

PREDICTORS OF MANDATORY THIRD GRADE RETENTION FROM HIGH-STAKES TEST PERFORMANCE FOR LOW-INCOME, ETHNICALLY DIVERSE CHILDREN

by

Tanya Tavassolie  
A Thesis  
Submitted to the  
Graduate Faculty  
of  
George Mason University  
in Partial Fulfillment of  
The Requirements for the Degree  
of  
Master of Arts  
Psychology

Committee:

\_\_\_\_\_ Director

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_ Department Chairperson

\_\_\_\_\_ Dean, College of Humanities  
and Social Sciences

Date: \_\_\_\_\_ Summer Semester 2015  
George Mason University  
Fairfax, VA

Predictors of Mandatory Third Grade Retention From High-Stakes Test Performance for  
Low-Income, Ethnically Diverse Children

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of  
Arts at George Mason University

by

Tanya Tavassolie  
Bachelor of Arts  
Franklin and Marshall College, 2011

Director: Adam Winsler, Professor  
Department of Psychology

Summer Semester 2015  
George Mason University  
Fairfax, VA



This work is licensed under a [creative commons attribution-noncommercial 3.0 unported license](https://creativecommons.org/licenses/by-nc/3.0/).

## **DEDICATION**

This is dedicated to my loving fiancé Ken, and my wonderful mother, father, and two sisters.

## **ACKNOWLEDGEMENTS**

I would like to thank the many friends and family who have supported me throughout my graduate career thus far. My loving family and friends helped make this possible. Drs. Winsler, Curby, and Williams were of invaluable help, and without them this thesis would not have been possible. Finally, thanks go out to the Fenwick Library for providing a clean, quiet, and well-equipped repository in which to work.

## TABLE OF CONTENTS

	Page
List of Tables .....	vii
Abstract .....	viii
Predictors of mandatory 3 <sup>rd</sup> grade retention from high-stakes test performance for low-income ethnically diverse children .....	1
High Stakes Tests .....	2
Traditional Retention Policies .....	4
Chicago’s Traditional Policy .....	5
Baltimore’s Traditional Policy .....	7
Florida’s Kindergarten Policy .....	8
Florida .....	10
The Current Study .....	15
Method .....	19
Participants .....	19
Measures .....	20
Florida Comprehensive Assessment Test (FCAT) .....	20
Norm-Referenced Test (NRT) .....	21
Special Education status .....	21
Grade Point Average (GPA) .....	22
B average or Better .....	22
Retained in Grade 3 .....	22
English use at home (ELL status) .....	23
English proficiency in third grade .....	23
Free/reduced lunch status .....	24
Gender .....	24
Ethnicity .....	24
Analytic Plan .....	25

Results.....	27
Research Question 1.....	27
Retention policy exemption categories.....	27
Research Question 2.....	30
Failing FCAT Reading Test. ....	30
Retention in Third Grade.....	31
Performance in Third Grade.....	32
Research Question 3.....	33
Predictors of Failing the FCAT. ....	33
Predictors of Third Grade Retention. ....	37
Discussion.....	41
Failing the FCAT Reading Test.....	42
Retention in Third Grade.....	47
Limitations.....	50
Policy Implications.....	52
References.....	58

## LIST OF TABLES

Table	Page
Table 1 Bivariate Correlates of Failing FCAT in Third Grade.....	54
Table 2 Bivariate Correlates of Being Retained after Failing FCAT in Third Grade.....	55
Table 3 Logistic Regressions Predicting Failing Reading FCAT.....	56
Table 4 Logistic Regression Predicting Retention (Among Students who Failed Reading FCAT).....	56

## **ABSTRACT**

### **PREDICTORS OF MANDATORY THIRD GRADE RETENTION FROM HIGH-STAKES TEST PERFORMANCE FOR LOW-INCOME, ETHNICALLY DIVERSE CHILDREN**

Tanya Tavassolie, M.A.

George Mason University, 2015

Thesis Director: Dr. Adam Winsler

As a result of the No Child Left Behind (NCLB) policy, Florida, like many other states, has mandatory third grade retention for children who fail the high-stakes reading test, the Florida Comprehensive Assessment Test [FCAT]. Since implementation, the percentage of third graders retained has increased dramatically. Various exceptions and loopholes exist to help with retention and promotion decisions. Little research exists, however, on actual practices in enforcing mandatory retention policies from high-stakes test results. Using data from the Miami School Readiness Project (MSRP), I examined a large ( $N = 27,980$ ) and ethnically diverse (59.2% Latino, 33.3% Black, 7.4% White/other) sample of third graders over five cohorts who completed third grade between 2006-2010. I ask the following research questions: (1) What proportion of children take and fail the FCAT reading test, and of those children, how many actually repeat 3<sup>rd</sup> grade? (2) To what extent is failing the FCAT, and subsequent retention,

related to demographic variables and children's academic performance in 3<sup>rd</sup> grade? (3)

When considering all demographic factors, what are the unique and combined predictors of failing the FCAT, and actually being retained? Multiple logistic regression analyses revealed that even after accounting for demographic variables (e.g. free/reduced lunch, gender, ethnicity, ELL and special education status), those receiving free/reduced lunch, who were not English proficient, and in special education in third grade were more likely to fail the FCAT reading test, and be retained after failing. Latinos and Blacks were less likely to be retained even after failing the FCAT, compared to Whites. Implications for high-stakes testing policy in Florida are discussed.

## **PREDICTORS OF MANDATORY 3<sup>RD</sup> GRADE RETENTION FROM HIGH-STAKES TEST PERFORMANCE FOR LOW-INCOME ETHNICALLY DIVERSE CHILDREN**

High-stakes testing has gained popularity among school districts across the United States as a way of objectively measuring school, teacher, and student performance. The results from these tests are used to determine student promotion and retention, the status and earnings of K-12 teachers within the public school system, and the reputation and rankings of the public schools within the community (Duffy, Giordano, Farrell, Paneque, & Crump, 2008). Some states require students to demonstrate a specific level of academic achievement on a state-mandated standardized test in order to determine student grade promotion and/or graduation (Greene & Winters, 2007; Horn, 2003). About 25 states in the United States operate under this type of policy (Johnson, Thurlow, Stout, & Mavis, 2007). In 2002, Florida joined this list and began using a test-based promotion policy for third graders (Greene & Winters, 2007).

Part of this push for high-stakes testing and objective measures of student and school performance is due to the No Child Left Behind Act (NCLB) in 2001. This act enforced accountability of schools, teachers, and students, and was intended to improve student performance and teaching quality, and ultimately, enhance learning within the schools (Duffy et al., 2008). This thesis focuses on the third-grade high-stakes test policy in Miami-Dade County, Florida.

## **High Stakes Tests**

High-stakes tests are standardized assessments that almost all students within a particular grade level are required to take. These tests are built on the idea that children's learning and knowledge should be held against a particular standard that's assumed to be readily and reliably measured (Heubert & Hauser, 1999). There are important and severe consequences based on test performance, sometimes for the school itself, but students may also receive negative consequences for low performance. These large-scale tests are designed to assess the improvements made by the school in reaching their educational and instructional goals, and importantly, there is pressure on schools to deliver demonstrable improvements in the performance of their students throughout the course of the year and over time (Plake, 2002). These test results are intended to be useful for schools to make decisions about how well students are learning, how effective the curriculum and teaching programs are working, as well as how well the school as a whole performs in educating their students (Plake, 2002).

Notably, recently students and parents in states that have high-stakes testing policies are boycotting and protesting taking the test. There are national "opt out" organizations that help teachers and students navigate the world of testing policies and promote a more well-balanced and fair testing atmosphere among public schools in America (United Opt Out, 2014). Many students and teachers have opted out of high-stakes tests, mentioning that testing takes too much time away from learning in school. Even so, states that still have their testing policies are still having issues. Atlanta has recently been under some pressure for a number of teachers inflating their student's test

scores of children in struggling schools (Fantz, 2015). Lastly, and of importance, Texas, among other states, has recently removed their high-stakes testing policies for third grade students (Texas Education Agency, 2013).

High-stakes testing has many implications for students and their academic trajectory in K-12. The most common student-related consequence of low performance on these high-stakes tests is grade retention (Greene & Winters, 2007). There is much controversy in the literature regarding whether or not high-stakes tests are actually beneficial for students. The American Educational Research Association's (AERA) position statement cautions researchers and policy makers about the potential harmful effects of high-stakes testing programs that have been implemented in areas without adequate educational resources (AERA, 2000). They also advise that schools should consider more than just test scores when making decisions about promotion (AERA, 2000), yet many districts, even in predominantly low-income areas, continue to have high-stakes testing programs. A common fear among researchers is that high-stakes testing will encourage teachers and administrators to develop curricula that are in line with the standardized test, thus excluding other areas that may be important for students' education but are not on the exam (Horn, 2003), in other words "teaching to the test" (Siegel, 2004, p. 225).

Additionally, there have been concerns expressed in the literature regarding the disproportionate negative effects of high-stakes testing on ethnic minorities and historically disadvantaged groups (Frey, 2005; Greene & Winters, 2009; Horn, 2003; Penfield, 2010). There is a greater likelihood that families from low-income, diverse

backgrounds will be disadvantaged and perform at a lower level compared to their more advantaged counterparts, regardless of the teaching and learning taking place in the classroom (Heubert & Hauser, 1999). Hedges and Nowell (1998) found the proportion of Black students who perform in the lower 5% of the distribution is four to eight times the proportion of Whites; likewise the proportion of Whites who score in the top 5% is ten to twenty times the proportion of Blacks (Hedges & Nowell, 1998). However, Hedges and Nowell did not control for income or poverty status among their sample, therefore it is unclear if this effect is occurring primarily because of ethnic differences or socioeconomic discrepancies.

