants included 1,164 low-income, mostly African-American seventh- and eighth-grade students, of which 315 had been retained. Of the students retained, 296 had been retained once, and 19 had been retained more than once. The study included four outcomes measures: a) reading comprehension, b) mathematics achievement, c) perceived school competence, and d) school-reported delinquency. The participants were given the Iowa Test of Basic Skills (ITBS) mathematics and reading comprehension subsections at the age of 14.

The students' scores were analyzed using a hierarchical multiple regression model. Retention significantly predicted student reading comprehension ($R^2 = .47$) and mathematics performance ($R^2 = .57$) when comparing same age peers. According to the study, retention explained 47% of the variability in the participants' ITBS reading comprehension scores and 57% of the variability in the participants' ITBS mathematics scores. After the researchers included demographic factors (e.g., gender, parent education, free-lunch eligibility, and SES) and early adjustment indicators (e.g., classroom adjustment, first-grade reading and mathematics achievement), retention predicted lower mathematics scores ($ES = -.481, p < .001$) and reading comprehension scores ($ES = -.424, p < .001$). The time at which students were retained (e.g., early vs. late retention) seemed to have an impact on reading scores. Early retention (grades 1-3) was associated with lower reading achievement than later retention (grades 4-7). The authors did note, however, that more than 50% of the differences between the groups were explained by other factors.

In addition to comparing same age peers, the authors compared seventh- grade retained students with their same grade peers. Again, analysis indicated that retention

predicted negative reading comprehension ($\beta = -4.6$ standard score points, $p < .001$), however, the prediction was not significant for the mathematics ITBS scores. According to the authors, the results of this study indicate that retention is not an effective intervention for low achieving students. Retention did not seem to benefit students more than other less expensive alternatives.

A similar study conducted by Jimerson (1999) attempted to answer the question “To retain or not to retain?” (p 243). This study set out to determine the association between grade retention and high school academic achievement, later high school dropout rates, and post-secondary education. Participants in the study fell into three groups including a retention group ($n = 29$), a low-achieving, promoted group ($n=50$), and a control group ($n = 100$). The retained group had a larger number of males than did either the low-achieving, promoted group or the control group. The percentage of minority students was highest in the retained group (35%), followed by the low-achieving, promoted group (31%) and finally the control group (16%). Students in the retained group were retained once in kindergarten through third grade. The researchers conducted teacher interviews, child interviews, child testing, parental interviews and parental testing (e.g., Wechsler Adult Intelligence Test, Home Inventory) during the early childhood years for participants in the retained group, low-achieving, promoted group, and the control group. In addition, mother and home assessments were collected before birth and soon after birth, at 30 and 48 months, and at first grade. This information included SES, age of the mother at the child’s birth, education completion, and intelligence assessment. The participants were followed into adolescence (eleventh grade) and early adulthood (19 and 20-years-old). Several outcome measures were investigated

including high school achievement as measured by grade point average, and attendance, high school completion, and post-secondary education enrollment.

The results of the study indicated that the retention group had significantly lower high school achievement than the low-achieving, promoted group ($F = 6.59, p < .01$) and the control group ($F = 13.95, p < .001$). The participants in the retention group were also more likely to drop out of high school ($F = 3.57, p < .05$), and were less likely to receive a certificate for high school completion ($F = 5.44, p < .01$) relative to the low achieving, promoted group. The researcher suggested that the results of the study indicated that retention is not an effective early intervention practice. Students who were retained in early primary grades were more likely to eventually drop out of school. However, the additional academic support that the retention students received during the retention year was not assessed or discussed. Students who were retained may not have received any academic remediation interventions.

An earlier study conducted by Jimerson et al. (1997) examined the effects of retention on achievement in elementary school and at sixth grade. The study included 179 students from Minnesota that consisted of three groups: a retained group, a low-achieving, promoted group, and a control group. The retained group consisted of 29 students who were retained in either kindergarten, first, second, or third grade. The participants in the low achieving, promoted group were selected because their academic performance was similar to that of the participants within the retention group. These students were identified by Peabody Individual Achievement Test (PIAT) scores that fell within the bottom quartile of the entire sample. Finally, the control group consisted of 100 participants randomly selected from a larger pool of participants who were not

eligible for the retained or the low-achieving, promoted groups. The participants were enrolled in kindergarten (25), first (25), second (25) and third (25) grades. The study utilized several outcome measures to compare the three groups. Each participant was given each of the measures during the primary grade of the student and at sixth grade. The outcome measures included teacher interview measures, attendance reports, the Child-Behavior-Checklist—Teacher Edition (CBCL-T), peer acceptance measures, the PIAT, the Wechsler Preschool and Primary Scales of Intelligence (WPPSI), the Wechsler Intelligence Scale for Children-Revised (WISC-R), the Woodcock-Johnson Achievement Test-Revised (WJ-R), maternal interviews, and a life events inventory.

The results of the study reported both the short-term effects and long-term effects of retention. The short-term effects of retention were measured by running contrasts comparing the PIAT scores of the retained and the low achieving, promoted groups the next school year. The age at which the participant was retained was also considered. Participants who were retained in kindergarten were compared with same aged peers after the completion of first grade. After controlling for previous achievement, contrasts indicated that the PIAT math, reading comprehension, and spelling scores of students who were retained during kindergarten and low achieving, promoted students were not statistically significant. This indicates that retained students academic achievement did not improve relative to same age peers. Similarly, the PIAT reading comprehension and spelling scores of first and second grade retained group did not differ from the low achieving, promoted group. However, retained students did have significantly higher PIAT math scores than low achieving, promoted students ($F = 6.05, p < .05$).

In addition to examining the short-term effects of retention the authors also examined the long-term effects of retention on academic achievement. The long-term effects of retention were measured by total PIAT score performance for each group at the completion of sixth grade and WJ-R total scores at age 16. The contrasts between the retained and low achieving, promoted groups PIAT scores at the completion of sixth grade were not significant. Furthermore the contrasts between the retained and the low achieving, promoted groups WJ-R scores were not significant. The results of this study indicate that students who were retained did not perform significantly better than their peers when prior achievement was controlled. Short-term positive effects for math were evident for first and second grade participants, however these effects washed out after the completion of sixth grade. Critiques of this study are that it included a small sample size and omitted a discussion regarding academic interventions employed during the retention year. It is possible that students who were retained received no additional academic support services and completed another year of curricula that they already had received.

A meta-analysis conducted by Jimerson, et al. (2002) examined 17 studies in order to determine the relationship between retention and high school drop-out. Each of the studies that identified retention as a potential predictor of later high school dropout indicated that retention was significantly associated with later high school dropout. Furthermore, retention was consistently one of the most powerful predictors of dropout. Moreover, students who were retained more than once are more at-risk for later dropout. The studies that were reviewed indicated that students who were retained once were 40 to 50% more likely to drop out of high school and students who were retained more than once were 90% more likely to drop out of high school when compared with promoted

peers. The authors conclude that retention should not be considered a direct cause of dropout. Rather, retention and other factors such as low SES, immaturity, and low achievement place students at risk for future dropout. The relationship between retention and later high school dropout is transactional. Retention leads to other negative conditions such as absenteeism, low school engagement, and low self-esteem contributing to later dropout.

A recent study by Fine and Davis (2003) investigated the long-term effects of grade retention. Specifically, the authors were interested in the relationship between grade retention and later post-secondary education enrollment. The study included 11,637 (5605 male, 6031 female) participants derived from the National Educational Longitudinal Survey (NELS) database. The data used for this study were gathered in 1988, 1992, and again in 1994. Retention status for each participant was determined by responses to survey questions. Likewise, the outcome measure, post-secondary enrollment, was determined by participant responses to surveys in 1994. The participants were matched on demographic variables, SES and academic achievement. Odds ratios reported the likelihood of the outcome measure among students with different characteristics. Each of the regressions used promoted students as the comparison group.

The results of the study indicated that boys were almost twice as likely to be retained than girls with an odds ratio of 1.89 ($p < .01$) and students with low SES were almost twice as likely to be retained as high SES participants with an odds ratio of 1.87 ($p < .01$). However, males were slightly less likely to be retained more than once (odds ratio = .88). Students who were retained were one-half less likely to enroll in a four-year college (odds ratio = .47, $p < .01$) when compared to promoted students. Retained

students were also less likely to enroll in any type of post-secondary education (odds ratio = .617, $p < .01$). Interestingly, students who were retained more than once were more likely to enroll in a four-year college (odds ratio = .75) than students who were retained once (odds ratio = .45) when compared to promoted peers. However, students who were retained once were more likely to enroll in any type of secondary education (odds ratio = .64) than students who were retained more than once (odds ratio = .37) when compared with promoted students.

Overall, the results of this study indicate that retention has negative long-term effects on students. The author suggests that even when retained students overcome the odds and graduate from high school the effects of retention are still evident. Students who were retained were significantly less likely to enroll in post-secondary education than their promoted peers. According to the investigators, teachers, school staff and school psychologists should be cautious when recommending retention for students because the effects of retention may not be evident until early adulthood. Limitations of this study include the age of the NELS database. More current data would have been more desirable. Also, there was no control for the quality of instruction, interventions, and curriculum to which retained and promoted students were exposed. Finally, this study did not include students who did not graduate from high school, but who did obtain their GED. These students could have impacted the results of the study.

A meta-analysis performed by Jimerson (2001) attempted to provide a thorough review of 20 studies published between 1990 and 1999. The author's goal was to summarize the most current research on retention, and to recommend alternatives to both

retention and social promotion. Each study was examined by (a) determining the variables used to match retained students with comparison peers, (b) determining the outcomes associated with the grade at which students were retained, (c) examining the statistical outcomes of retention on academic achievement, and (d) the author's conclusions regarding the use of retention as an intervention for academic difficulties. The studies reviewed matched students on various variables including IQ scores, previous academic achievement, SES, and sex. Most studies compare students who were retained with promoted students and measured academic achievement by relative gains on norm-referenced achievement tests.

The results of the analysis indicated that retention had a negative effect on the academic outcomes of the participants. Included within the 20 studies were 91 statistically significant analyses, of which 82 favored low achieving, promoted students over retained students. In addition, 84 statistical analyses on academic achievement were not significant. Subtracting the mean of the retained group from the mean of the comparison group and dividing by the standard deviation of the comparison group calculated the effect sizes (ES) for each academic achievement analyses. The effect sizes ranged from -2 to 1.25 . The overall ES was $-.39$ indicating that on average, the retained group scored $.39$ standard deviation units lower than the comparison group. The author concludes by suggesting that neither grade retention nor social promotion will remediate the academic difficulties of students. The costs that are associated with retention do not justify its use. Professionals and researchers should begin to focus on alternative strategies to enhance academic achievement such as prevention and early intervention, and progress monitoring.

Positive Outcomes of Student Retention

Although a large portion of literature contends that retention negatively affects student achievement, some research has documented positive effects of retention. A study conducted by Mantzicopoulos (1997) investigated the long-term academic effects of retention on kindergarten students with attention problems. The study included 40 participants (28 males, 12 females) of which 25 were retained and 15 were promoted. Each student had attention difficulties as measured by teacher interviews and was matched for school, sex, at-risk status, reading achievement, and math achievement. Two measures were used to assess achievement: the Stanford Achievement Test, and the California Test of Basic Skills (CTBS). Specifically, reading and mathematics achievement were assessed at the end of the kindergarten, first and second grade.

The results of the study included comparisons of same-grade participants and same-age participants. Same-grade comparisons indicated that retained students earned higher mathematics achievement scores ($F = 5.63, p < .05$). However, same-grade comparisons of reading achievement scores did not favor retained or promoted students. Same-age comparisons for mathematics achievement was also significant ($F = 4.95, p < .05$) indicating that retained students outperformed their same-age peers on the mathematics achievement tests. Same-age comparisons of reading achievement did not yield significant results. Adjusted mean differences were calculated for both retained and promoted students from national averages. Retained students' performed higher than the national average in mathematics ($M_{Ad} = .21, .41$ for first and second grades respectively) while the promoted group remained below the national average. Adjusted mean differences were also calculated for reading achievement scores. After an initial

improvement in the first grade ($M_{Ad} = .72$), mean differences decreased by second grade ($M_{Ad} = .18$) for retained students. This decrease was not evident for promoted students.

Based on the results of this study the author suggested that retention does not benefit students as an intervention for delayed academic progress. This assertion is made because the participants reading achievement scores improved the first year after retention, but faded by second grade. Retention did however seem to benefit the participants' mathematics achievement scores. Even at the end of first grade, retained students mathematics achievement scores remained above the national adjusted average, and were significantly higher than same-age and same-grade peers. As with most studies investigating retention, this study did not control for instructional strategies that were used during the retention and subsequent school years. The decline in reading achievement score gains could be explained by discontinued academic interventions after the retention year. Additionally, the sample size used in this study was relatively small, making it difficult to generalize the results to other populations.

While Mantzicopoulos (1997) offered limited evidence of the possible positive effects of retention, Alexander, et al. (1994) conducted a longitudinal study with more conclusive positive findings. The study consisted of a stratified random sample of 800 children in Baltimore. At the beginning of the study, each of the participants were entering first grade. Each were then followed for eight subsequent years. At the end of the first year of the study, 127 first-grade students were retained. The authors noted that some students who were retained were later promoted mid year ($n = 17$), or were double promoted ($n = 12$). By the eighth year of the study, 142 children were at least one year behind grade level (7th grade). Data were collected on each participant via test scores,

grade reports, interviews with the participants and their parents, and questionnaires completed by teachers. In analyzing the data, the researchers controlled for prior achievement, demographic variables (race, SES), parent education, and school readiness. Three groups were compared: the retained group, the low achieving, promoted group, and the rest of the students who were promoted. In addition to comparisons, the retained groups academic trajectory was determined by comparing pre- and post-retention academic achievement.

As expected, the participants who were retained at end of first grade had significantly lower test scores than did their same age non-retained peers at the beginning of first grade ($M = -33.5, p < .01$), and at end of first grade ($M = -59.5; p < .01$). After the completion of the retained year, the achievement gap was less between the retained group and their same age non-retained peers. The retained group participants gained seven points ($p < .01$) on reading test scores and 4 points ($p < .01$) on math test scores when compared to non-retained same age peers. Comparisons of math test scores between retained students and same-grade non-retained students at the completion of the retention favored the retained students with a 17-point relative gain ($p < .01$). However, this gap decreased after the initial follow up. Students who were retained in second grade seemed to fair better than the participants who were retained in the first grade. The second grade retained students gained on average 17.3 test score points relative to their same age peers. Students who were retained in third grade seemed to show the most achievement gains. At the end of the initial third grade year, the retained students on average scored 28.4 points lower than their same-age peers. At the end of the retention year, the retained student's test scores were on average 15.9 points lower than their same-

age peers, and by the end of seventh grade, the retained student's scores were on average only 9.7 points lower than their non-retained peers.

The results this study indicated, that on average, the students who were retained in the third grade had better outcomes than the students who were retained in the first grade. The authors argue that if the retained students would not have been retained, the achievement gap between them and their same-age peers would have widened rather than decreased. Although this may be the case, other factors may have influenced the results of the study. No attempt was made to determine whether supplemental services were provided to the students who were retained. The gains in tests scores could be attributable to intense remediation interventions that were provided during the retention year rather than merely repeating a particular grade.

The State of Florida's Retention Model

According to the Florida Department of Education, Florida's retention practices are different from those of the past in two key domains. In the past, retention has included repeating a grade with no emphasis on supplemental services. Students experienced the same materials, instruction, and teachers for an additional year. Moreover, past models of retention were conceptualized as an intervention to remediate students' academic skills. It has been a response to academic failure and if retained, students will catch up to their same-grade peers. The state of Florida's retention policy is focused on (a) providing students with increased amounts of time to engage in academic instruction and (b) preventing the academic failure of students.

Supplemental Instruction

According to the Florida Department of Education (2002a) each retained student is required to have an individualized academic improvement plan (AIP). This program delineates the type, difficulty, amount, and intensity of instruction each retained student needs in order to reach academic standards. Many times, AIPs specify the use of supplemental instructional services for students. Supplemental instruction is that which is beyond what students typically obtain during allotted instructional time. Some examples are as follows: reading instruction is increased from 45 minutes to 90 minutes, other types of instruction (e.g., mathematics, art) are suspended giving more time for reading instruction, and the use of after-school tutors. Providing more allocated time to students may result in more opportunities for the retained student to be engaged in instruction and academic activities than their peers. The extra time allows for at-risk students to close the achievement gap (Aronson, Zimmerman, & Carlos, 1998; Nelson, 1990).

In addition to providing more time, supplemental instruction is intended to provide students with intensive evidence-based academic interventions matched to students' individual needs. Numerous studies have documented the effectiveness of supplemental instruction in promoting academic achievement (Gredler, 1997). A recent meta-analysis conducted by Jimerson, Kaufman, Anderson, Whipple, Figueroa, Rocco, & O'Brien, (2002) provides a comprehensive review of academic interventions in an attempt to persuade educators and professionals to move beyond discussing retention and social promotion and focus on supplemental instruction strategies.

Early Intervention

In addition to supplemental instruction, early intervention is a critical component to Florida's model of retention. Students are retained in third grade, before what many call the curriculum shift at fourth grade. The retention gate is set at third grade forcing educators and teachers to focus academic support on grades 1-3. The overall goal is to *prevent* students from being retained in third grade. Therefore, more resources are given to early intervention in order to prevent academic failure. Early intervention research has suggested that reading difficulties are more easily prevented than remediated (Coyne et al., 2001, National Association of School Psychologists, 2003b), and the remediation of reading difficulties is most successful if interventions are employed early in the development of the problem (Haager & Windmueller, 2001, Jenkins & O'connor, 2002; Phillips, et al., 2002).

A study conducted by Lennon & Slesinski, (1999) evaluated the impact of early intervention on later reading development. The study included 156 students who were assessed in reading based on their letter-naming proficiency. The participants in the study fell into three groups; low-scoring ($n= 80$), middle-scoring ($n= 56$), and high scoring ($n= 40$). Students in the low-scoring group were randomly assigned into two subgroups: 1:2 tutoring during 20 weeks (low-scoring A), or 1:2 tutoring during the second 10 weeks (low-scoring B). The middle-scoring group was also assigned to one of two tutoring sessions (middle-scoring A and middle-scoring B). The high-scoring group did not receive tutoring but served as a control group. Each tutoring session lasted 30 minutes, 5 times a week and consisted of explicit instruction in letter naming, letter sounds, phoneme segmentation, the alphabetic principle, print awareness and sight words. The

outcome measures assessed letter naming fluency, letter sound, phoneme segmentation, decoding fluency, sight words, and concepts of print. Each of the outcomes was measured during baseline, after 10 weeks, and again after 20 weeks.

The results indicated that all students who received instruction during the first 10 weeks performed better on each of the outcome measures than did students who waited 10 weeks for tutoring. The low-scoring group A outperformed the low-scoring group B on letter naming (ES = .63), letter sounds, (ES = .98), decoding (ES = .69), phoneme segmentation, (ES = .67), sight words (ES = .78), and concepts of print (ES = .67) outcome measures. Results also indicated that middle-scoring participants benefited from receiving instruction during the first 10 weeks when compared to their peers who waited 10 weeks for tutoring on each outcome measure. Finally, results indicated that after intervention the low-scoring A subgroup performed similarly to the middle-scoring group did at baseline. Moreover, the middle-scoring A subgroup performed similarly to the high-scoring group did at baseline. The researchers concluded that early intervention promotes academic achievement for low performing students, and average performing students.

Conclusion

In conclusion, a preponderance of research studies on retention does not find support for retention as an effective intervention for the remediation of academic delays of primary grade students. Furthermore, studies examining the long-term effects of retention suggest that students who are retained are more likely to drop out of school, and are less likely to obtain post-secondary education (Jimerson, Anderson, & Whipple, 2002). In addition to negative academic effects, research has suggested that retention has

negative behavioral and emotional effects (National Association of School Psychologists, 2003a). A few studies (Alexander, et al., 1994; Gottfredson, Fink, & Graham, 1994; Mantzicopoulos, 1997, Peterson, DeGracie, & Ayabe, 1987) offer some evidence that in some instances retention can help promote the academic achievement of students. However, these academic gains are often reported to be short-term (Pagani, et al., 2001).

According to researchers (e.g., Ferguson et al., 2001; Jimerson, 2001; Jimerson et al., 1997) research that has examined the effects of student retention on academic achievement have common experimental design flaws. Isolating the effects of retention is difficult because one cannot randomly assign groups of students to be either promoted or retained and examine the achievement differences. Therefore, many studies are quasi-experimental attempting to control for potentially relevant variables (e.g., demographic characteristics, prior achievement). A second potential flaw in literature is the definition used by researchers for retention. Retention is broadly defined within the literature as a practice that requires any student to repeat a grade as a result of academic difficulties. This broad definition makes it difficult to determine what characteristics of the retention year are potentially effective or ineffective. The definition does not control for the quality of instruction and interventions that students engage during the retention year. No study could be found that addressed the types of services that were offered during the retention year. Some students may have received intensive academic interventions during the retention year, while others may have been exposed to the same curriculum that they had received the previous year.

Florida's model of retention can be considered to be retention with reading remediation. Students who are retained are provided with AIPs that are individually

developed for each student and provide evidence that retained students received supplemental services during the retention year. In addition to individualized AIPs, students are provided with extra opportunities to become proficient in reading due to the extra time afforded by retention.

Chapter Three

Method

Introduction

Chapter III contains information regarding the method and procedures that were used in the present study. Specifically, the population and sample, instrumentation, data collection, and data analysis are discussed.

Sample

The sample for this study was drawn from the total population of third- and fourth-grade students in the public schools of the state of Florida who took the FCAT reading Test in 2003 and 2004. In the state of Florida, 28,028 third-grade students were retained during the 2002-2003 school calendar year (Florida Department of Education, 2004), however the final sample of students used for the current study consisted of two select groups of retained and promoted third-grade students.

Higher Performing, Retained Sample. Third-grade students whose 2003 FCAT reading score fell just “under” (score of 248-258) the cut score required for promotion to fourth grade and who were retained in third grade for the 2003-2004 school year constituted the retained group. These students scored at a Level 1 on the 2003 FCAT Reading Test. According to the Florida Department of Education (2004) students who score at Level 1 on the FCAT-reading test will experience limited success with the Sunshine State Standards Curriculum. Therefore, these Level 1 students are retained

because it is believed that they will not benefit from fourth-grade instruction. According to Florida educational guidelines, any student who exhibits a substantial reading, writing, math, or science deficiency must have an Academic Improvement Plan (AIP). The purpose of an AIP is to ensure an individualized intervention plan for each student. These plans are required to identify specific areas of deficiency (e.g., fluency, phonemic awareness, comprehension) and include instructional and environment interventions designed to remediate academic deficiencies in students. Examples of AIP interventions are (1) the use of a peer tutor, (2) one-on-one instruction with a reading coach, and (3) 30 minutes extra time for reading instruction. AIPs also include measurable academic goals that are linked to previous assessment. Teachers and school staff are required to monitor the progress of these goals on a frequent basis. Retained students as well as at-risk students are required to have an AIP.

Promoted, Low Achieving Sample. This sample consisted of students whose 2003 3rd grade FCAT reading score ranged from 259-269, who received a Level 2 designation and were promoted to the fourth grade for the 2003-2004 academic year. For the purposes of this study, this sample of students is referred to as the “promoted, low achieving group”. Students who obtained an FCAT reading score just above the required promotion score received a Level 2 designation. According to the state of Florida, these students will experience little success with the Sunshine State Standards curriculum, but possess the skills necessary for promotion to fourth grade. FCAT reading scores from 12,948 students (grades three and four) in the Florida Department of Education database constituted the final sample. The score of each student who took the 2002-2003 and

2003-2004 FCAT was archived in the FLDOE database, along with the gender, race, and size of district that the student attends.