Heubert and Hauser (1999) studied high-stakes testing within K-12 on a national level and explored the outcomes for students. They discuss how African-Americans, Hispanics, and youth in poverty are more likely to perform poorly on these tests and are therefore subjected to the negative consequences (such as grade retention, high school drop-out, etc.) at a higher rate compared to other ethnic groups who are more affluent. In the 1970s, Florida used a high-stakes test for high school graduation decisions, and with it 20% of Black students were failing high school (ten times the proportion of White students, Heubert & Hauser, 1999). Clearly, without partialling out the effects of poverty, high-stakes tests seem to be putting minority students at an even greater disadvantage.

### **Traditional Retention Policies**

Grade retention requires children to repeat a specific grade level, also referred to as nonpromotion (Frey, 2005; Meisels & Liaw, 1993). In 2007, about 10% of students in kindergarten through eighth grade in the United States were retained (National Center for

Education Statistics, 2009). However, most of the research on grade retention predates the implementation of mandatory retention due to high-stakes tests, therefore the literature on retention decisions is based on school performance more generally, and the views of parents, teachers, and administrators.

When examining retention on a national level, Meisels and Liaw reported the proportions of parent-reported retention, distributed across different racial and SES groups, between kindergarten and eighth grade (Meisels & Liaw, 1993). Using a nationally representative sample from the National Education Longitudinal Study (NELS), Meisels and Liaw find a larger percent of low-income and ethnic minority children reported by their parents to have been retained. They also found that the largest number of retention decisions occurred in the earlier grade levels, kindergarten through third grade. The same pattern seems to be true based on studies that explore retention policies in specific cities and states, such as Chicago, Baltimore, and Florida.

**Chicago's Traditional Policy.** Extensive research has been conducted on earlier retention policies. One such study was conducted in the Chicago school district. The Chicago Longitudinal Study began as a way to assess Chicago's retention policies in the early 1990s. This Chicago sample included mostly Black (95% Black, 5% Hispanic), low-income children who attended government-subsidized kindergarten programs (Reynolds, 1992). Under the original policy, Chicago public school districts had a subjective retention procedure where the retention decisions seemed to target more disadvantaged students (Reynolds, 1992). Boys, children who were among the lower socioeconomic brackets, and minority children were more likely to be retained, compared

to their peers who performed similarly (Reynolds, 1992). Additionally, over 20% of the Chicago students were retained before fourth grade, which is higher than the national average and suggests that retention rates may be higher among low-income, at-risk communities. McCoy and Reynolds (1999) explored the predictors of grade retention within this Chicago sample; among their predictors were gender, reading and mathematics achievement, and reading course grades. When a child's first grade school performance dropped one letter grade in reading (i.e. going from a C to a D), the likelihood of retention increased by 10.9 percentage points. Additionally, children who were retained were more likely to have parents without a high school diploma; and children who experienced frequent school moves were more likely to be retained; alternatively, children with parents who were highly involved in school were less likely to be retained (McCoy & Reynolds, 1999).

In Chicago, retained students seemed to initially benefit from the extra year in the grade, but over time these benefits declined. Long-term outcomes for retained children were grim, with many children dropping out of high school all together. Additionally, students who were retained tended to have a decreased perception of their competence in school (McCoy & Reynolds, 1999).

It is important to note here that Chicago operated under a traditional retention policy until 1996, when they implemented a high-stakes test policy. This policy required third, sixth, and eighth graders to take the Iowa Tests of Basic Skills in order to determine promotion and retention (Roderick & Nagoaka, 2005). However, after overwhelming evidence that making promotion decisions based on high-stakes test scores did not

improve student performance in the long run, Chicago put an end to their high-stakes testing policy in 2013.

**Baltimore's Traditional Policy.** In a separate study conducted in Baltimore City Public Schools, Alexander and colleagues discuss predictors of retention (Alexander, Entwisle, & Dauber, 1994). Similarly to Chicago pre-1996, Baltimore, at the time, operated under a regular retention policy where decisions about retention were made at the discretion of the school, teacher, and parents. Alexander et al. (1994) explored a group of Baltimore students who had been retained. They used a sample of roughly 800 students who were entering first grade in 1982. These students were selected from 20 different Baltimore City Public Schools that were chosen based on integration status (percentage of the school's population that was White and African-American) and the socioeconomic level of the community to become part of the "Beginning School Study" (BSS). They found that males, children in poverty with parents who do not have a high school diploma, and African-American children were all overrepresented in the retention group, compared to the never-retained group. Students who were retained performed significantly lower on average reading and math scores during their first quarter, compared to their promoted peers (Alexander et al., 1994). From this study alone, however, it is difficult to determine whether students are being retained at a greater rate because of their poor performance in school or if ethnic minorities and at-risk children are truly being over-selected for retention even controlling for actual performance in school. Most of the existing literature on retention in public schools does not adequately

control for child-level predictors of retention, such as socioeconomic status, gender, and prior performance in school.

Students in Baltimore who were retained had similar outcomes to children in Chicago. Many retained students were catching up to their promoted peers during their repeated year, but were still behind in school generally. Likewise, their performances tended to decline over time and any gains that retained students made diminished greatly over time (Alexander et al., 1994).

**Florida's Kindergarten Policy.** In a recent paper by Winsler et al. (2012), drawing from the same population as the current study, predictors and prevalence of delayed kindergarten entry and kindergarten retention were examined in a low-income, ethnically diverse sample. Using a large community sample in Miami, Florida, this study found that children who were retained in kindergarten were typically struggling academically and/or behaviorally (Winsler et al., 2012). Explained bivariately, retained kindergarteners scored lower on cognitive, language, and fine motor assessments one year earlier. Additionally, girls were about half as likely, compared to boys, to be retained in kindergarten; and Black kindergarteners were twice as likely to be retained compared to White kindergarteners. These findings alone seem to suggest that ethnic minorities are being put at a greater disadvantage compared to Whites, however those analyses did not control for student income or language ability.

In a multi-step logistic regression, the authors included demographic variables (child ethnicity, gender, free/reduced lunch and ELL status) in step 1 in order to examine the predictive value of these variables. They found that when simply including these

variables into the model, gender and free/reduced lunch status were the only significant predictors of kindergarten retention. Girls were less likely to be retained in kindergarten compared to boys, and those receiving free/reduced lunch were more likely to be retained compared to those who were not. These are similar findings to what we would expect and what others have found.

In the second step, however, the authors included children's school readiness (cognitive, language, fine motor, social skills, and behavior problems) and preschool experiences (i.e. preschool type) into the model. All of the school readiness measures significantly predicted kindergarten retention, with children who perform better having reduced odds of being retained (Winsler et al., 2012). Here also, when step 2 was included in the model, the gender effect was no longer significant, suggesting that the reason boys were more likely to be retained compared to girls was because of their initial lower school readiness. However, free/reduced lunch continued to be a significant predictor of kindergarten retention, even after controlling for demographic variables and school readiness. With only steps 1 and 2 in the model, the authors noted that there appears to be no evidence for a gender bias but there was a bias against those who were in poverty in kindergarten. Of meaningful interest, ELL status significantly predicted kindergarten retention when school readiness measures were added to the model, with ELL students being *less* likely to be retained compared to native English speakers.

In the third step of the regression model, the authors included children's performance in kindergarten, which significantly predicted kindergarten retention, with children performing worse in school being more likely to be retained. ELL status

significantly predicted kindergarten retention, in the similar trend as before. However, of particular interest, is even with school performance in kindergarten in the model, black children were less likely to be retained compared to white children (Winsler et al., 2014). This is evidence that there doesn't seem to be an ethnic bias in the Miami sample for early retention decisions, after accounting for all demographic and performance variables. These particular findings can also be explained by the cultural context in which Miami is positioned. In Miami, Whites are the minority and it is perhaps this ethnic diversity that protects Latinos and Blacks from retention in earlier years. Importantly, here, we see that when simply examining the data bivariately Blacks are being retained more often than Whites, however this difference seems to be due to poorer performance on school readiness and due to a variety of other risk factors associated with poverty found among African American students.

The major difference between previous research and the current thesis is that prior studies have analyzed regular retention throughout kindergarten and elementary school, that is, retention decisions that are not based on high-stakes testing. There is a need in the field to perform similarly rigorous studies on the predictors of performance on high-stakes tests, and to explore the outcomes for students in states where these policies being practiced. There are only a handful of studies have been performed on Florida's policy, which will be examined below.

## **Florida**

Since 2002, Florida has been using the Florida Comprehensive Assessment Test (FCAT) to help make promotion decisions in grades 3-12. The state of Florida enacted this

practice in order to help bring an end to social promotion, which allows students to advance to the next grade level for social developmental reasons, regardless of academic performance (Greene & Winters, 2009). In Florida, third graders are required to achieve at least a level 2 (out of a possible 5) on the reading FCAT in order to advance to the fourth grade (Greene & Winters, 2007, 2009; Siegel, 2004). Students who do not perform up to this set standard are mandated by the State Department of Education to repeat third grade. The implementation of this policy is intended to enhance student learning, school performance, teaching quality, and increase expectations for students (Siegel, 2004). Those students who are retained are assigned to a “high-performing teacher” for the repeated year at the same school (Greene & Winters, 2007; Stewart, 2011). These teachers are designated high-performing by their student performance data and above-satisfactory ratings on their teaching evaluations. Additionally, students who are retained must also receive additional reading time each day throughout the repeated year, and the school must provide an individualized improvement plan for the student (Greene & Winters, 2007).