Several students were excluded from the study due to various reasons. Students who scored just below the required FCAT reading score for promotion but were promoted because of one of the previously stated good cause exceptions were not included. Students who attended lab schools in the State of Florida were excluded, as well as students who attended home school. Finally, students with missing data were also excluded from the sample.

Instruments

The Florida Comprehensive Assessment Test (FCAT) is a criterion-referenced test developed by a panel of curriculum specialists from the Harcourt Educational Measurement Company (Florida Department of Education, 2003). It was developed to assess student achievement of the higher-order cognitive skills represented in the Sunshine State Standards (SSS). The FCAT reading 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 (Florida Department of Education, 2003). 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. During the months of February and March, the FCAT is administered to over 1.5 million students. The tests are then sealed and sent to the Florida Department of Education. The FCAT contains both multiple choice and performance questions. The multiple choice questions are scored by computers while the performance tasks are hand scored.

Scoring of the FCAT is based on item response theory (IRT, Lord & Novick, 1968). The 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 the fourth-grade reading portion of the FCAT scores was reported by total score ($r = .88$), literary text ($r = .79$) and informational text ($r = .79$) (Florida Department of Education, 2002b). The third-grade reliability coefficients were not provided. Unfortunately, score reliability estimates for the proposed sample will not be available. Construct, criterion, and content validity coefficients of the FCAT could not be found. However, the items on the FCAT were reviewed by the Florida Department of Education for style, content and match to SSS benchmark. Community sensitivity committees, bias committees, and content committees then reviewed the FCAT items (Florida Department of Education, 2002b). Currently, the only resource that has provided the technical characteristics of the FCAT is provided by the state of Florida (Florida Department of Education, 2002b).

The possible range of scaled scores on the FCAT reading achievement test is 100 to 500 (Florida Department of Education, 2003). Based on these scaled scores students are placed within one of five levels. Each level represents a different level of proficiency in reading. In 2003, Level 1 scores fell within the scaled scores of 100-258, Level 2 scores fell within the 259-283 range, Level 3 scores fell within the 284-331 limits, Level 4 scores fell within the 332-393 range, and Level 5 scores fell within the 394-500 limits. In reading, a student who achieves a Level 1 score on the FCAT is predicted to experience limited success with the content of the Sunshine State Standards. A Level 2 score represents little success (Florida Department of Education, 2003). Currently,

Florida Statute (Florida Department of Education, 2002a) requires that students in third grade who scored Level 1 in Reading must be retained (with noted exceptions).

Procedure

A proposal for the current study was submitted to the Institutional Review Board (IRB) at the University of South Florida and the FLDOE for approval before any data were analyzed. After the approval of the IRB and the FLDOE, the procedure for the present study was carried out in the following manner.

Step 1: The primary investigator identified the potential pool of third- and fourth-grade students who obtained 2003 3rd-grade FCAT reading scaled scores of 248-269 from the data provided by the Florida Department of Education.

Step 2: Classifications of size for each school district in the state of Florida was obtained from the Florida Department of Education. The size classifications are Very Large, Large, Medium, Small/Medium, and Small.

Step 3: Participants who obtained a scaled score of 248-258 on the 2003 3rd grade FCAT reading test, obtained a Level 1 designation, and were retained in third grade due to academic reasons were selected. This group of students was designated as the “higher performing retained group”. The 2004 3rd grade FCAT reading scores for the higher performing retained group were then obtained in the spring of 2004.

Step 4: Participants who received a scaled score of 259-269 on the reading portion of the 2003 3rd grade FCAT, obtained a Level 2 on the 2003 FCAT , and were promoted to fourth grade due to academic reasons were placed in a “promoted, low achieving group”. These participants represent students who received the lowest scores on the FCAT reading test of Level 2 students and were promoted to fourth grade. The

2004 4th grade FCAT reading scores for the promoted, low achieving group were then obtained in the spring of 2004 from the FLDOE.

Step 5: Students who attended school at home, attended laboratory schools, or who had missing data were excluded from both samples. In addition, students with data that were not consistent with Florida's student progression plan (e.g., promoted due to academic reasons, attained Level 1 on the FCAT) were also excluded from the study.

Step 6: The size of each group was determined upon analysis of third graders' 2003 FCAT reading scaled scores. As predicted, there were an adequate number of participants in each group to conduct inferential statistical procedures on the data. The 3rd grade higher performing retained group consisted of 3,886 students and the 4th grade promoted, low achieving group consisted of 9,062 students.

Step 7: The information from the database obtained from the FLDOE were transferred into the Statistical Analysis Software (SAS, 2004) by the primary investigator for analysis.

Step 8: The alpha significance level used for the present study was delineated at .05, two-tailed. The rationale for a two-tailed analysis is that it was not certain what the effect retention will have on students FCAT performance. Therefore analysis of the variables were sensitive to both negative (lower FCAT attained Levels) and positive (higher FCAT attained Levels) outcomes.

Data Analysis

1. What proportion of higher performing retained third-grade students who had reading scaled scores of 248-258 on the 2003 FCAT reading test subsequently scored at Level Two or higher on the 2004 3rd-grade FCAT reading test?

This research question was addressed from the data obtained from the 2003 and 2004 3rd grade FCAT reading scaled scores for students for the higher performing retained group. The demographic characteristics of the groups were reported in terms of race, gender, and size of school district. The number and percent of the 3rd grade higher achieving retained students who scored at Levels 1 through 5 on the 2004 3rd grade FCAT reading test were computed. In addition, the percentages of these retained students who attained a Level 2 or higher and a Level 3 or higher on this test were reported. A bar graph was developed to visually communicate the percent of higher performing retained students who scored at Level 2 or higher and those students who scored at Level 1 on the 2004 3rd-grade FCAT reading test.

2. There is a significant difference in the obtained 2004 FCAT reading levels of higher performing retained students by a) gender, and by b) race/ethnicity.

Descriptive statistics for each subgroup (e.g., retained male, female, African-American, Asian, Caucasian, and Hispanic students) were computed including means, standard deviations and skewness and kurtosis coefficients for the 2003 FCAT reading scaled scores. The percentages of the 2004 FCAT Level designations by gender were reported for higher retained students. Chi-square procedures were used to determine if the obtained proportion of higher performing retained male and female students who scored at Levels 1, 2, and 3 through 5 on the 2004 3rd-grade FCAT reading subtest was significantly different than the expected proportions. Similarly, the percentages of the 2004 FCAT level designations were reported for higher performing retained African-American, Asian, Caucasian, and Hispanic students. Chi-square procedures were also used to determine if the obtained proportions of higher performing retained African-

American, Asian, Caucasian, and Hispanic students who scored at Levels 1, 2, and 3 through 5 on the 2004 3rd-grade FCAT reading subtest was significantly different than the expected proportions.

3. There is no difference in the obtained post-retention 3rd-grade FCAT reading levels of higher performing retained students as a function of district size.

Data from 2003 and 2004 FCAT reading scores, retention group classification, and the size of students' attended district were examined. District size classifications that were used in the analysis were Very Large, Large, Medium, Medium/Small, and Small, following the criteria that are used by the FLDOE for such designations. Descriptive statistics included the 2003 FCAT reading score means and standard deviations from each district size designation were calculated for the higher performing retained students. Skewness and kurtosis of FCAT scores of these retained students by size of attended district were also calculated. The percentages of higher performing retained students who attained Levels 1, 2 and 3 through 5 designations on the 2004 3rd-grade FCAT reading test were calculated and reported. A chi-square measure of association analysis was conducted for the higher performing retained group to determine if the obtained proportions of retained third-grade students who achieved state reading standards significantly differed from the expected proportions relative to the size of district attended.

4. For higher performing students retained in the third grade, there is a significant relationship among gender, race/ethnicity, prior performance on the FCAT reading test and attaining state reading standards on the 2004 3rd-grade FCAT reading test.

To test this hypothesis, the data from the higher performing retained student sample were subjected to a logistic regression procedure. The logistic regression statistical procedure allows for a dichotomous outcome variable (achievement of state reading standards vs. non-achievement of state standards), and both dichotomous (e.g., male and female) and continuous (2003 FCAT scaled scores) variables as predictor variables. In the present study, gender, race/ethnicity, and prior performance on the 2003 FCAT reading test were entered into the model to predict meeting state standards for reading achievement in 2004 as measured by the 2004 3rd-grade FCAT reading level designations for higher performing retained third-grade students. The nominal variables were coded to allow for within group comparisons among the higher achieving retained male students and likewise for the promoted, low performing students. The coding was as follows: (a) Males were coded as 0 and female students were coded as 1, (b) the race/ethnicity variable was dummy coded so that African-American, Asian, and Hispanic students were compared to Caucasian students, (c) student who received the FCAT reading Levels 2 through 5, which was considered to have met state reading standards, was coded as a 1 and the FCAT Level 1 designation which does not meet state standards and was coded as a 0, and (d) students who received free and reduced lunch in the 2003-2004 school year were coded as a 0, and students who did not receive free and reduced lunch during the 2003-2004 were coded as a 1. The overall likelihood ratio for the model was reported to determine if the model predicted retained student outcomes better than chance along with the Hosmer and Lemeshow goodness-of-fit statistic which indicates if model is a good fit for the data. The weights and standard error for each of the variables were reported and odds ratios for race/ethnicity, gender, and size of district attended for higher performing

retained third-grade students were also calculated and reported to determine if retention differentially benefited groups of students.

5. What proportion of promoted third-grade students who had reading scaled scores of 259-269 on the 2003 FCAT reading test subsequently scored at Level 2 or higher on the 2004 4th-grade FCAT reading test?

To analyze this research question, descriptive statistics were reported including the means for each group including the standard deviation and the skewness and kurtosis values. First, the percentage of students who attained a 2004 4th-grade FCAT of Level 2 or higher and Level 1 were reported. The percentages of promoted students attaining designations at each specific achievement level were also reported. A bar graph was utilized to display percentages of promoted, low achieving students who scored at Level 2 or higher and at Level 1 on the 2004 4th-grade FCAT reading section.

6. There is a difference in the obtained 2004 FCAT reading levels of promoted, low achieving students by a) gender, and by b) race/ethnicity.

Descriptive statistics for each subgroup (e.g., promoted, low achieving male, female, African-American, Asian, Caucasian, and Hispanic students) were reported including 2003 FCAT reading scaled score means, standard deviations and skewness and kurtosis coefficients. The percentages of the 2003 FCAT level designations were reported for promoted, low achieving male and female students along with a pie graph representing these percentages. Chi-square procedures were used to determine if the obtained proportion of promoted, low achieving male and female students who scored at Levels 1, 2, and 3 through 5 on the 2004 4th-grade FCAT reading test were significantly different than the expected proportions. Similarly, the percentages of the 2004 FCAT level

designations were reported for promoted, low achieving African-American, Asian, Caucasian, and Hispanic students. Chi-square procedures were also used to determine if the obtained proportions of promoted, low achieving African-American, Asian, Caucasian, and Asian students who scored at Levels 1, 2, and 3 through 5 or higher on the 2004 4th-grade FCAT reading test was significantly different than the expected proportions.

7. There is a difference in the obtained post-retention FCAT reading levels of promoted, low achieving students as a function of district size

This hypothesis was tested utilizing the database provided by the FLDOE. Specifically, 2003 3rd-grade FCAT reading scaled scores, the 4th-grade 2004 FCAT Levels, retention group classification, and the size of students' attended district were examined. Descriptive statistics including the 2003 FCAT reading score means and standard deviations from each district size designation were calculated for the promoted, low achieving students. The skewness and kurtosis of FCAT scores of promoted students by size of attended district were also calculated. A chi-square measure of association analysis was conducted for the promoted, low achieving students to determine if the obtained proportions of students who achieved state standards on the 2004 4th-grade FCAT reading test significantly differed from the expected proportions relative to the size of district attended.

8. For students promoted in third grade, there is a relationship among gender, race/ethnicity, prior performance on the FCAT reading test and attaining state reading standards on the 2004 FCAT reading test.

To test this hypothesis, the data from promoted, low achieving student sample were subjected to a logistic regression procedure. The logistic regression statistical procedure allows for a dichotomous outcome variable (achievement of state reading standards vs. non-achievement of state standards), and both dichotomous (e.g., male and female) and continuous (2003 FCAT scaled scores) variables as predictor variables. In the present study, gender, race/ethnicity, and prior performance on the 2003 FCAT reading test were entered into the model to predict meeting state standards for reading achievement in 2004 as measured by the 2004 4th-grade FCAT reading level designations for promoted, low achieving fourth-grade students. The nominal variables were coded to allow for within group comparisons among the higher achieving retained male students and likewise for the promoted, low performing students. The coding was as follows: (a) Males were coded as 0 and female students were coded as 1, (b) the race/ethnicity variable was dummy coded so that African-American, Asian, and Hispanic students were compared to Caucasian students, (c) student who received the FCAT reading Levels 2 through 5, which was considered to have met state reading standards, was coded as a 1 and the FCAT Level 1 designation which does not meet state standards and was coded as a 0, and (d) students who received free and reduced lunch in the 2003-2004 school year were coded as a 0, and students who did not receive free and reduced lunch during the 2003-2004 were coded as a 1. The overall likelihood ratio for the model was reported to determine if the model predicted retained student outcomes better than chance along with the Hosmer and Lemeshow goodness-of-fit statistic which indicates if model is a good fit for the data. The weights and standard error for each of the variables were reported and odds ratios for race/ethnicity, gender, and size of district attended for higher performing

retained third-grade students were also calculated and reported to determine if retention differentially benefited groups of students.

Chapter Four

Results

Introduction

The purpose of the present study was to explore the relationship between Florida's model of third-grade student progression and reading performance as measured by the FCAT-reading test. This chapter reports the results of the present study as delineated in the previous chapter. Specifically, this chapter includes a brief description of how the data were screened, student demographic information, the characteristics of the samples, and the results of the data analysis to answer the research questions and hypotheses posed.

Data Screening

Before the data were analyzed, students who attained a scaled score between 248 and 269 on the 3rd-grade 2003 FCAT-Reading test were identified. Students who attended home school and laboratory schools, and students who were younger than six-years-old and older than 16-years-old were excluded from the final sample. Based on these parameters, 14,139 students were identified. For the analyses that required gender, ethnicity, and district information, an additional 1,191 (457 retained, 734 promoted) students from the study were excluded due to inconsistencies within the database (e.g., obtaining a scaled score within the Level 1 designation yet receiving a Level 2), student attrition, and for missing data on the gender and ethnicity variables for a total of 12,948

students. For the analyses that required information regarding SES data, an additional 90 (35 retained, 55 promoted) students were excluded due to missing data. Thus, the overall sample for the study consisted of 12,858 students.

Characteristics of Final Samples

The final subsamples for this study included the following:

- a) The first subsample consisted of 3,886 students who were retained in third grade at the end of the 2003 academic year, attained a scaled score of 248-258 on the 2003 FCAT reading test, and who retook the 3rd-grade FCAT reading test in the 2003-2004 academic year.
- b) The second subsample was comprised of 9,062 students who were promoted to fourth grade at the beginning of the 2003-2004 academic year, attained a scaled score of 259-269 on the 2003 FCAT reading test, and who took the 4th-grade FCAT reading test in the 2003-2004 academic year.

Demographic characteristics of the students in the first subsample (hereafter, referred to as the higher performing retained group sample), and second subsample (hereafter, referred to as the promoted, low achieving sample) for academic year 2003-2004 are reported in Table 1. As is shown in the table, 40% of the higher performing retained students were African-American while 37% of promoted students were Caucasian. A great majority of the promoted, low achieving and higher performing retained students received free and reduce lunch (70% and 71%, respectively). As expected, the students attending Very Large districts comprised a majority of the higher performing retained (58%) and promoted, low achieving (59%) samples.

Table 1

Demographic Characteristics of Retained and Promoted Students

Characteristic	2004 Status					
	Total (N = 12,949)		Retained (N = 3886)		Promoted (N = 9062)	
	N	%	n	%	n	%
Gender						
Male	6720	51.90	2027	52.16	4,693	51.78
Female	6228	48.10	1859	47.83	4,369	48.21
Race/Ethnicity						
African-American	4715	36.41	1558	40.09	3157	34.84
Asian	163	1.26	37	.95	126	1.39
Caucasian	4532	35.00	1175	30.24	3357	37.04
Hispanic	3538	27.32	1116	28.72	2422	26.73
Free and Reduced Lunch						
Yes	9477	73.71	3087	80.16	6390	70.94
No	3381	26.29	764	19.84	2617	29.06
District Size						
Very Large	7619	58.84	2273	58.49	5346	58.99
Large	2090	16.14	674	17.34	1416	15.63
Medium	2051	15.84	595	15.31	1456	16.07
Small/Medium	717	5.54	205	5.25	512	5.65
Small	471	3.64	139	3.58	332	3.66

Performance of Samples on the 2003 3rd-Grade FCAT Reading Test

Descriptive statistics for the 2003 FCAT reading scaled scores by gender and race/ethnicity for promoted, low achieving and higher performing retained students are presented in Table 2.

Table 2

Means and Standard Deviations of 2003 3rd grade FCAT-Reading Scaled Scores for Retained and Promoted Students

Characteristic	N	Mean	SD	Skew	Kurtosis
Retained Students (N = 3886)					
Gender					
Male	2027	252.98	3.18	-.004	-1.23
Female	1859	252.99	3.15	-.012	-1.18
Race/Ethnicity					
African-American	1558	252.91	3.16	.013	-1.22
Asian	37	252.97	3.28	.215	-1.35
Caucasian	1175	253.15	3.12	-.074	-1.17
Hispanic	1116	252.93	3.21	.024	-1.21
Promoted Students (N = 9062)					
Gender					
Male	4693	264.17	3.16	-.078	-1.22
Female	4369	264.24	3.12	-.075	-1.20
Race/Ethnicity					
African-American	3157	264.05	3.18	-.017	-1.24
Asian	126	264.51	3.16	-.149	-1.26
Caucasian	3357	264.33	3.10	-.126	-1.17
Hispanic	2422	264.19	3.16	-.080	-1.22

The ranges, means, skewness and kurtosis values of scores on the 3rd-grade 2003 FCAT reading test were computed for both the higher performing retained and promoted, low achieving samples. The mean score for the higher performing retained group was 252.96 ($SD= 3.16$) with a range of 248 to 258. The skewness value for the higher performing retained group FCAT suggested a relatively normal distribution, however, it was platykurtic ($k= -1.21$), which was expected because the 2003 FCAT reading scaled scores

were restricted. As is shown in Table 2, the means and standard deviations of scores for males and females and across race/ethnicity were approximately equal.

The obtained mean for the 2003 FCAT reading scaled score variable for the promoted group was 264.19 ($SD= 3.14$) with a range of 259 to 269. An examination of the distribution of scores indicated that the skewness value was minimal (-0.07) but was platykurtic ($k= -1.22$). The promoted, low achieving gender and race/ethnicity subgroups had similar means, and also had platykurtic distributions.

Performance of Higher Achieving Retained Students

1. What proportion of higher performing retained third-grade students who had reading scaled scores of 248-258 on the 2003 FCAT-Reading test scored at Level Two or higher on the 2004 3rd-grade FCAT-Reading test?

The performance levels of the third-grade higher performing retained students are reported in Table 3.

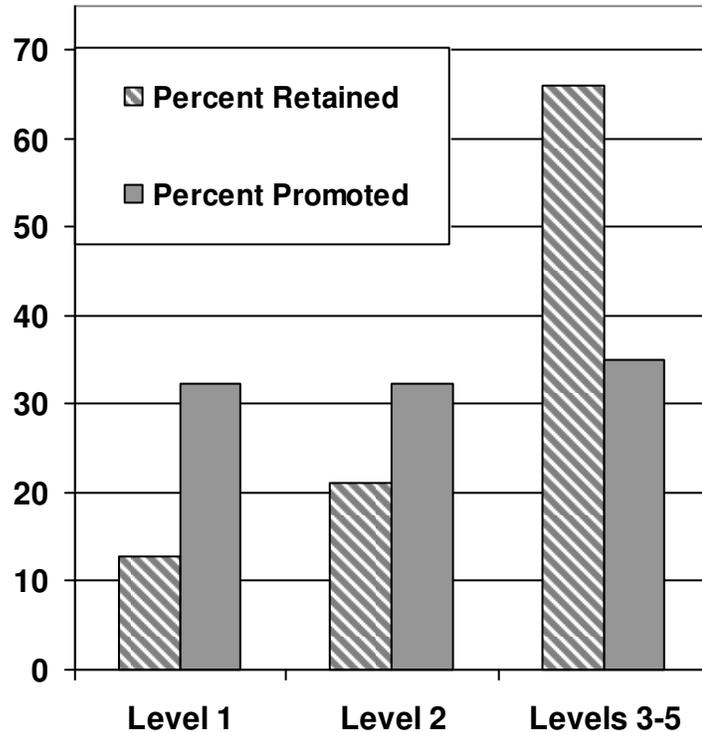
Table 3

Number and Percent of Retained Students on the 2004 3rd grade FCAT Reading Test by Performance Level

Performance Level	N	%
1	495	12.74
2	823	21.18
3	2015	51.85
4	527	13.56
5	26	.67
Total	3886	100

As is shown, a majority (87.26%) of higher performing retained third-grade students scored at Level 2 or higher on the 2004 3rd-grade FCAT reading test, while 12.74% (n = 495) scored at Level 1. A more detailed examination of students scoring at Level 2 or higher reveals that 66% of the higher performing retained students scored at Levels 3 through 5. Students scoring at Level 3 or higher on the FCAT reading tests are considered proficient in reading. Students who score at Level 1 are predicted to experience little success with future reading instruction. Although an attained Level 2 on the FCAT test meets state standards for reading, students scoring at this level are predicted to experience limited success with future reading instruction (Florida Department of Education, 2004). Figure 1 visually displays the percentage of higher performing retained and promoted, low achieving students scoring at Levels 1, 2, and 3 through 5 on the 2004 3rd-grade FCAT reading test.

Figure 1. Percentages of retained students attaining Levels 1, 2, and 3 through 5 on the 2004 3rd-grade FCAT reading test and promoted, low achieving students attaining Levels 1, 2, and 3 through 5 on the 2004 4th-grade FCAT reading test.



- There is a significant difference in the obtained 2004 FCAT reading levels of higher performing retained students by a) gender, and by b) race/ethnicity.

Gender. This hypothesis was rejected. There were 2027 males and 1859 females in the higher performing retained group. To test the hypothesis, a χ^2 test of Association was conducted. The number and percent of higher performing retained male and female students by performance level on the 3rd-grade FCAT reading test are presented in Table 4. The obtained χ^2 statistic was not statistically significant ($\chi^2 = 5.04, p > .05$), indicating that there was no difference in reading performance between

male and female higher performing retained students as measured by the 3rd-grade 2004 FCAT Reading test.

Table 4

Number and Percent of Retained Students by Performance Level and Gender on the 3rd-Grade FCAT-Reading in 2004.