The state of Florida has developed ways for students to bypass the FCAT retention policy. According to Florida state law, if a student has demonstrated reading deficiency, as revealed by performance on the FCAT reading test, the student must be provided intensive reading interventions (Florida Department of Education, 2013). This reading intervention can be provided through a district summer reading camp, however, if the student does not demonstrate enough improvement after this summer program, he/she must be retained. If the student demonstrates successful independent reading and is

performing at or above grade level in reading, based on other standardized tests and/or a portfolio that the teacher creates, the school may re-evaluate the student using alternative assessments in order to promote the student midyear, during the repeated year (Florida Department of Education, 2013).

Greene and Winters evaluated Florida's retention policy and focused on the students who were formally exempt from being retained due to failure on the FCAT (Greene & Winters, 2009). Exemptions from the retention policy include: (1) English language learners who have had fewer than two years of instruction in the English for Speakers of Other Languages (ESOL) program, (2) students with disabilities with a specified individual education plan (IEP) that removes the student from taking the test, (3) students who perform at or above the 51<sup>st</sup> percentile on the Reading SAT-10 or another approved standardized test (either before or midyear during the repeated year), (4) students whose teachers develop a portfolio demonstrating reading competence up to a Level 2 on the Reading FCAT (either before or midyear during the repeated year), (5) students in special education who were previously retained and have already participated in an intensive reading program for more than two years, yet still show a deficiency in reading abilities, and (6) students who have received additional intensive reading attention for two or more years who still demonstrate a deficiency, and who have already been retained for a total of two years (Stewart, 2011). Exemptions three and four are used when students have shown to be capable of the necessary skills needed for promotion in all other areas, other than the FCAT, whereas the other exemption categories are more objective and based on the assumption that retention would be detrimental to these

specific students (Greene & Winters, 2009). Using their statewide dataset, Greene and Winters discovered that while 54.1% of students who failed the FCAT were retained, the other 45.9% were promoted because they fell into one of the formal exemption categories (e.g. 3.1% of students who failed the FCAT were promoted because of a student portfolio, and 7.8% of students who failed the FCAT were promoted due to having a disability and receiving extensive reading instruction (Greene & Winters, 2009)).

Since the implementation of this policy, retention among third graders in Florida overall has increased from 2.9% in 2000 (two years before the policy took effect) to 11.7% in 2002 (two years after the policy was in place) (Greene & Winters, 2009). There is some evidence to suggest there could be a disproportionately large minority student population who are being retained (Greene & Winters, 2009; Penfield, 2010). Using a statewide dataset from the Florida Department of Education, Green and Winters (2009) found that compared to White students, African-American students were 4% more likely to be retained, and Hispanic students were 9% more likely to be retained, even though all groups performed similarly on the FCAT. This is true, even controlling for other factors such as language proficiency, eligibility for free and reduced lunch, and IEP status.

Though Green and Winters, along with many others, have contributed extensively to the research on this particular topic, there are still many gaps that need to be addressed. Green and Winters (2007, 2009) did not assess other predictors of retention and poor performance on the FCAT, specifically, overall school grades. They also examined the immediate effects of the high-stakes testing policy on third graders, right after the initial implementation of the policy. Over time, schools and districts may change the way they

use the policy once the policy doesn't feel so new to the schools and teachers. This thesis will fill in these gaps by using a variety of predictors of FCAT performance and third grade retention, including ELL and special education status. Importantly, this thesis examines this policy, using school records, a few years after its initial implementation in order to capture some of the more recent, real-world implementation of such practices.

In a technical report produced by the Florida Department of Education using descriptive measures, performance on the FCAT varied by ethnicity in fourth grade (Florida Department of Education, 2002). African-American students were consistently performing below Hispanic-American students, who were performing behind White, non-Hispanic students (Florida Department of Education, 2002; Siegel, 2004). This is further supported by the Florida State Department of Education statistics that report a larger percentage of African-American and Hispanic students being retained, in general across grades (Florida Department of Education, 2011).

A major concern with these studies is that they neglect to control for various confounding variables that we know influence differences we see among ethnic groups. Many of the findings from these studies make it seem like ethnic minorities are being discriminated against, when in fact, it could very well be that ethnic minorities are more likely to be in lower socioeconomic brackets that make it difficult to succeed academically. This illustrates the first potential issue regarding minority ethnic groups and retention policies. It may be that minority groups are being retained more often *because* of their poorer performance in school and the variety of risk factors stacked against them (e.g. poverty). If this is the issue, then the educational system in the United

States is putting ethnic minority children at an extreme disadvantage because of their particular circumstances. However, there may be a different, larger, issue at hand. There may be active discrimination with the implementation of retention policies. This would suggest that even when comparing children who perform similarly on tests and school coursework, and even with partialling out these differences, we still see minority ethnic groups having a greater likelihood of being retained. In order to describe this, studies would have to control for various demographic and child-level measures in order to truly see this kind of discrimination. This second and different problem would suggest implementation of educational policy that is particularly discriminatory.

### **The Current Study**

Since the implementation of Florida's high-stakes test policy in 2002, surprisingly little research has been conducted regarding the effectiveness of this policy and how it has impacted children in Florida. The current thesis takes a closer look at what is happening in Miami, specifically, in order to accurately examine the pervasiveness and predictors of retention in third grade. Unlike the rest of Florida, Miami is an especially ethnically and linguistically diverse area. The majority of the Miami population is Hispanic/Latino (65.6%, US Census Bureau, 2013), and even more speak a language other than English at home (72.2%). In Miami, children might grow as old as 5 years of age before they begin speaking, or are exposed, to English (Winsler et al., 2012). This culturally enriched community provides a unique window through which we can examine how statewide educational policies are implemented for this population.

The current study uses a sample of children that belong to the largest school district in the state of Florida (Miami-Dade County), the fourth largest school district in the country, that serves about 348,021 children in Florida (Miami-Dade County Public Schools, 2014). This low-income, ethnically, and linguistically diverse sample is a part of a large-scale university-community initiative that began in 2002 called the Miami School Readiness Project (Winsler, Karkhanis, Kim, & Levitt, 2013; Winsler et al., 2008). Participants were initially recruited based on childcare services; they were receiving government subsidies to attend childcare, or attending fee-supported or Title-1 public school pre-kindergarten programs, and the researchers have continued to follow them throughout their public school educational trajectory in the county. This study uses a unique group of children, where many live in extreme poverty. Using this sample allows us to capture the diversity of the low-income population in Miami and what their experiences are with the FCAT and third-grade retention.

The aims of the current study are to first, examine the degree to which demographic variables are related to failing the FCAT reading test, and subsequently being retained in third grade. To the extent possible, we will also examine whether or not some students fall into the retention policy-exemption categories. Second, we examine the role that these demographic variables play in predicting failing the FCAT reading test and third-grade retention. And finally, since many prior studies have simply studied these comparisons bivariately without rigorous data analysis and controls, we compare the extent to which these demographic variables are still considered good predictors of failing and retention, after controlling for performance in third grade. A primary goal of

the study is to predict retention, while being diligent about controlling for specific child- and family-level variables, so that we can assess whether this particular policy discriminates against specific minority groups. The following research questions were addressed:

(1) What proportion of an ethnically and linguistically MSRP sample of third graders take and fail the FCAT, and of those children, how many actually repeat third grade?

(a) To what extent are children in this sample appearing to be falling under one or more of the retention policy exemption categories?

(2) To what extent is failing the FCAT, and subsequent retention, related to child ethnicity, gender, poverty status, disability status, English language learner (ELL) status, English proficiency, and children's academic performance in third grade?

(a) To what extent are children failing the FCAT but doing fine in school?

(3) When considering all demographic factors, what are the unique and combined predictors of both failing the FCAT, and actually being retained?

a) Does performance in third grade explain part of the predictive relationship between ethnicity and failing the FCAT, and subsequent retention?

Based on previous research, we expected our sample to share similar demographic variables represented in the fail (FCAT reading test) and third grade retention groups as the former research conducted on the Chicago, Baltimore, and Florida samples (Alexander et al., 1994; Greene & Winters, 2009; Reynolds, 1992). We expected there to be a larger number of students in poverty, special education, and of ethnic minorities, represented in the failed and retained categories. We anticipated that some, but not all, the students who failed the FCAT reading test and were promoted would fall into one or more of the exemption categories. Further, we expected that ethnicity, special education, and poverty status would each have a large impact on predicting failing the FCAT and subsequently being retained. ELL status was expected to be a protective factor against failing and retention, based on prior research with this sample (Winsler et al., 2012).

## METHOD

### Participants

This study explored a subset of the larger sample of five MSRP cohorts (from 2006-2010) of students in Miami Dade County, Florida. Our sample consists of an ( $N = 27,980$ , 50.3% male) ethnically diverse (59.2% Latino, 33.3% Black, 7.4% White/other) sample of students who completed third grade from the Miami School Readiness Project (Winsler et al., 2008). This sample reflects essentially the entire consenting population (92%) of four-year-old children in Miami-Dade County during the 2006-2011 academic years who were receiving government subsidies in order to attend non-Head Start childcare facilities, and made it to third grade in the county (De Feyter & Winsler, 2009). The majority of the sample was in poverty, as measured by receiving free and reduced lunch in kindergarten (79.4%). About 57% of the students in this sample were considered ELL by the school district, and 23.1% were not English proficient even in third grade. For my sample, since I am concerned with performance and retention in 3<sup>rd</sup> grade and with performance on the FCAT, I have three large subgroups that need to be clarified. The largest sample that I am concerned with is all students who have began and completed third grade an initial time represented by receiving end-of-the-year grades in third grade ( $N = 27,980$ ). For some analyses, however, I am limited to the number of students who took the FCAT ( $N = 27,703$ ).