	<u>N</u>	<u>Performance Level</u>					
		<u>Level 1</u>		<u>Level 2</u>		<u>Levels 3-5</u>	
<u>Gender</u>		<u>n</u>	<u>%</u>	<u>n</u>	<u>%</u>	<u>n</u>	<u>%</u>
Male	2027	266	54	401	49	1360	53
Female	1859	229	46	422	51	1208	47
Total	3886	495		823		2568	

$\chi^2(2, N = 3886) = 5.04, p = .08$

Race/Ethnicity. This hypothesis was supported. There were a total of 1,558 African-American, 37 Asian, 1,175 Caucasian, and 1,116 Hispanic students in the higher performing retained group. To address the hypothesis, the data were subjected to a χ^2 test of Association. The number and percent of African-American, Asian, Caucasian, and Hispanic students by performance level on the 2004 3rd-grade FCAT reading test are presented in Table 5. As is shown, there is a significant difference in the attained reading levels by race/ethnicity $\chi^2(4, N = 3886) = 70.21, p < .01$, indicating that the expected outcomes of retention varied as a function of race/ethnicity.

Table 5

Number and Percent of Retained Students by Performance Level and Race/Ethnicity on the 3rd-Grade FCAT-Reading Test in 2004

Race/Ethnicity	N	Performance Level					
		Level 1		Level 2		Levels 3-5	
		n	%	n	%	n	%
Caucasian	1175	111	22	226	27	838	33
African-American	1558	266	54	371	45	921	36
Hispanic	1116	115	23	214	26	787	31
Asian	37	3	1	12	1	22	1
Total	3886	495		823		2568	

$\chi^2(4, N = 3886) = 70.21, p < .01$

Specifically, more African-American students scored at Levels 1 and 2 than expected while fewer African-American students scored at Levels 3 through 5 than expected. In addition, more Caucasian students scored at Levels 3 through 5 than expected.

3. There is a significant difference in the obtained post-retention 3rd-grade FCAT reading levels of higher performing retained students as a function of district size.

This hypothesis was rejected. The number and percent of higher performing retained students by performance level on the 3rd-grade FCAT-Reading in 2004 by size of district are reported in Table 6.

Table 6

Number and Percent of Retained Students by Performance Level and Size of District on the 3rd-Grade FCAT-Reading in 2004

District Size	N	Performance Level					
		Level 1		Level 2		Levels 3-5	
		n	%	n	%	n	%
Small	139	16	3	29	4	94	4
Medium-Small	205	30	6	43	5	132	5
Medium	595	62	13	128	16	405	16
Large	674	72	15	132	16	470	18
Very Large	2273	315	64	491	60	1467	57
Total	3886	495		823		2568	

$\chi^2(8, N = 3886) = 11.44, p = .18$

A review of the Table 6 reveals that 2,273 higher performing retained students attended Very Large districts while only 139 students attended Small districts. The obtained χ was not statistically significant $\chi^2(8, N = 3886) = 11.44, p = .18$, indicating that the expected outcomes of retention did not vary significantly as a function of the size of attended districts.

4. For higher performing students retained in the third grade, there is a significant relationship among gender, race/ethnicity, prior performance on the FCAT reading test and attaining state reading standards on the 2004 FCAT reading test.

This hypothesis was accepted. A logistic regression was conducted. An additional 15 students were excluded due to missing data on the SES variable. Thus, a total of 3,851 students were included in the analysis. The variables race/ethnicity, 2003 FCAT reading

scaled scores, gender, and SES were entered into the regression model as independent variables. Relating to the gender variable, males were coded as 0, and female students were coded as 1. The race/ethnicity variable was subjected to a dummy coding procedure, comparing African-American, Asian, and Hispanic students with their Caucasian peers. The dependent variable was defined as student attainment of state standards on the 2004 3rd-grade FCAT reading test and was treated as a dichotomous variable (achieving state standards = 1, not achieving state standards = 0). As mentioned in previous chapters, students must score a level 2 or higher on the 3rd-grade FCAT-reading test to meet state standards for reading proficiency. The results of the regression are presented in Table 7. The results of the logistic regression showed that

$$\begin{aligned} \text{Predicted logit (Achieving State Standards)} = & -4.7679 + .0948*\text{Female} + \\ & (-0.5944*\text{African American}) + (-0.0196*\text{Hispanic}) + (0.1817*\text{Asian}) + \\ & (-0.2715*\text{LowSES}) + (0.0283*\text{Prior FCAT Reading Scaled Scores}). \end{aligned}$$

The overall likelihood ratio was statistically significant ($\chi^2 = 51.33, p < .01$) indicating that the model with five factors was significantly more effective in predicting students' achievement of state standards than a constant only model. The Wald and Score tests support this conclusion. The Hosmer and Lemeshow goodness-of-fit test was insignificant ($\chi^2 = 4.69, p = .78$) indicating that the obtained regression was a good fit for the data.

Table 7

Logistic Regression Analysis of Retained Students Meeting State Standards on the 2004 3rd-grade FCAT-Reading

Predictor	<i>B</i>	SEB	Wald's χ^2	<i>df</i>	<i>p</i>	e^B (odds ratio)
Constant	-4.768	3.924	1.4763	1	.2243	
Gender (1 = Female)	.0948	.0982	.9330	1	.3341	1.099
Race						
African American	-.5944	.1283	21.4574	1	<.0001	.552
Hispanic	-.0196	.1462	.0180	1	.8934	.981
Asian	.1817	.6115	.0883	1	.7664	1.199
SES (1 = Low SES)	-.2715	.1454	3.4885	1	.0618	.762
SSR	-0.028	.0155	3.3351	1	.0678	1.029
Test			χ^2	<i>df</i>	<i>p</i>	
Overall model evaluation						
Likelihood Ratio test			51.3274	6	<.0001	
Score test			51.5561	6	<.0001	
Wald test			50.2783	6	<.0001	
Goodness-of-fit test						
Hosmer & Lemeshow			4.6921	8	.7899	

The Goodman-Kruskal's Gamma statistic, which accounts for ties on both the outcomes and predictor variables (as are present in these data), is .206. This is interpreted as 21% fewer errors were made in predicting which of two students would achieve success on the FCAT-Reading by using the estimated probabilities than by chance alone. In addition, the c statistic, which for this model is .601, means that for 60% of all possible pairs of students – one successful and the other unsuccessful – the model correctly

assigned a higher probability of success to the student who was successful. This indicates that the model is better at assigning outcomes than one that randomly assigns probabilities to observations. In addition to the measures of association, a measure of classification was conducted, and results are displayed in Table 8.

Table 8

Observed and Predicted Frequencies for Attainment of State Standards in Reading by Logistic Regression with the Cutoff of .87

Observed	Predicted		% Correct
	Successful	Unsuccessful	
Successful	2119	233	63.0
Unsuccessful	1246	253	52.1
Overall % Correct			61.6

Note. Sensitivity = $2119/(2119+1246)\% = 62.9\%$. Specificity = $253/(253+233)\% = 44.6\%$. False positive = $233/(233+2119)\% = 9.9\%$. False negative = $1246/(1246+253)\% = 83.7\%$.

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. The false positive rate (9.9%) measures the proportion of observations misclassified as events over all those classified as events while the false negative rate (83.7%) measures the proportion of observations misclassified as nonevents over all those classified as non-events. The overall correction prediction was 61.6%, which is improved from chance.

Of the independent variables, only race was significantly associated with the performance level obtained on the 2004 FCAT Reading-Test. Specifically, African-American students were less likely to meet state standards ($B = -.594, p < .01$) than their

Caucasian peers. The Odds Ratio for African-American students was .552 (Confidence Limit = .43-.71) indicating that African-American students achieved state standards at approximately half the rate of their Caucasian counterparts.

Performance of Low Achieving Promoted Students

5. What proportion of promoted, low achieving fourth-grade students who had reading scaled scores of 259-269 on the 2003 FCAT reading test subsequently scored at Level 2 or higher on the 2004 fourth-grade FCAT reading test?

Overall, the percentage of promoted fourth-grade students who scored at Level 2 or higher on the fourth-grade 2004 FCAT reading test was 67.68% while 32.32% of students scored at Level 1 on this test. A more detailed analysis of students scoring at Level 2 or higher indicates that approximately 36% of students scored at Levels 3 through 5 on the fourth-grade test. The number and percentage for promoted students at each achievement level is presented in Table 9.

Table 9

Number and Percent of Promoted Students on the 2004 4th-grade FCAT-Reading by Performance Level

Performance Level	N	%
1	2929	32.32
2	2930	32.33
3	2906	32.07
4	291	3.21
5	6	.07
Total	9062	100

Figure 1 visually displays the percentage of higher performing retained and promoted, low achieving students scoring at Levels 1, 2, and 3 through 5 on the 2004 3rd-grade FCAT reading test. Students who scored at Level 3 or higher are considered proficient in reading (Florida Department of Education, 2004).

6. There is a difference in the obtained 2004 FCAT reading levels of promoted, low achieving students by a) gender, and by b) race/ethnicity.

Gender. This hypothesis was accepted. A total of 4,693 male and 4,369 female students were included in the promoted, low achieving sample. To test the hypothesis, the data were subjected to a Chi-square Test of Association. Table 10 reports the number and percent of promoted students by performance level on the 3rd-grade FCAT-Reading in 2004 by gender.

Table 10

Number and Percent of Promoted Students by Performance Level and Gender on the 4th-Grade FCAT-Reading in 2004

	<u>N</u>	<u>Performance Level</u>					
		<u>Level 1</u>		<u>Level 2</u>		<u>Levels 3-5</u>	
		n	%	n	%	n	%
Gender							
Male	4693	1573	54	1464	50	1656	52
Female	4369	1356	46	1466	50	1547	48
Total	9062	2929		2930		3203	

$$\chi^2 (2, N = 9062) = 8.21, p = .02$$

The obtained Chi-square statistic for promoted male and female students was statistically significant $\chi^2 (2, N = 9062) = 8.21, p = .02$. This indicated that promoted, low achieving

females and males 2004 reading achievement levels were significantly different than expected. Specifically, fewer promoted female students than expected scored at Level 1 on the 2004 4th-grade FCAT-reading test. Conversely, more male students attained Level 1 than expected.

Race/Ethnicity. This hypothesis was accepted. A total of 3,157 African-American, 126 Asian, 3,357 Caucasian, and 2,422 Hispanic students were included in the sample. In order to test the hypothesis, the data were subjected to a Chi-Square Test of Association. The number and percent of promoted students by race/ethnicity and performance level on the 4th-grade 2004 FCAT-Reading Test in 2004 are reported in Table 11.

Table 11

Number and Percent of Promoted Students by Performance Level and Race/Ethnicity on the 4th-Grade FCAT-Reading in 2004

Race/Ethnicity	<u>N</u>	Performance Level					
		<u>Level 1</u>		<u>Level 2</u>		<u>Levels 3-5</u>	
		n	%	n	%	n	%
Caucasian	3357	1000	34	1030	35	1327	41
African-American	3157	1209	41	1047	36	901	28
Hispanic	2422	690	24	810	28	922	29
Asian	126	30	1	43	1	53	2
Total	9062	2929		2930		3203	

$\chi^2 (7, N = 9062) = 124.59, p < .0001$

The obtained χ^2 was statistically significant $\chi^2 (7, N = 9062) = 124.59, p < .0001$

indicating that the expected outcomes of student promotion varied among race/ethnicity.

Specifically, more African-American students scored at Level 1 on the 2004 4th-grade FCAT reading test than expected, and fewer scored at Levels 3 through 5 than expected. Fewer Caucasian and Hispanic students scored at Level 1 and more Caucasian than expected scored at Levels 3 through 5.

7. There is a difference in the obtained post-retention FCAT reading levels of promoted, low achieving students as a function of district size.

This hypothesis was accepted. The number and percent of promoted students by performance level on the 4th-grade 2004 FCAT-Reading test by district size are presented in Table 12.