In order to tell if a student was retained, he/she needs to be around the entire subsequent year (either in third grade again, or in fourth grade). For the questions regarding retention as the outcome, only those who started and completed third grade and who started and completed the next year of school (either third grade again or fourth grade) are involved in the analyses ( $N = 25,739$ ). Due to natural longitudinal attrition, we lose some children just because this is a longitudinal study following children for many years.

## **Measures**

**Florida Comprehensive Assessment Test (FCAT).** The FCAT is a mandatory high-stakes standardized exam given to third through twelfth graders in Florida, in English only. This exam has a reading and a mathematics portion, however mandatory retention/promotion decisions for third grade are made based on performance on the reading portion of the exam only. Standard scores for this exam range from 100 to 500 points, and the ordinal proficiency scale ranges from 1 to 5. A score of 1 indicates “little success with the challenging content” and a score of 5 indicates “success with the most challenging content” (FCAT, 2011). Performance at level 3 is considered satisfactory. In third grade, a level 2 and above will permit a student to be promoted to fourth grade, while receiving a level 1 results in “mandatory” retention (Stewart, 2011).

Most of the questions on the FCAT are multiple choice, however there are also some short answer and detailed responses as well (Florida Department of Education, 2004). Some of the reading test questions require students to give answers responding to an excerpt that they read. The FCAT has been shown to be a reliable test with high

internal consistency. The Cronbach alpha's for the reading and math tests are 0.91 and 0.88, respectively (Florida Department of Education, 2004).

**Norm-Referenced Test (NRT).** Based on school records, we also have a measure of student performance on the Stanford Achievement Tests (SAT-10, 2011) given to second through fourth graders in the state of Florida, with  $n = 9,396$  with third-grade SAT scores. This is a reading and math battery that allows comparison to test scores of children nationally.

**Special Education status.** In order for a student to receive a code for special education, they are required to have a listed primary exceptionality code in third grade. These codes include the following groups: intellectual disability ( $n = 164$ , 0.4%), speech impaired ( $n = 486$ , 1.3%), language impaired ( $n = 113$ , 0.3%), visually impaired ( $n = 11$ , <0.01%), deaf or hard of hearing ( $n = 30$ , 0.1%), specific learning disabled ( $n = 1,505$ , 3.9%), orthopedically impaired ( $n = 57$ , 0.1%), autistic ( $n = 506$ , 1.3%), severely emotionally disturbed ( $n = 14$ , <0.01%), emotionally handicapped ( $n = 337$ , 0.9%), mentally handicapped ( $n = 98$ , 0.2%), traumatic brain injured ( $n = 5$ , <0.01%), other health impaired ( $n = 227$ , 0.6%), and gifted (who were excluded from the analyses,  $n = 4145$ , 10.7%), with the rest of third graders ( $n = 30,920$ , 80%) not in special education. If any of these codes were present in third grade, the child was coded (0 = does not have ESE code in third grade; 1 = has primary exceptionality code in third grade, excluding gifted students) as receiving special education. Students with gifted classifications were intentionally excluded (and given a 0) from this category because they are given this classification based on a different set of exceptional strengths.

**Grade Point Average (GPA).** At the end of each academic year, students receive grades from their teachers based on their performance in their subject domains. Third-grade subjects include art, science, social studies, music, reading, language arts, English as a second language, math, and physical education. However, not all students take all these subject areas. Grades are based on a 5-point scale, where 5.0 = A, 4.0 = B, 3.0 = C, 2.0 = D, 1.0 = F. Composite scores were calculated by averaging all grades that children received across all their subjects within their third grade year.

**B average or Better.** A “B average or Better” variable was calculated by using the average grades of all coursework in third grade. A categorical variable was created where students who received at or above a 4.0 average are coded as 1, and at or below a 3.9 average are coded as 0. This allows for the examination of those students who failed the FCAT in third grade, but still performed well in their coursework. This variable was also created in order to distinguish those who performed well in school, from those who received poorer grades.

**Retained in Grade 3.** Children were classified into a dichotomous variable that distinguishes retained students from non-retained students. A “yes” (1) on this variable indicates the student appeared in third grade on time, continued through the year and received end-of-the-year grades in third grade, and then, in the following year, the student returned and completed third grade again and received end-of-the-year grades the following year for third grade again. A “no” (0) on this variable indicates the student appeared on time in third grade, continued through the year and received end-of-year-grades in third grade, and appeared the following year in the next grade (fourth grade),

and continued through the year and received end-of-year grades in fourth grade. Students who may have been promoted midyear during the retained year would look the same as children who were promoted to fourth grade, due to the nature of the information in this dataset. Since we are defining promotion with the completion of the subsequent year's courses, we are not able to tell the difference between a midyear promotion and a regular promotion. From preliminary data received from the district, it seems that this midyear promotion option is very rarely being used as a promotion option.

**English use at home (ELL status).** A student was given a 1 on the ELL variable if the student had ever been considered an ELL prior to third grade. This was acquired from parent-reported home language at kindergarten/school entry. A student who received a 0 on this variable was not taking ESOL classes in kindergarten or ever.

**English proficiency in third grade.** Students' classification as ELL by the school district was based on whether their parent indicated another language was spoken at home in kindergarten, and these students were coded as a 1 (vs. zero). For these individuals, if they still received ESOL services in third grade, meaning they were not yet proficient in third grade, they received a 1. ESOL services were first provided to kindergarteners who complete the Miami-Dade County Oral Language Proficiency Scale-Revised (M-DCOLPS-R) assessment (Abella, Urrita, & Schneiderman, 2005). This test measures aural comprehension and oral production, and those students who are determined to be ELLs are provided with English for Speakers of Other Languages (ESOL) instruction (Abella et al., 2005). The ESOL levels are marked 1-5 with levels 1 and 2 indicating beginning English learners who still have much difficulty, levels 3 and 4

being advanced stages of English learning, and level 5 is considered proficient in English. If parents answer “yes” to questions that ask about the child’s non-English language use at home, then the child is subject to be placed into one of five levels (Level 1 for beginners, Level 5 for fully proficient in English) according to the test. Students who were not considered proficient in English are required to take ESOL classes until their English proficiency is at a Level 5 (Winsler, Kim, & Richard, 2014).

**Free/reduced lunch status.** Children in poverty are eligible to receive free or reduced price lunch. Those families who are 130% of the federal poverty line are eligible for free lunch, and those who are 185% of the federal poverty line are eligible for reduced-price lunch. At the beginning of each school year, parents and families apply for free/reduced lunch. A select group of students may be eligible for direct certification (USDA, 2010). In Florida, the districts obtain information about TANF (Temporary Assistance for Needy Families) and then match this information to their enrollment. If students are listed on both lists, then they receive free/reduced lunch through direct certification. In our sample, if students received free or reduced lunch in third grade then they received a 1, if they did not apply or did not receive free or reduced lunch in third grade, they received a 0.

**Gender.** The parents reported their child’s gender. Based on school records received from the district, females were given a 0, and males were given a 1.

**Ethnicity.** Parent-reported child ethnicity is provided by the school district every year. For the children in this sample, a number of ethnic categories were collapsed in order to obtain a three-level ethnic variable. “Hispanics” included anyone who reported

Hispanic/Latino or Hispanic and some other racial group, on self-reported school registration forms, which closely aligns with the U.S. Census Bureau definition of Hispanic as including anyone who responds with Hispanic, Latino, or Spanish origin (U.S. Census, 2010). “Black” included anyone who reported African-American/Black/Caribbean, or Black and some other racial group, which also closely aligns with the U.S. Census as reporting anyone who marks African-American, Black, or anyone who has origins in Black racial groups in Africa (U.S. Census, 2010). “White/other” included anyone who reported White, Asian/Pacific, or a combination of other racial categories.

### **Analytic Plan**

First, I examined the data in a univariate descriptive manner, by determining the proportion of children who took and failed the FCAT, and of those children, the number that actually repeated third grade. Next, I determined the extent to which failing the FCAT was related to demographic variables bivariately (ethnicity, gender, poverty and ELL status, and academic performance), in third grade, using chi-square analyses and t-tests. Lastly, I determined the unique and combined multivariate predictors of failing the FCAT by including all demographic variables in as predictors. Subsequently, I selected only those students who failed the FCAT reading test and entered all demographic variables as predictors, and the dichotomous retained variable as the outcome. These analyses were preliminarily analyzed in a multi-level hierarchical linear model (HLM),

and similar results were found, therefore for the purposes of this thesis, HLM was not used.<sup>1</sup>

---

<sup>1</sup> For the publication we will use multi-level HLM to accurately describe the relations between the predictors and the outcomes for these two logistic regression analyses.