Table 12

Number and Percent of Promoted Students by Performance Level and Size of District on the 4th-Grade FCAT-Reading in 2004

District Size	N	Performance Level					
		Level 1		Level 2		Levels 3-5	
		n	%	n	%	n	%
Small	332	136	5	96	3	100	3
Medium-Small	512	171	6	164	6	177	6
Medium	1456	509	17	457	16	490	15
Large	1416	423	14	462	16	531	17
Very Large	5346	1690	58	1751	60	1905	59
Total	9062	2929		2930		3203	

$$\chi^2(8, N = 9062) = 22.21, p = .005$$

As shown in Table 12, the obtained χ^2 for promoted students by district attended was statistically significant $\chi^2(8, N = 9062) = 22.21, p = .005$, indicating that the expected outcomes of promotion varied among size of the students' district attended.

8. For students promoted to fourth grade in 2003-2004, there is a relationship among gender, race/ethnicity, prior performance on the FCAT reading test and attaining state reading standards on the 2004 4th-grade FCAT reading test.

This hypothesis was accepted. A logistic regression was conducted for the promoted group who took the fourth-grade FCAT in 2004. A total of 9,007 promoted students were included in the analysis, 58 students that were excluded due to missing data on the SES variable. Data reflecting race, 2003 FCAT reading scaled scores, gender, and SES were entered into the regression model as independent variables. For gender, males were coded as 1, and female students were coded as 0. The race/ethnicity variable was dummy coded, comparing African-American, Asian, and Hispanic students with their Caucasian peers. The dependent variable was student attainment of state standards on the 2004 fourth-grade FCAT reading test and was treated as a dichotomous variable (attainment of state standards = 1, non attainment of state standards = 0). The results of the logistic regression are presented in Table 13.

The results of the logistic regression showed that

$$\begin{aligned} \text{Predicted logit (Attaining State Standards)} = & -12.7390 + 0.1355 * \text{Female} + \\ & (-0.2663 * \text{African American}) + (0.1651 * \text{Hispanic}) + (0.3734 * \text{Asian}) + \\ & (-0.2730 * \text{LowSES}) + (.0517 * \text{Prior FCAT Reading Scaled Scores}). \end{aligned}$$

Table 13

Logistic Regression Analysis of Promoted Students Meeting State Standards on the 2004 4th-grade FCAT-Reading

Predictor	<i>B</i>	SEB	Wald's χ^2	<i>df</i>	<i>p</i>	e^B (odds ratio)
Constant	-12.7390	1.9118	44.3989	1	<.0001	
Gender (1 = Female)	.1355	.0456	.8.8296	1	.003	1.145
Race						
African-American	-.2663	.0569	21.9216	1	<.0001	.766
Hispanic	.1651	.0618	7.1226	1	.0076	1.179
Asian	.3734	.2162	2.9814	1	.0842	1.453
SES (1 = Low SES)	-.2730	.0557	24.0051	1	<.0001	.761
SSR	-.0517	.00724	51.0777	1	<.0001	1.053
Test			χ^2	<i>df</i>	<i>p</i>	
Overall model evaluation						
Likelihood Ratio test			167.0246	6	<.0001	
Score test			166.0687	6	<.0001	
Wald test			163.3335	6	<.0001	
Goodness-of-fit test						
Hosmer & Lemeshow			7.6385	8	.4696	

The overall likelihood ratio test was statistically significant $\chi^2(6, 9007) = 167.02$, $p < .01$). The Hosmer and Lemeshow goodness-of-fit Test was insignificant ($\chi^2 = 7.64$, $p = .47$) indicating that the obtained regression was a good fit for the data. The Goodman-Kruskal's Gamma statistic, which accounts for ties on both the outcomes and predictor variables (as are present in these data), is .168. This is interpreted as 17% fewer errors made in predicting which of two students would achieve success on the FCAT-Reading

by using the estimated probabilities than by chance alone. The c statistic, which for this model is .583, means that for 58% of all possible pairs of students – one successful and the other unsuccessful – the model correctly assigned a higher probability of success to the student who was successful. This indicates that the model is better at assigning outcomes than one that randomly assigns probabilities to observations. In addition to the measures of association, a measure of classification was conducted, and results are displayed in Table 14.

Table 14

Observed and Predicted Frequencies for Success by Logistic Regression with the Cutoff of .67

Observed	Predicted		% Correct
	Successful	Unsuccessful	
Successful	3616	2474	59.4
Unsuccessful	1377	1540	52.8
Overall % Correct			57.2

Note. Sensitivity = $3616/(3616+2474)\% = 59.4\%$. Specificity = $1540/(1540+1377)\% = 52.8\%$. False positive = $2474/(2474+3616)\% = 27.6\%$. False negative = $1377/(1377+1540)\% = 61.6\%$.

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. The false positive rate (27.6%) measures the proportion of observations misclassified as events while the false negative rate (61.6%) measures the proportion of observations misclassified as nonevents. The overall correction prediction was 57.2%, which is improved from chance.

According to the results of the logistic regression gender, race, prior achievement and SES were significantly associated with the level attained on the 2004 FCAT reading test for this select group of promoted students. When other variables were controlled, female students ($B = .135, p < .01$) were significantly more likely to achieve state standards than male students. The obtained odds-ratio for female students achieving state standards to male students was 1.15 (Confidence Limits 1.04-1.52). African-American students were less likely ($B = -.266, p < .01$) to achieve state standards than their Caucasian peers. The obtained odds-ratio for African-American students to Caucasian students achieving state standards was .766 (Confidence Limits .685-849). Hispanic students ($B = .165, p < .01$) were more likely to achieve state reading standards than their Caucasian peers. The obtained odds-ratio for Hispanic students to Caucasian students achieving state standards was 1.18 (Confidence 1.05- 1.33). Students of low socio-economic status were less likely ($B = -.273, p < .01$) to achieve reading state standards than students of high socio-economic status. The obtained odds-ratio for low SES students to high SES students for achieving state standards was .761 (Confidence Limits.682-.849). Finally, students who had higher prior achievement ($B = .052, p < .01$) were more likely to achieve state standards for reading.

CHAPTER FIVE

Discussion

Introduction

The purpose of the present study was to examine the academic outcomes of Florida's student progression policy on retained and promoted students. Specifically, the academic outcomes for two select samples of students were examined: students whose 2003 3rd-grade FCAT reading scaled scores fell just "under" (248-258) the criteria for promotion to fourth-grade (and were subsequently retained), and students whose 2003 3rd-grade FCAT reading scaled scores fell just "above" (259-269) the criteria for fourth-grade promotion (and were promoted). Chapter V contains a discussion of the results that were reported in Chapter IV. Specifically, the purpose of this chapter is to discuss the relevance of the present study in the context of past research, present a summary of research findings, and discuss the implications for educational policy in the State of Florida. The limitations of the study will be discussed in detail and directions for future research will be addressed.

Student Characteristics

Overall, the sample characteristics of the present study suggested that male and minority students were more likely to be retained than female and Caucasian students, respectively. This is consistent with previous research conducted by Abidin, et al. (1971), Alexander et al. (1994), Fine and Davis, (2003) and Jimerson et al. (1997). Upon closer

examination, the percentage of males in the retained group in the present study was lower than what has been found in previous research. The distribution of males in the higher achieving retained and promoted, low achieving groups in the present study was not congruent with the findings of Jimerson et al. (1997). Jimerson found that males constituted 74% of the higher achieving retained group, and 56% of the promoted, low achieving group, while males comprised 52% of the retained group and 51% of the promoted group for the present study.

Minority students represented the majority of the higher achieving retained (70%) and promoted, low achieving (63%) groups in the present study, while only constituting 51% of the third-grade population in the State of Florida. This represents a 19% overrepresentation of minority students in the higher achieving retained sample. This was congruent with previous research (McCoy & Reynolds, 1999; Reynolds, 1992) that indicated that minority students were more likely to be retained than their Caucasian peers. Minority students were also overrepresented in the promoted, low achieving sample by 12%. African-American students were the most overrepresented ethnic group in the higher achieving retained and promoted, low achieving groups. African-American students consisted of 24% of all third-graders in the State of Florida in the 2002-2003 school year, while African-American students comprised 40.23% of all higher achieving, retained students in the present study.

The observed student characteristics of the higher achieving retained and promoted, low achieving groups may have been directly related to how the groups were identified for the study. Students were selected based upon a narrow range of 2003 3rd-grade FCAT reading test scaled scores. The percentages of minority and male students

may be different for the entire population of students who scored at Levels 1 and 2 on this test. In addition, the retention rates for minority students may have been influenced by socio-economic status (SES). The students who were selected for the study scored at Levels 1 and 2. Previous research would suggest that SES may account for more of the variance in reading performance than race/ethnicity. Furthermore, the educational opportunities for minority students may have put those students at risk for retention.

Retained Student Outcomes

Overall, higher achieving third-grade retained students were likely to meet state standards for third-grade reading proficiency at the conclusion of the retention year. This conclusion was supported by the result that 87.6% of higher achieving retained students attained a Level 2 or higher designation on the 3rd-grade FCAT reading test. Perhaps of greater interest is that approximately two-thirds of the higher achieving retained students scored at Level Three or higher on the 2004 FCAT reading test. Many of the higher achieving retained students not only met the minimum state standards for reading, but were proficient in reading at the third-grade level as defined by the State of Florida. Many of the retained third-grade students in this study would not be considered at-risk for future academic failure at the conclusion of their retention year.

The general outcomes for the higher achieving retained students in this study are not consistent with the overall results of previous research. The meta-analysis conducted by Jimerson (2001) clearly indicated that the majority of retention literature has suggested that grade retention was not an effective intervention for addressing academic deficiencies for students. In fact, retention was often associated with negative academic outcomes. The few studies that reported initial positive results were not longitudinal and

indicated that the academic gains were not sustained over time. Furthermore, the few academic gains that were demonstrated by retained students were not related to reading proficiency. Rather, a limited number of students demonstrated higher achievement in mathematics after the retention year.

It is difficult to determine why the majority of higher achieving retained students in the present study demonstrated increased reading proficiency when previous research would predict otherwise. It may be due in part to the educational climate that exists in Florida. The FCAT reading test is linked to third-grade promotion in the State of Florida. Students must pass the test in order to be promoted. If third-grade students who were retained fail for a second time, they will be retained in third grade again, repeating the grade for a third time. This type of atmosphere in the State of Florida has influenced the priorities for teachers, principals, parents and students. The emphasis on reading achievement, and the high stakes associated with the FCAT, may have had an impact on the quality of reading instruction that was provided to students in Florida's third-grade classrooms.

The intent of the retention year is to adapt instruction in order to maximize success in the acquisition of reading skills during that retention year. Students who were retained were required to receive differentiated instruction intended to remediate reading deficiencies. According to the State of Florida's policy (Florida Department of Education, 2002a) retained students are required to receive prescriptive and intensive remedial reading instruction delineated on an Academic Improvement Plan (AIP). Students also received additional instructional time which allowed for students to increase their academic engaged time in the third-grade curriculum. Previous research has

linked academic engaged time (AET) to higher academic achievement (Aronson et al., 1998). The extra AET afforded by the retention year may help explain, in part, the academic achievement of the higher performing retained students in the study. The combination of additional academic instructional time and instruction organized through AIPs may have provided the higher achieving retained students with the opportunity to improve reading skills. Moreover, it is possible that the retained students were successful because they benefited from third-grade instruction and would have not been successful within the fourth-grade curriculum. Research has indicated that students benefit most from instruction within their instructional level (Shapiro & Elliot, 1999). Fourth-grade regular education teachers may have not been able to provide the higher achieving retained students with differentiated instruction at their instructional level that would be necessary for success.

Although the results of the present study indicated gains in academic achievement for the higher achieving retained students, past research (Pagani et al., 2001) has suggested that these gains may not continue after the retention year. Students who were retained and subsequently scored at a proficient level on the FCAT reading test may not continue to receive supplemental academic services in addition to the core curriculum. If and when the extra academic learning time and academic supports are withdrawn, the observed reading gains for the higher achieving retained group may fade over time. Additional research is needed to confirm or reject this hypothesis.

More perplexing questions occur when the retained students reach secondary education. A recent study conducted by Jimerson et al. (2002) indicated that while retention may have initial benefits on academic achievement, it was a significant

predictor of future school dropout. Additionally, the research conducted by Fine and Davis (2003) suggested that students who were retained once were half as likely to enroll in post-secondary education as their promoted peers. These previous studies did not delineate the academic remediation strategies that were employed during the retention year, and it not clear if students received supplementary academic instruction. Therefore, it remains to be seen if the student outcomes from previous research pertain to the students in the present study.

Promoted Student Outcomes

Previous research (Jimerson et al., 2001) has indicated that when compared to similarly performing retained students, promoted students attained higher scores on standardized reading tests. The results of the present study may suggest otherwise. Like their retained peers, promoted, low achieving students who were promoted to fourth grade were also likely to meet state reading standards on the 2004 4th-grade FCAT-Reading test. Approximately two-thirds of promoted students attained a Level Two designation or higher on the FCAT. However, only 35% of low-achieving promoted students scored proficiently (Levels 3 through 5) on the 4th-grade 2004 reading FCAT as compared with the 60% of students who scored proficiently on the 3rd-grade FCAT-reading test. Promoted students were relatively evenly distributed among Levels 1, 2 and 3, with very small percentages of students scoring at Levels Four and Five on the 2004 4th-grade FCAT-Reading test. Although a majority of the promoted, low achieving students passed the 4th-grade FCAT, only 35% of the students attained a level of proficiency in reading from the previous year. This indicates that approximately two-thirds of the promoted students may be considered at-risk for future academic failure as

opposed to one-third of the higher achieving retained group. More explicitly, a majority of the promoted, low achieving students maintained their position as relatively low-achievers at the end of fourth grade.

The previous research (Jimerson et al., 2001) that has suggested the negative academic effects of retention has typically compared the achievement outcomes of retained students and similar low achieving peers. It is difficult to directly compare the higher achieving retained and promoted, low achieving students in the present study because the two groups of students took different levels (reflecting different skill sets) of the FCAT. The FCAT proficiency rate for the low achieving, promoted students was somewhat less than the proficiency rate of higher achieving retained students but, the retained third-grade students took the 3rd-grade FCAT and the promoted fourth-grade students took the more difficult 4th-grade FCAT. Therefore, the reading proficiency differences that were observed in higher achieving retained and promoted, low achieving students may not represent the retained group closing the achievement gap. The higher achieving retained group may subsequently not score as proficiently on 4th-grade measures of reading achievement. However, the results of the present study seemed to suggest that within their respective curricula, retained students were more likely to be proficient in reading than promoted, low achieving students.

Student Outcomes by Gender

In this study, meeting state reading standards as measured by the FCAT reading test did not vary by gender for the higher achieving retained students. However, meeting state reading standards was moderated by gender for the promoted, low achieving students in the present study. The χ^2 was statistically significant, but it is difficult to

determine if this outcome was due to the large sample size of the group ($n = 9,062$), or if it was due to clinically significant differences in promoted student outcomes. Similarly, the χ^2 for the higher achieving retained group may have been statistically significant if it would have contained a larger sample size.

Student Outcomes by Race/Ethnicity

African-American students were the most overrepresented racial group in the higher achieving retained and promoted, low achieving groups. In addition, higher achieving retained African-American students attained a Level 1 designation on the 3rd-grade 2004 FCAT reading test at a significantly higher rate than was expected. The logistic regression analysis indicated that higher achieving retained African-American students were significantly less likely to achieve state standards in reading than their Caucasian peers, even after other variables such as gender, SES, and prior reading achievement were controlled. This trend was similar among the promoted African-American students. More promoted African-American students did not meet state reading standards on the 2004 4th-grade FCAT reading test than expected. Moreover, African-American students were significantly less likely to achieve state standards than their Caucasian peers. These results seem to suggest that African-American students tended to be less likely to achieve state standards in reading regardless of promotion status. Neither retention nor promotion seemed to improve the reading outcomes for African-American students when compared with their peers. The implications of this finding for retention policies and future research are discussed later in the chapter.

The reading outcomes for Hispanic students seemed to be more encouraging. According to the results of the logistic regression, higher achieving retained Hispanic

students were just as likely to achieve state standards on the 3rd-grade FCAT reading tests as their Caucasian peers. The results from the χ^2 statistical procedure indicated that more Hispanic students scored at Levels 3 through 5 (proficient in reading) than expected. Given these results, it seems as if higher achieving retained Hispanic students benefited from the extra year of instruction more so than other minority groups, and on the same level as their Caucasian peers. Interestingly, promoted, low achieving Hispanic students were *more* likely to achieve state standards in reading than their Caucasian peers when SES, prior achievement, and gender were held constant. In addition, more Hispanic students scored at a proficient level on the 4th-grade FCAT reading test than was expected. This seems to suggest that the majority of promoted Hispanic students outperformed their peers on the 2004 4th-grade FCAT reading test. Based upon these and the previous results, it seems that student characteristics did contribute to the outcomes of retention.

Limitations

The present study has a number of limitations, many of which are present in previous retention literature. The students in the present study were not randomly assigned to the retained and the low achieving, promoted groups. Therefore, this study was not truly experimental and it was impossible to isolate the effects of retention or student progression on students' academic achievement. Moreover, the retained students and the promoted, low achieving students took different versions of the FCAT. Therefore it is impossible to directly compare the 3rd-and 4th-grade 2004 FCAT levels for the two subsamples of students.

A second limitation was the use of the FCAT-Reading test as the sole criteria for reading achievement. The FCAT-Reading test has excellent technical characteristics (Florida Department of Education, 2002b), yet it may not accurately represent the curricula that were delivered to retained and promoted students in Florida's classrooms. The study did not account for other methods that may be used to demonstrate students' reading performance (e.g., other standardized achievement measures, student portfolio, curriculum-based measures). The State of Florida allows some students to demonstrate reading proficiency through student portfolios and alternative assessment.

A third limitation is related to the database that was used. The database contained information from a large sample of retained and promoted third-grade students. The coding system that was used may have caused errors in the database which may threaten the validity of the study. Inter-rater agreement on the codification of the data was not conducted by districts or the Florida Department of Education. In addition, students' SES data was determined through eligibility for participation in the free and reduced lunch program defined by the State of Florida. Students who were eligible for free and reduced lunch were coded as low SES. Students who were not eligible for free and reduced lunch were considered to not be of low SES. This definition of SES limited the range and the continuous nature of the variable and may not have accurately measured the true socio-economic status of the students. However, virtually all educational research studies conducted in the United States that include SES as a variable use this definition.

A fourth limitation of the study is that the independence assumption required for the logistic regression procedure was violated. The students in the present study were nested within various schools across the State of Florida. It is likely that different schools

had different reading curricula to prepare students for the FCAT. It is also possible that schools had differing educational policies, resources, and services that may have differentially affected students FCAT scores. In other words, the FCAT-Reading scores of students who attended the same elementary school may not be independent of each other. Therefore, the fidelity of the logistic regression procedure may have been compromised and should be interpreted with caution and with supplemental evidence (e.g., previous research, χ^2 analyses).

The final limitation is that the present study only examined the short term student outcomes within the context of Florida's student progression plan. It is impossible to determine if the academic gains that were demonstrated by the retained and promoted, low achieving groups will continue. The research that was previously mentioned (Fine & Davis, 2003; Jimerson et al., 2002; McCoy & Reynolds, 1999) has indicated that the effects of grade retention are long-term, and the short-term effects may distort the later outcomes for students.

Delimitations

The results of this study may be generalized to third-grade students in the State of Florida who scored within a narrow band (248-269) of 2003 3rd-grade FCAT scaled scores. The results of the study may not generalize to students who were retained or promoted in different grades, or third-grade students who were retained and promoted in third-grade but achieved an FCAT-reading scaled score of less than 248 and greater than 269. Additionally, the results may not generalize to states that do not have educational policies that mandate academic remediation of retained students, or require individualized academic support plans for retained students.

Implications for Florida's Student Progression Policy

This study evaluated Florida's student progression policy by examining the academic outcomes of retained and promoted third-grade students. The results suggested that Florida's policy was associated with increased rates of reading proficiency for a select group of higher performing retained and promoted, low achieving students. Students who scored at the higher end of the Level 1 designation cutoff (248-258) seemed to benefit (in the short term) from the mandated academic remediation that was provided during the retention year. The majority of higher performing retained students not only met the minimum state standards for reading achievement, but many of these students were also proficient readers by the end of the retention year. Over 60% of the retained students moved from a Level 1 designation in 2002-2003 to Level 3 designation or higher in 2003-2004. These results seem to suggest that the policy to retain these students was supported at least initially.

Unfortunately, Florida's retention policy did not benefit all students equally. Male students and minority students, especially African-Americans, were retained at disproportional rates. African-American males were the most likely group to be retained under the current retention policy. Reasons for this may include the criteria for which retention decisions are made. Florida uses the FCAT Reading test for retention decisions. Fewer African-American students may be retained if there was more flexibility within retention decisions. The use of alternative forms of assessment in conjunction with the FCAT Reading test such as curriculum-based measures, increased use of portfolios to document reading proficiency, attendance, and other norm-referenced reading achievement tests may reduce the number of minority students retained in third grade.

The State of Florida could also reduce the number of minority students retained in third grade by ensuring that these students have equitable access to evidence-based core and supplemental instruction *before* the retention year. By doing so, Florida may prevent the retention of students with moderate reading problems and the costs that are associated with it.

The majority of higher achieving retained and promoted, low achieving students achieved state standards on the 2004 FCAT reading Test. However, African-American students in the present study fared much worse on the 2004 3rd-grade FCAT-Reading test than their Caucasian peers. The State of Florida should ensure that all students have equal access to supplemental and intensive academic services during the retention year. A hypothesis for why retained African-American students did not perform well on the 2004 FCAT was that these students may not have received high-quality remedial instruction and interventions during the repeated year. Schools may not have had the resources to closely monitor the academic progress of low SES African-American students. Additionally, schools may not have had the resources to provide evidence-based instruction at the level of intensity that was required for these students. Funds, training, and staff should be equitably distributed according to student needs if the State of Florida expects high academic standards from all students.

Overall, Florida's retention policy also seemed to benefit the promoted, low achieving students. The majority of these students met state standards in reading, however, only a minority were considered proficient in reading at the end of fourth-grade. If this subsample would have included students who attained scaled scores higher than 269, it is likely that more students would have scored at Level 3 or higher on the

2004 4th-grade FCAT-Reading test. Interestingly, prior FCAT scores were a significant predictor of achieving state standards for the promoted, low achieving students. This was not expected given the restriction placed on the 2003 FCAT-Reading scores (259-269). Promoted, low achieving students who scored at the low end of this range were significantly less likely to meet state standards for reading. This may indicate that cutoff score for retention decisions was not high enough, and some of the promoted students may have benefited, at least initially, from services provided in association with grade retention.

African-American and male students were overrepresented in the promoted, low achieving group, continuing the trend that was observed in the retained group. They also were less likely to achieve state standards for reading. This adds credibility to the argument that the current service delivery of AIPs does not seem to be as effective with this population of students. Florida should ensure that all low achieving students have equitable access to evidence-based reading instruction and interventions.

In summary, many researchers agree that retention and social promotion are not sufficient for addressing the needs of students who do not demonstrate adequate yearly academic progress (Jimerson, et al. 2002). Educational policies should focus on preventing academic difficulties before they occur. Florida's student progression policy may have achieved this goal by pressuring school administrators, principals, teachers, and parents to focus on reading achievement before students reach third-grade. By providing early intervention services to students, fewer students will require supplemental academic instruction, and will be less at-risk for retention. Third-grade students who are at-risk for retention should receive differentiated instruction and additional academic learning time

to catch up to their peers. It is possible that students with reading difficulties may benefit from being retained in reading, but not in grade. In this scenario, students receive reading instruction in a third-grade classroom and still benefit from fourth-grade instruction in content areas other than reading.