## RESULTS

### Research Question 1

For the 27,980 MSRP students who completed third grade (between 2006-2011), almost all of them ( $n = 27,703$ , 99%) took the FCAT reading test. Out of the third graders who took the test and were on time to third grade, 15% ( $n = 4,284$ ) failed (received a 1 on the FCAT reading test). There were 25,739 children with data the following year after third grade, and of those children, 15% ( $n = 3,798$ ) failed the FCAT reading test. And of this group, 52% ( $n = 1,999$ ) were retained in third grade.

**Retention policy exemption categories.** To begin, I looked deeper at the children who did not take the FCAT. This is outside the requirements of the exemption category, and simply helps understand more about these students who never even took the FCAT test. About 2% ( $n = 627$ ) of third graders did not take the FCAT reading test but finished third grade. Of these students, 46% were diagnosed with Autism, 23% had an Intellectual Disability, 25% had some other disability, and 6.7% did not have any disability according to the school.

In order to explore the number of children affected by formal retention exemption policies, I examined the students who failed the FCAT but were *not* retained in third grade ( $n = 1,799$ ). The only two exemption categories that I was not able to address are the alternate portfolio, and the additional reading instruction that students may have

received. Our data do not allow for the analysis of these categories. However, I examine the degree to which those students who failed the FCAT but were not retained were in the ESOL program, were given an IEP, or performed well on an alternate standardized test, SAT-10. I analyzed this univariately by first selecting those students who failed the FCAT but were not retained in third grade and observed the extent to which they fall under the above categories.

*Designated, individualized education plan (IEP).* Of the students who failed the FCAT ( $n = 4,284$ ), 7.6% ( $n = 327$ ) were in special education. These students, presumably, fall under the exemption category of having a designated, individualized education plan that removes them from being retained even if they fail the FCAT reading test, but we do not know for sure. Unfortunately, we do not have any information about the specific individualized plan for the students who were in special education.

*Alternative standardized exam.* The next exemption category involves the students who took an alternative standardized exam and scored above the 51<sup>st</sup> percentile. Of the students who failed the reading FCAT, a subsample of 481 (11.2%) students had scores for another reading standardized test that was available to us in third grade. Important to note here, is that these students could have taken other standardized tests (either in school or out of school) and we would not know because we don't have access to those test scores. So, out of the students known to take another reading standardized exam as part of typical district testing, 12% ( $n = 60$ ) of students received a score above the 51<sup>st</sup> percentile, and presumably this made them eligible to be promoted to fourth grade even though they failed the FCAT reading test.

*Taking ESOL for less than two years.* For the next exemption category, I looked at the proportion of students who were in ESOL classes for less than two years, out of those students who failed the FCAT but were not retained. There was only one student for which third grade was his/her first time in the ESOL program. Since the students in our sample started public school in kindergarten, most of them are proficient by third grade (about 90%), so using our dataset it is difficult to address this question entirely (Kim, Curby, & Winsler, 2014). However, an important characteristic to note, outside of the formal exemption policy, is that of the students who failed the reading FCAT, 12.9% ( $n = 551$ ) were considered not yet proficient in English in third grade.

*Retained already and have individualized education plan.* The final exemption category that can be explored is those students who have an individualized education plan given by the school, and who were retained already before third grade. These students were not included in any analyses throughout the rest of this thesis; their data are in different columns in the dataset because they are off the normal trajectory of schooling since they have been previously retained. There were 3,329 students who were retained in either first or second grade, and 4,244 students who were in special education in either first or second grade. There were 363 students who had both been retained prior to third grade and had a disability. Of this group of students, 189 (52%) failed the reading FCAT. We can only assume that these 189 students were not retained a second time because of this FCAT reading test failure, but we cannot be sure.

From our dataset, it seems that not many students are being exempt from the retention requirement after failing the FCAT reading test. Only a small portion of

students are being promoted to fourth grade based on their placement into these categories, at most 9% of children in our sample seem to be promoted for either being in special education or scoring well on an alternate exam. Since we cannot capture all the reasons why students were promoted even though they failed the FCAT, there appear to be other external factors explaining why these students were promoted, which we are exploring in the remaining research questions.

## **Research Question 2**

**Failing FCAT Reading Test.** I conducted a bivariate 2x2 cross tabulation, chi-square test between the students who did and did not fail the FCAT (i.e. received a score of 1 or below) with various categorical demographic variables. It is important to conduct bivariate analyses first in order to be comparable to previous studies, and to demonstrate similar relationships without controlling for other relevant demographic variables, and especially to report the changes in these relationships once they are included in multivariate models. Table 1 shows row percentages, indicating the percent of children with different demographic characteristics who failed the FCAT reading test. Ethnicity, gender, English use at home, English proficiency, special education, free/reduced lunch, and doing well in school were all related to poor performance on the FCAT. As a reminder, out of the full sample of third graders who had FCAT reading scores, 15% failed.

Out of the total group of Black children, 22.9% of them failed, compared to 12.4% of Latino/Hispanic students and 6.4% of White children ( $\chi^2(2) = 639.76, p < 0.001$ ). This means that Blacks were nearly two times more likely to fail the FCAT

compared to Latinos and almost 4 times more likely to fail compared to Whites. Of the boys in our sample, 17.7% of them failed the FCAT, compared to 13.3% of girls ( $\chi^2(1) = 109.33, p < 0.001$ ). Among ELL students, 13.5% of them failed the FCAT compared to 18% of those who were non-ELL ( $\chi^2(1) = 101.663, p < 0.001$ ). Importantly, however, 22.1% of children not proficient in English in third grade failed the FCAT compared to 13.5% of those who were ( $\chi^2(1) = 284.33, p < 0.001$ ). This indicates that children who are not proficient in English in third grade are almost twice more likely to fail the FCAT compared to children who are proficient, but that being an ELL in general is a protective factor (as long as English proficiency is there). Children in special education ( $\chi^2(1) = 1665.74, p < 0.001$ ) were almost three times more likely to fail the FCAT (41.7% of those in special education versus 12.3% of those not in special education). Children receiving free/reduced lunch were almost three times more likely to fail the FCAT compared to those who were not receiving free/reduced lunch ( $\chi^2(1) = 446.35, p < 0.001$ ; 17.8% versus 6.4%).

**Retention in Third Grade.** Out of the children who failed the FCAT reading test in third grade and were around the following year, 52.6% were retained in third grade. Table 2 shows row percentages indicating the percent of children with different demographic characteristics who were retained in third grade, after failing the FCAT reading test. Gender, English proficiency in third grade, special education, and free/reduced lunch status were all significantly related bivariately to being retained in third grade having failed the FCAT reading test.

Students who were retained (after failing the FCAT) were more likely to be male ( $\chi^2(1) = 47.61, p < 0.001$ ) compared to female, receive free/reduced lunch in third grade ( $\chi^2(1) = 14.46, p < 0.001$ ) compared to not receiving free/reduced lunch, not be English proficient in third grade ( $\chi^2(1) = 16.97, p < 0.001$ ) compared to being fully proficient in English, and be in special education ( $\chi^2(1) = 164.92, p < 0.001$ ) compared to typically developing children.

**Performance in Third Grade.** Students who failed the FCAT and were retained also had a lower GPA ( $M = 3.15(0.58)$  vs.  $M = 3.59 (0.514)$ ) in third grade ( $t(3796) = -24.25, p < 0.001$ ), and a lower raw score on the FCAT reading test overall ( $M = 206.59(40.96)$  vs.  $M = 234.38 (26.65)$ ;  $t(3796) = -24.48, p < 0.001$ ) compared to those who were promoted. This suggests that children who were retained were performing poorer in school overall and on the FCAT reading test continuous score as well.

For all students who took the FCAT, continuous FCAT reading scores were correlated  $r = .667$  ( $p < 0.001$ ) with GPA in third grade. While these two variables are highly correlated, there are a fair number of students who are doing well in school performance, but are still failing the FCAT reading test. Fifteen percent ( $n = 658$ ) of the students who failed the FCAT reading test were actually doing fine in school, as marked by receiving a B average or better in all coursework in third grade. Bivariate analysis revealed that 30% of the Whites who failed the FCAT reading test were doing fine in school, compared to 18.6% of Hispanics and 12.1% of Blacks ( $\chi^2(2) = 53.289, p < 0.001$ ). Of the students who received a B average or better in their classes but also failed the FCAT reading test, there was no significant difference in retention outcomes among

ethnic groups. Also examining bivariately, 28.1% of those students who were not receiving free/reduced lunch, failed the FCAT but got a B average or better, compared to 14.6% of those who received free/reduced lunch, ( $\chi^2(1) = 45.43, p < 0.001$ ). When looking at the gender distribution, we see that 19.6% of females failed the FCAT but earned were doing fine in school, compared to 12.9% of boys ( $\chi^2(1) = 34.68, p < 0.001$ ). However, 17.4% of ELL students who failed the FCAT, were doing well in school coursework, compared to 14% of non-ELLs ( $\chi^2(1) = 8.98, p < 0.001$ ). No significant differences were found for those with and without a disability or children who were not proficient in English. There seems to be a trend that children with a number of risk factors are also likely doing poorly in school (so there are fewer of them who fail the FCAT but are doing OK in school more generally compared to White students, with 39% of Whites who failed the FCAT but were doing fine in school yet were still retained).

Simply explaining these data bivariately only tells part of the story. We see clear gender, ethnic, ELL, and income differences in the probability of failing and being retained. However, this is explaining these relationships without controlling for the other variables that likely influence the relationship. For example, we found that Blacks were more likely to fail compared to Whites, however, we do not know yet how much of this relationship is due to income differences.