Implications for Future Research

Although the results of the present study indicate that a majority of both the retained and promoted, low achieving groups achieved state standards the subsequent year, future research is needed to directly compare the reading achievement of both groups. This would provide for more definite conclusions of the impact of retention versus promotion on reading outcomes under the provisions of Florida's student progression policy. Additionally, a longitudinal design would be beneficial to determine if positive academic effects are sustained over time after the retention year. Previous research has consistently indicated the negative long term effects of retention; however, no study has examined the long term effects of retention within the context of a state wide retention initiative. Future research questions should determine if a large scale retention policy, such as Florida's, affects student drop out rates, and post-secondary education enrollment.

Due to the exploratory nature of this study, there are many variables that were not controlled. These included the integrity of AIPs, early intervention services, and quality of core instruction. Future research should attempt to determine the moderating effects of these variables on the outcomes of retention (Does the quality of an AIP during the retention year predict successful academic outcomes?). In addition, the impact of early intervention services and an evidence-based core instruction on the rates of student

retention should be examined. Some research questions may include: (1) Are retention rates lower for school districts with Reading First grants?, (2) Do evidence-based early intervention services for reading lower the number of students who are at-risk for grade retention?, and (3) Does a tiered model of service delivery impact number of students retained?

The present study focused exclusively on the academic outcomes of retention. However, retention may have other potentially important effects on students. Previous literature has suggested that retained students are more at-risk for mental health difficulties, poor attendance, and behavioral and social problems (National Association of School Psychologists, 2003a) Future research should examine the impact of grade retention on the social and emotional outcomes of students within the context of a statewide policy of retention.

Conclusion

The current educational climate in the United States has recognized the importance of reading for the educational outcomes of students. Florida has responded to this emphasis by enacting policies to help ensure that all students have pre-reading and reading skills by the end of the third grade. Students that do not meet state standards for reading at the end of third grade are retained and are provided with systematic interventions in the form of academic improvement plans (AIPs). In this study, higher achieving third-grade students who were retained were likely to succeed in the following year. These results suggest support for retention practices within the context of mandated academic remediation. However, more research is needed to determine the long-term academic and social impact of retention practices in Florida before more unequivocal

recommendations are made. Additionally, the retention policy did not seem to address, in an equitable fashion, the academic needs for certain groups of students (African-Americans, Males). These groups of students were identified as at-risk for repeated failure. More information about these groups of students should be gathered to identify, analyze, and develop solutions for this problem.

References

- Abidin, R.R., Golladay, W. M., & Howerton, A. L. (1971). Elementary school retention: An unjustifiable, discriminatory, and noxious policy. *Journal of School Psychology, 9*, 410-414.
- Alexander, K., Entwisle, D., & Dauber, S. (1994). *On the success of failure: A reassessment of the effects of retention in the primary grades*. New York: Cambridge University Press.
- Aronson, J. Zimmerman, J., & Carlos, L. (1998). Improving student achievement by extending school: Is it just a matter of time? Retrieved October, 11, 2003 from http://www.wested.org/online_pubs/timeandlearning/1_intro.html
- Canales, J., Frey, J., Walker, C., Walker, S., Weiss, S., & West, A. (2002), *No state left behind: The challenges and opportunities of ESEA 2001*. Denver, Colorado: Education Commission of the States.
- Clinton, W. (1999). *State of the Union Address (January 1, 1999)*. Washington, DC: U.S. Government Printing Office.
- Coyne, M., Kame'enui, E., & Simmons, D. (2001). Prevention and intervention in beginning reading: Two complex systems. *Learning Disabilities Research & Practice, 16*(2), 62-73.
- Dawson, P. (1998). A primer on student grade retention: What the research says. *Communique, 26*(8), 28-30.

Denton, D. (2001). Finding alternatives to failure: Can states end social promotion and reduce retention rates? Atlanta, Georgia. Southern Regional Education Board.

Donahue, P., Finnegan, R., Lutkus, A., Allen, N., Campbell, J. (2001). The nation's report card: Fourth grade reading 2000. *National assessment of educational progress*. National Center of Educational Statistics. United States Department of Education.

Donnelly, B. (2000). *Making sense of reading*. Austin, Texas. Southwest Educational Development Laboratory.

Education for All Handicapped Children Act of 1975, 20 U. S. C. Sec. 401 (1975).

Ferguson, P., Jimerson, S. R., & Dalton, M. (2001). Sorting out successful failures: Exploratory analyses of factors associated with academic and behavioral outcomes of retained students. *Psychology in the Schools*, 38(4), 327-342.

Fine, G. J., & Davis, J. M. (2003). Grade retention and enrollment in post-secondary education. *Journal of School Psychology*, 41, 401-411.

Florida Department of Education. (2002a). *Student Progression*. Retrieved

October 10, 2003 from

http://www.flsenate.gov/Statutes/index.cfm?App_mode=Display_Statute&search_String=&URL=Ch1008/SEC25.HTM&Title=-

- Florida Department of Education. (2002b). *Technical report: For operational test Administrations of the 2000 Florida Comprehensive Assessment Test*. Retrieved October 11, 2003 from <http://www.firn.edu/doe/sas/fcat/pdf/fc00tech.pdf>.
- Florida Department of Education. (2003). FCAT reading SSS. *The Florida Comprehensive Assessment Test*. Retrieved October 11, 2003 from <http://www.firn.edu/doe/sas/fcat.htm>.
- Florida Department of Education. (2004). Third grade reading 2003 state reading Profile. Retrieved January 7, 2004 from <http://www.firn.edu/doe/commhome/pdf/3statesum.pdf>
- Gottfredson, D. C., Fink, C., & Graham, N. (1994). Grade retention and problem behavior. *American Educational Research Journal*, 31(4), 761-784.
- Graue J., & DiPerna, E. (2000). Redshirting and early retention: Who gets the “gift of time” and what are its outcomes. *American Educational Research Journal*, 37(2), 509-534.
- Gredler, G.R. (1997). Intervention programs. *Psychology in the Schools*, 34, 161-169.
- Greenwood, C. R., Maheady, L., & Delquadri, J. (2002). Classwide peer tutoring programs. In M. Shinn, H. Walker, & G. Stoner, (Eds.), *Interventions for academic and behavioral problems II: Preventative and remedial strategies*. (pp. 611-649) Bethesda, MD: National Association of School Psychologists.

- Haager, D., & Windmueller, M., (2001). Early reading intervention for English Language learners at-risk for learning disabilities: Student and teacher Outcomes in an urban school. *Learning Disability Quarterly*, 24(4), 235-249.
- Jenkens, J., & O'conner R. (2002). Early identification and intervention for young children with reading/learning disabilities. In R. Bradley, & L. Danielson, (Eds.), *Identification of learning disabilities: Research to practice*. (pp. 99-149). Mahwah, NJ: Lawrence Erlbaum Associates.
- Jimerson, S. R. (1999). On the failure of failure: Examining the association of early grade retention and late adolescent education and employment outcomes. *Journal of School Psychology*, 37(3), 243-272.
- Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, 30 (3), 420-437.
- Jimerson S. R, Anderson G E., & Whipple A. D. (2002). Winning the battle and losing the war: examining the relation between grade retention and dropping out of high school. *Psychology in the Schools*, 39(4) 441-457.
- Jimerson, S., Carlson, E., Rotert, M., Egeland, B. & Sroufe, L. (1997). A prospective, longitudinal study of the correlates and consequences of early grade retention. *Journal of School Psychology*, 35(1), 3-25.

- Jimerson, S.R., Kaufman, A. M., Anderson, G. E., Whipple, A. D., Figueroa, L. R., Rocco, F. & O'Brien, K. M. (2002). Beyond grade retention and social promotion: Interventions to promote social and cognitive competence. Manuscript submitted for publication.
- Jimerson, S. R., & Kaufman, A. M. (2003). Reading, writing, and retention: A primer on grade retention research. *The Reading Teacher*, 56(8), 622-635.
- K-20 Education Code, 48 FL. Gen. § 1008 *et seq.* 2003.
- Lennon, J. E., Slesinski, C. (1999). Early intervention in reading: Results of a screening and intervention program for kindergarten students. *School Psychology Review*, 28(3) 353-64.
- Linn, R., Baker, E., & Betebenner, D. (2002). *Accountability systems: Implications of Requirements of the NCLB Act of 2001. CSE technical report.* California.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental scores.* Reading, MA: Addison-Wesley.
- Mann, D., & Shakeshaft, C. (2003). In God we trust; All others bring data. *School Business Affairs*, 69(1) 19-22.
- Mantzicopoulos, P. V. (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 the second grade. *American Educational Research Journal*, 29, 182-198.

- McCoy, A. R., & Reynolds, A. J. (1999). Grade retention and school performance: An extended investigation. *Journal of School Psychology, 37*, 273-298.
- National Association of School Psychologists, (2003a). Position statement on student grade retention and social promotion. Retrieved October 10, 2003 from http://www.nasponline.org/information/pospaper_graderetent.html.
- National Association of School Psychologists, (2003b). Position statement on early intervention services. Retrieved May 10, 2004 from http://www.nasponline.org/information/pospaper_eis.html
- National Reading Panel, (2000). Teaching children to read: An evidence-based assessment of the scientific research literature of reading and its implications for reading instruction. Washington DC: United States Department of Education.
- Nelson, S. (1990). Instructional time as a factor in increasing student achievement. Portland, OR: Northwest Regional Educational Laboratory.
- Niklason, L. (1984). Nonpromotion: A pseudoscientific solution. *Psychology in the Schools, 21*(4), 485-499.
- No Child Left Behind, (2002). Retrieved October 10, 2003 from <http://www.ed.gov/nclb/landing.jhtml?src=pb>.
- Pagani, L., Tremblay, R. E., 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*, 24(1), 107-18.
- Phillips L., Norris S., Osmond W., & Maynard, A. (2002). Relative reading achievement: A longitudinal study of 187 from first through sixth grades. *Journal of Educational Psychology*, 94(1), 3-13.
- Rafoth, M., (1991). A survey of state-level contacts for school psychology regarding retention/promotion practices: Are we evaluating the risks and benefits? *Psychology in the Schools*, 28(1), 35-42.
- Reynolds, A., (1992). Grade retention and school adjustment: An explanatory analysis. *Educational Evaluation and Policy Analysis*, 14(2), 101-121.
- Shapiro, E.S, & Elliot, N.S. (1999). Curriculum-based assessment and other performance-based assessment strategies. In C. Reynolds and T. Gutkin (Eds.), *Handbook of school psychology* (pp. 383-408). New York: John Wiley.
- Shapiro, E. S., Durnan, S. L., Post, E. E., & Levinson, T. S. (2002). Self-monitoring procedures for children and adolescents. In M. Shinn, H. Walker, & G. Stoner, (Eds.), *Interventions for academic and behavioral problems II: Preventative and remedial strategies*.(pp. 611-649) Bethesda, MD: National Association of School Psychologists.
- Shepard, L.S., & Smith, M. L. (1987). Effects of kindergarten retention at the end of first grade. *Psychology in the Schools*, 24, 346-357.

Sindelar, P. T., Lane, H. B., Pullen, P. C., & Hudson, R. F. (2002). Remedial Interventions for students with reading decoding problems. In M. Shinn, H. Walker, & G. Stoner, (Eds.), *Interventions for academic and behavioral problems II: Preventative and remedial strategies*.(pp. 611-649) Bethesda, MD: National Association of School Psychologists.

Statistical Package for Social Sciences. (2003). Retrieved December 1, 2003, from <http://www.spss.com>.

Tomchin, E. M., & Impara, J. C. (1992). Unraveling teachers' beliefs about grade retention. *American Educational Research Journal*, 29, 199-223.

United States of America Constitution. (1787).

United States Department of Education, (2002). *Why NCLB is Important to America*. Retrieved October 12, 2003 from <http://www.ed.gov/nclb/overview/importance/edlite-index.html>

United States Department of Education (2003). Introduction: No Child Left Behind. Retrieved October 9, 2003 from <http://www.ed.gov/nclb/overview/intro/index.html>