### **Research Question 3**

**Predictors of Failing the FCAT.** A hierarchical logistic regression was used in order to determine the unique and combined predictors of failing the FCAT, and subsequently being retained. Results, including odds ratios and standard errors, are

reported in Table 3. The first step assessed the predictive value of child and family demographic variables (child gender, ethnicity, free/reduced lunch status, ELL status, English proficiency in third grade, and disability status in third grade). Next, in the second block of the regression (the two right most columns of Table 3), average GPA in third grade was added. This method was chosen in order to control for variables that likely influence the likelihood of not passing the FCAT, and to explore the unique contributions of each variable (ELL status, gender, ethnicity, income, and disability status) while holding all other predictors in the model constant. Additionally, adding GPA in third grade gives us more information about how the children are doing in school and allows us to determine if there is any bias in the system of retaining and promoting students (i.e., if after controlling for actual performance in 3<sup>rd</sup> grade, there is still ethnic difference, this would be more suggestive of overt discrimination in the system).

The child/family demographic variables of ethnicity, free/reduced lunch, gender, English use at home, ELL status in third grade, and disability status, significantly predicted failing the FCAT reading test in third grade. Notably, we were able to correctly identify students who passed and did not pass the FCAT reading test 85.2% of the time, simply by using demographic variables and with no knowledge about what is actually going on in the schools (i.e. teaching quality data, classroom data, child learning data, etc.). While controlling for the other demographic variables (most importantly income), the odds of failing the FCAT reading test were 50% higher for Latinos and 336% higher for Blacks compared to Whites. Additionally, when controlling for the other demographic variables, males had 18% greater odds of not passing the FCAT reading test compared to females,

and those receiving special education in third grade still had 560% greater odds of failing, compared to those typically developing students. Interestingly, ELLs (those who did not report speaking English at home) were less likely to fail the FCAT reading test, compared to non-ELLs. However, ELLs not yet proficient in third grade had 283% greater odds of failing, compared to those proficient in English.

By simply looking at these variables in step 1, we examined the child and family demographic variables that contributed to the prediction of failing the FCAT reading test, while controlling for the others. In step 2, we explored the predictive value of GPA in third grade. When adding third grade performance to the model (the two right most columns in Table 3), our prediction accuracy increased to 87.5%, suggesting an improvement in our ability to accurately predict failing the FCAT based on these variables. When accounting for general performance in third grade, Blacks still had 54% greater odds of failing the FCAT compared to Whites. This effect was reduced in half compared to step 1, illustrating that some of the reason for why Blacks were more likely to fail the FCAT compared to Whites is because their GPA/performance in third grade is lower. However, Latinos did not have significantly greater odds of failing compared to Whites in step 2 meaning that the reason Latinos were more likely to fail the FCAT compared to Whites in step 1 was because of their poorer performance in school in third grade. Similar to step 1, Latinos were less likely to fail compared to Blacks, even controlling for ELL status, income, and performance in school. This demonstrates a level of bias in the FCAT that works against Black students controlling for other factors.

When controlling for demographic variables and third grade performance, those receiving free/reduced lunch still had 32% greater odds of failing compared to those who did not receive free/reduced lunch. Because this effect is still large, this is further evidence for bias against those who are from less fortunate backgrounds, regardless of actual performance in school.

Also notable in step 2 is that males became less likely to fail, compared to females. This change between step 1 and step 2 suggests that males were more likely to fail the FCAT due to their poor performance in school, but once we control for that, we see this relationship flip with girls being more likely.

Those with a disability were almost four times more likely to fail the FCAT compared to those who do not have a disability, with all other demographic variables held constant. This trend was similar to the previous step, however somewhat reduced. This suggests that part of the reason that students in special education were more likely to fail the FCAT was because of their poor performance in school. But, even controlling for performance in school, we still see differences in the odds of failing for children who are in special education. Those with disabilities who take the FCAT appear to be particularly at a disadvantage with high-stakes tests.

ELL students were still less likely to fail compared to non-ELL, even with third grade performance controlled. However, children who were considered not proficient in English in third grade were still 77% more likely to fail the FCAT with third grade performance in the model. The ELL trend was similar in both step 1 and 2, suggesting a potential protective quality of being an ELL. However, if ELL children are not proficient

in English by third grade, then they are more likely to fail even when controlling for the child and family level demographic variables and performance in school. This suggests a potential problem in the educational system because it requires children who are not proficient in English to take (and do well on) the FCAT.

Finally, when considering all these child and family level predictors in the first block, when GPA was added to the second block, those with higher GPA's were less likely to fail the FCAT reading. For every unit increase in GPA (e.g. going from a 3.0 to a 4.0) children have a 90% decrease in the odds of failing the FCAT.

Overall, when considering just demographic variables alone, we see significant trends that disadvantage ethnic minorities and children from low-income backgrounds on the FCAT. However, when GPA was added in the second step, some of these relationships changed. This kind of analysis allows us to explore the relationship between these variables and failing, when controlling for the other variables in the model.

**Predictors of Third Grade Retention.** A similar second hierarchical logistic regression was performed to predict retention in third grade after failing the FCAT. This analysis was performed identically to the previous regression (see Table 4 for results), however only selecting those students who failed the FCAT reading test. We chose this method in order to further explore those students who failed this test and see which demographic variables could significantly predict retention in third grade. In the second block, third grade school performance was added later to the model in order to determine if GPA helped predict retention after controlling for demographic variables. Using this hierarchical model also helps determine if third grade performance contributes to

retention decisions and helps determine with actual performance in 3<sup>rd</sup> grade in the model, does systematic bias in decisions related to retention appear to be present.

In the first block, child and family level demographic variables (ethnicity, free/reduced lunch, gender, ELL status, English language proficiency, and disability status) were added in as predictors. Of these predictors, free/reduced lunch, Latino group membership (compared to Blacks), male gender, English language proficiency, and disability status were significant. Using this model, we were able to correctly predict 59.7% of children's retention outcomes, using just demographic variables alone (this analysis does not include any information about who falls into the exemption categories, which would be a systematic way of determining if children were retained or not). Latinos were less likely to be retained compared to Blacks after failing the FCAT, even while controlling for ELL status, and poverty, suggesting a possible protective quality for Latinos. Being an ELL did not significantly predict retention after failing, further suggesting that there is no overt bias against non-English speakers for retention, and the protective quality that Latinos seem to have does not seem to be due to language, and may be explained by other factors. Children who received free/reduced lunch in third grade had 72% greater odds of being retained compared to those not receiving free/reduced lunch in third grade after failing the FCAT. Males had 37% greater odds of being retained compared to females. ELLs still not proficient in English in third grade had 47% greater odds of being retained compared to those proficient by then, and third graders in special education had 2.65 times the odds of failing compared to typically developing children.

When simply considering the demographic variables in step 1, we are perhaps painting an incomplete picture of what is really going on in the school system. Similar to the previous analysis, by adding third grade performance into the model in step 2, we can see if GPA/school performance more generally explains some of the relationship between the demographic variables and our retention outcome, and whether this mediates the effect of other variables.

When third grade performance was added to the model in step 2 (the two right most columns in Table 4), our prediction accuracy increased to 68.7%. This means that we can now accurately determine whether or not a child will be retained, based on this new model, including performance in third grade, 68.7% of the time.

Latinos now have 35.6% decreased odds of being retained if they failed, and Blacks had 36.9% decreased odds of being retained if they failed, compared to Whites. This suggests that most of the differences between Latinos and Blacks was due to third grade GPA/performance. Most notable here is that there does *not* seem to be a bias against minority groups in retention decisions once performance in third grade is appropriately controlled.

Still, those who receive free/reduced lunch in third grade had 32% greater odds of being retained, compared to those who didn't, even after controlling for third grade performance and ethnicity, suggesting a potential bias in the system in terms of retention decisions for children in poverty. Children who were not English proficient in third grade had 36% greater odds of being retained compared to those who are proficient, and those in special education in third grade had 2.73 times greater odds of being retained

compared to those who do not. Lastly, for every single point increase in GPA, children had 76.9% decreased odds of being retained.

Overall, when considering only those demographic variables in the first step, there are clear differences in the relationship between these demographic variables and retention decisions after failing the FCAT, however in order to get a more complete picture of how this policy is functioning in the school system, we added third grade performance to the model. Even after controlling for all the other demographic variables and third grade performance, poverty, lacking English proficiency in third grade, and special education status were all significantly related to being retained after failing the FCAT.

## **DISCUSSION**

During a time when high-stakes testing is at the forefront of the minds of policy makers, school administrators, and educational researchers, it is imperative that we take a closer look at the educational policies that are being put to practice in districts across the United States. The goals of this thesis were to (a) describe the proportion of ethnically and linguistically diverse children in an urban community in Miami, Florida who fail the FCAT reading test, and subsequently get retained, and (b) to determine the child-level predictors of these outcomes. Previous studies typically looked at data in a bivariate manner, without properly controlling for confounding variables. However, this study not only explained the relationships between demographic variables bivariately, but also in a logistic regression where we were able to control for confounding variables that might explain the likelihood of failing and being retained. Additionally, prior research in the area of elementary school retention has been based on retention policies that do not include data from high-stakes tests. This thesis allows us to look at the consequences of these high-stakes test policies at the child-level, especially years after it has been put in place, and during a current time when many states are reconsidering whether to continue their mandatory high-stakes test retention policies.

This study explores what happens when implementing a high-stakes retention policy in a particularly low-income and ethnically diverse area. By using such a diverse

and urban population, we are able to witness the effects of policies and the potential problems and biases that can arise from such policies. Importantly, since we are still seeing major differences that indicate biases within such a diverse urban community, one can only imagine the disparities this kind of policy can create in an area that is rural and potentially more segregated by income and ethnicity.

In our dataset in Miami, only about half of the students who failed the FCAT were retained in third grade. This means that the mandatory policy is not being enforced for every student who fails the FCAT. In the current paper, we were able to assess the proportion of students who appeared to fall into the exemption categories, and for the current sample, this totaled about 9% of the group of children who failed the FCAT reading test. According to Greene and Winters, 26.4% of children who failed the FCAT fell into one or more of the exemption categories, with about 8% of them having a disability and about 7% of their sample passed an alternative standardized test as their cause for exemption (Greene & Winters, 2009). However, it is important to note that our sample is unique to Miami-Dade County, and Greene and Winters used Florida statewide data. This distinction helps us understand how this policy is being implemented in Miami-Dade County, a uniquely ethnically and linguistically diverse, low-income group of children. By just exploring how policies affect states overall we might ignore crucial differences that counties face. This study allows us to explore the impacts of statewide policies, on a large proportion of an individual county.

### **Failing the FCAT Reading Test**

A number of interesting findings emerged pertaining to child ethnicity. We found that when strictly examining race bivariately, Blacks were more likely to fail the FCAT test compared to Latinos or Whites. This result mirrors other previous studies that simply look at the bivariate differences across ethnic groups (Frey, 2005; Greene & Winters, 2009; Meisels & Liaw, 1993). However, this thesis continued and explained this relationship in greater detail. While controlling for income, gender, third grade performance and ELL status, Blacks were *still* more likely to fail compared to Whites. This suggests that the FCAT may be biased against Black children and putting them at an even greater disadvantage compared to their White peers. Importantly, this difference *cannot* be explained by poverty and ELL status because those variables were controlled for in the analyses.

Latinos seem to be consistently less likely to fail the reading FCAT in third grade, compared to Blacks. Interestingly, these findings were still significant in the presence of controlling for English proficiency. In this dataset, we have the chance to tease apart the influence of ethnicity and the effects of English proficiency. It is possible that the community support helps Latinos in Miami succeed in school. Lastly, we found that Latinos (compared to Whites) were more likely to fail the FCAT reading test, but since this effect disappeared when GPA was added to the model, this suggests that Latinos were more likely to fail the FCAT because they were not performing as well in third grade in general.

Students who received free/reduced lunch in third grade were more likely to fail the FCAT, even accounting for third grade GPA. Importantly, it would appear that some

of the relationship between poverty and failing the FCAT is due to performance in school generally, as evidenced by the odds ratio decreasing when GPA was added to the model. Children in poverty seem to be doing poorer in school in general, and this contributes to their likelihood of failing the FCAT. In a recent study exploring the performance on the Nebraska high-stakes test, Beckman, Messersmith, Shepard, and Cates found that students who were receiving free/reduced lunch performed significantly behind their peers who were not receiving free/reduced lunch (Beckman et al., 2012), however they did not control for child- or family-level variables. Students who are in poverty typically have trouble advocating for resources they deserve (Lacour & Tissington, 2011). Children in poverty do not have access to adequate resources, and this plays a critical role in why they are underachieving in school (Lacour & Tissington, 2011). Our study found similar results where children in poverty were still performing behind their more affluent peers even after controlling for relevant demographic variables like gender and ethnicity. Notably, in our regression model, we were able to accurately predict failing the FCAT for 85.2% of third graders using demographics (including poverty status) alone. This value is far too high. It is important to stress that standardized tests are intended to level the playing field among children across school, within any given state. We should not be able to predict performance so accurately using just demographic information alone. This suggests an even larger problem pointing to the kinds of demographic variables that are determining performance on these high-stakes tests. The best predictor of performance on high-stakes tests appears to be SES, not the quality of teaching and learning that is going

on in the classroom (Baker & Johnston, 2010), so it does not appear that these should be used to measure teacher, school, and student performance.

Boys were also more likely to fail the FCAT compared to girls, but this effect is reversed when GPA in third grade was added to the model. This indicates that boys were more likely to fail the FCAT because of their poorer performance in school, generally. This mirrors the gap in standardized test performance among boys and girls (Cornwell, Mustard, & Van Parys, 2012). Cornwell and colleagues used a nationally representative sample and found that girls consistently outperformed boys on reading tests, regardless of ethnicity. It seems that a similar trend is happening in Miami.

Students who were considered ELLs were less likely to fail the FCAT even after controlling for general performance in third grade. This suggests that ELL students were demonstrating a slight advantage compared to those who were non-ELL. De Feyter and Winsler (2009), using a similar sample as this thesis, found that children from immigrant backgrounds had better school readiness skills especially in the initiative, self-control, attachment, and behavior, compared to non-immigrant children. Also, first-generation immigrant children who were Latino demonstrated higher social-emotional protective factors compared to second-generation children and non-immigrant children (De Feyter & Winsler, 2009). This suggests that Latino immigrants may have a slight advantage over non-immigrant children in Miami. This could explain the finding we see in third graders. It could be that ELL third graders in our sample are better at these social and emotional skills compared to non-ELL children and this is why they are less likely to fail the FCAT.

The difficulty emerges for ELLs when they are still not proficient in English by third grade. These students not proficient in English in third grade were more likely to fail the FCAT, regardless of general performance in third grade. Students who have still not mastered English in third grade have limited English language ability and this seems to affect performance on the FCAT reading test, since it is administered in English. Previous studies have found that the gap between ELLs and non-ELLs in standardized test performance widens with subjects that are more linguistically challenging, such as reading (Abedi, 2002). ELL students in a Nebraska sample were also perform significantly behind their English proficient peers (Beckman et al., 2012). This supports our findings that students who are behind in English proficiency are being pushed further behind their peers by high-stakes testing. It is worth mentioning that perhaps schools and testing companies should consider a Spanish version of standardized exams for areas that have a large population of Hispanics/Latinos. This would ensure a more culturally sensitive testing environment and perhaps increase the likelihood that more students could perform well. Research has shown that when schools support an ELL student's dominant language, they perform better in school (García & Jensen, 2013). Our findings should encourage schools to provide intensive English instruction for students early on so that by third grade (especially in cases where there are high-stakes tests in third grade), ELL students are more proficient in English.

Similarly, students who were in special education in third grade were more likely to fail the FCAT reading test, even when controlling for performance in third grade. This finding suggests that this test is further disadvantaging students with disabilities. Students

in special education, like students who are limited English proficient, perhaps should not be required to take the FCAT reading test or any high-stakes test, or they should receive special accommodations.

### **Retention in Third Grade**

While it seems that the trend we saw for failing the FCAT above might follow the same pattern with regard to retention, however, the results in this current paper are not the same. It seems that even though Blacks and Latinos are performing poorly on the FCAT reading test, they are *not* being differentially biased against in terms of being retained in third grade, compared to Whites. Similarly, Latinos are less likely to be retained compared to Blacks. When considering possible reasons for this trend, a number of explanations can be considered. In Miami, there is a unique ethnic make-up where Latinos and Blacks are the first and second majority, and Whites are the minority. This may provide positive community support that helps Latino and Black children move forward in school (Lee & Klugman, 2013). Other explanations for this trend may be outside the scope of this study. Unfortunately, our data do not allow us to see what goes on behind the scenes in the school and the families. We only receive information about what actually occurred in the school using administrative resources. We do not have any information about encouragement or discouragement that children may have received from the school or teachers or parents regarding retention decisions that might have influenced the outcomes. It could be that Whites encourage their children to repeat a grade if they are falling behind, and Black and Latinos do not want that for their own children (Winsler et al., 2012). White parents might want to avoid later problems, and

may want to hold their children back on year earlier in elementary school before the problems persist too long. Alternatively, it could be that the school system is purposely not retaining Blacks and Latinos as much as Whites because they do not want to be perceived as discriminatory.

Children who receive free/reduced lunch in third grade are more likely to be retained after failing the FCAT reading test, even after controlling for performance in third grade. It appears that some (but not all) of the relationship between free/reduced lunch and being retained, after failing the test, is also due to GPA/performance in third grade. It seems probable that children who are from lower resourced families are performing behind their more affluent peers in school generally and this contributes to the decision to be held back. Students who come from low-resourced backgrounds do not have access to adequate test preparation or external resources that may help them succeed (i.e. tutors, additional instruction outside of school). Additionally, there are many other factors that influence families from low-income backgrounds. There tends to be increased stress, poorer nutrition, worse sleep patterns, and increased chaos in families who are struggling financially (Baum, Garofalo, & Yali, 1999). This puts them at a significant disadvantage when taking standardized exams and learning in school more generally, and this needs to be considered when using standardized exams to make serious educational decisions for students from at-risk backgrounds. While there is no obvious solution to this problem, there are steps we can take to ensure that children in poverty are not being held back more than their more affluent peers. Providing extra support (extra reading instruction, tutoring, reading materials, etc.) for families who are from lower-income

backgrounds would be one method of ensuring children have more equal opportunities of success in school.

Males were more likely to be retained compared to females when considering just demographic variables, but when controlling for third grade performance, this effect was no longer significant. This is a similar trend as what we found with failing the FCAT suggesting that the reason why boys are retained more often is due to their generally poorer performance in third grade.. Generally, boys are performing behind girls across subject areas in elementary school (Cornwell et al., 2012). Our findings provide further support for this trend that is consistent across the nation. More research needs to be done to determine ways of improving the current gender gap.

Children who were considered ELLs by the district were not more likely to be retained in third grade due to high-stakes test performance. From a policy perspective, this finding reveals that the school district does not systematically retain ELL students, and in fact, being an ELL might serve as a protective factor. These results are similar to prior research using this same dataset. Winsler and colleagues demonstrated that ELL students were not being retained in kindergarten as often as non-ELL students, an identical finding to ours, only 4 years earlier in school (Winsler et al., 2012). From this study, it seems that there is an ELL advantage, especially in this Miami sample. Notably, since 90% of third-grader ELL students were considered proficient in English, our sample has a rather quick English proficiency rate with half of them reaching proficiency within 2 years (Kim et al., 2014).

When children are still not proficient in English by third grade, however, further issues emerge in regard to retention. Third graders in our sample who were not English proficient were more likely to be retained in third grade after failing the FCAT, even after controlling for performance in third grade. This finding suggests that there should be more strict exemption categories for children who are not English proficient. Perhaps these students should not be required to take the FCAT, or at the very least, not be held to the same accountability standard with the retention consequence.

Students who were in special education who failed the FCAT were more likely to be retained, even with controlling for GPA in third grade. This indicates that perhaps this high-stakes testing policy doesn't adequately make exceptions for students who have a disability, and are in fact demonstrating a bias against individuals in special education. It is widely accepted that students with special needs have exceptional difficulties with reading-related assessments, and even the latest Common Core State Standards recommend teachers provide additional resources for students in special education who are experiencing difficulties (Haager & Vaughn, 2013). This provides even further evidence that students with a disability should either not be taking high-stakes tests or should be exempt from the high-stakes retention policies.

### **Limitations**

There are a number of limitations in this study that need to be considered. First, since we are using secondary data, we have no information on the actual decision-making processes for any of the children in our sample. We only receive the final outcomes for the students, and our data reflect what actually happened. We know that the schools are

required to retained children who failed the FCAT reading test, and we know that schools did not strictly adhere to this policy, as demonstrated by only half the children who failed the FCAT reading test actually being retained. So we do not know if there were other factors that went into the decision-making process for the other half of students who failed and were promoted (like parents or teachers advocating strongly for individual children to be or not be retained). Likewise, qualitative data that would help support or describe our findings are not available but clearly needed in future research.

Second, we also do not have much information on the exemption categories. We are able to address parts of them and determine, to a small extent, the number of children who may be exempted from the retention policy. For example, we do not have information about the exact IEP that students who are in special education have, and therefore we cannot determine if they were excluded from the policy based on an actual diagnosis or based on a personal decision. However, these are simply speculations as we are not given any information from the school about the formal exemptions that their students were given.

Third, the sample used in this paper was not statewide or fully countywide in its representation. We examined a sample of children who were in the MSRP (which is not the entire Miami-Dade County), and while our sample is quite large, it does not completely encompass the entirety of the county. However, this has its advantages and disadvantages. While we cannot comment on how the state of Florida is dealing with their high-stakes test policy, we can describe how this policy is being put to practice within a very low-income and diverse area within Florida. It is still important to measure

statewide policies at the county-levels in order to determine if there is regional variation in the implementation of these policies. As an aside, it is a particular strength of this study to explore how a statewide high-stakes test policy gets put into practice within an extremely diverse and low-income at-risk sample. If we are seeing differences among students who are primarily low-income and ethnic minorities within our MSRP sample, we can only imagine how difficult it must be for low-income and minority students in other areas around the country that have high-stakes test policies.

### **Policy Implications**

This project informs the policy sphere of how Florida's statewide high-stakes test policy is being put to practice in Miami-Dade County. In a post-NCLB world, we must be sensitive, conscientious, and educated about the potential negative consequences certain educational policies can have, across the entire nation, and especially among various vulnerable ethnic and socioeconomic groups.

This study has clear implications for the policy in Miami, Florida, and extensions to other high-stakes policies in the United States. Importantly, we could accurately categorize children by who failed the FCAT and who did not simply using demographic factors fully 85% of the time. This is striking evidence that this test may be measuring something other than simply reading abilities and teaching skill. It seems that this test may be pushing the divide even further among children in different ethnic and socioeconomic groups.

The key conclusion from this study is that even after controlling for performance in school, being from a low-income family puts children at greater risk for failing the

FCAT and being retained, compared to their more advantaged peers. Therefore, there is a potential bias that exists in this policy that negatively affects those in poverty. By using the FCAT for making retention decisions, the existing achievement gaps between income-level groups is only expanding. Therefore, based on the findings from this study, we would advise that this exam be removed as the main method of determining the retention and promotion of students. By further disadvantaging certain students who are taking this exam, the FCAT is directly contradicting the initial goals of the NCLB act. This high-stakes test policy in Florida is producing issues related to discrimination and further disadvantaging already at-risk students which runs counter to the efforts put forth by NCLB. By implementing policies that increase the divide among economic and racial groups, how can we say that we are “closing the achievement gap”? Instead, mandatory retention policies on the basis of high-stakes testing appear to be raising serious civil rights and social justice issues that need to be carefully addressed.

**Table 1 Bivariate Correlates of Failing FCAT in Third Grade**

Variable (N)	Failed FCAT (%)
Third Graders with FCAT scores (27,703)	15 (n = 4,284)
Ethnicity***	
White/other (2,029)	6.4 (n = 129)
Black (9,291)	22.9 (n = 2,125)
Latino/Hispanic (16,373)	12.4 (n = 2,024)
Gender***	
Male (13,800)	17.7 (n = 2,447)
Female (13,897)	13.2 (n = 1,833)
English use at Home***	
Yes (15,885)	13.5 (n = 2,156)
No (11,816)	18 (n = 2,127)
English Proficiency***	
Not proficient (6,409)	22.1 (n = 1,419)
Proficient (21,294)	13.5 (n = 2,865)
Special Education***	
Yes (2,778)	41.7 (n = 1,158)
No (24,575)	12.3 (n = 3,021)
Free/reduced lunch status***	
Free/reduced (22,040)	17.8 (n = 3,921)
Not (5,663)	6.4 (n = 363)

*Note.* These are row percentages (i.e. percent of white students who failed the FCAT reading test)

\*\*\* $p < .001$

**Table 2 Bivariate Correlates of Being Retained after Failing FCAT in Third Grade**

Variable (N)	Retained in Third Grade (%)
Third Graders around the Next Year (25,739)	52.6 ( $n = 1,999$ )
Ethnicity	
White/other (111)	55.9 ( $n = 62$ )
Black (1,840)	52.2 ( $n = 960$ )
Latino/Hispanic (1,843)	53 ( $n = 976$ )
Gender***	
Male (2,171)	57.5 ( $n = 1,248$ )
Female (1,624)	46.2 ( $n = 750$ )
English Use at Home	
Yes (1,964)	53.9 ( $n = 1,058$ )
No (1,833)	51.3 ( $n = 941$ )
English Proficiency in Third Grade***	
Not proficient (1,290)	57.3 ( $n = 739$ )
Proficient (2,508)	50.2 ( $n = 1,260$ )
Special Education***	
Yes (1,020)	69.8 ( $n = 712$ )
No (2,778)	46.3 ( $n = 1,287$ )



**Table 4 Logistic Regression Predicting Retention (Among Students who Failed Reading FCAT)**

Variable	Step 1		Step 2	
	Odds ratio	SE(B)	Odds ratio	SE(B)
Latino/White	0.76	0.212	.644*	0.223
Black/White	0.925	0.21	.631*	0.222
Latino/Black	0.822*	0.098	1.02	0.105
Free/Reduced Lunch	1.717***	0.127	1.32*	0.135
Male	1.37***	0.069	1.11	0.074
ELL ever	0.913	0.114	0.969	0.122
Not Eng. Proficient	1.47***	0.099	1.36**	0.105
Disability Status	2.65***	0.081	2.73***	0.087
GPA in 3rd Grade			.231***	0.072

\* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$

## REFERENCES

- Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometrics issues. *Educational Assessment, 8*(2), 231-257.
- Abella, J., Urrita, J., & Schneiderman, A. (2005). An examination of the validity of English-language achievement test scores in an English language learner population. *Bilingual Research Journal, 29*, 127-144.  
doi:10.1080/15235882.2005.10162827
- Alexander, K. L., Entwisle, D. R., & Dauber, S. L. (1994). *On the success of failure: A reassessment of the effects of retention in the primary grades*. New York, NY: Oxford University Press.
- American Educational Research Association (AERA). (2000). *AERA position statement on high-stakes testing in pre-K – 12 education*. Washington, DC: Author.  
Retrieved from  
<http://www.aera.net/AboutAERA/AERARulesPolicies/AERAPolicyStatements/PositionStatementonHighStakesTesting/tabid/11083/Default.aspx>
- Baker, M., & Johnston, P. (2010). The impact of socioeconomic status on high stakes testing reexamined. *Journal of Instructional Psychology, 37*(3), 193-199.
- Baum, A., Garofalo, J. P., Yali, A. M. (1999). Socioeconomic status and chronic stress: Does stress account for SES effects on health? *Annals New York Academy of Sciences, 131*-144.
- Beckman, T. O., Messersmith, K. Sheperd, J., & Cates, B. (2012). Ethnicity, language and poverty predicting scores on Nebraska state accountability reading test. *International Journal of Psychology: A Biopsychosocial Approach, 11*, 31-47.
- Cornwell, C., Mustard, D. B., & Van Parys, J. (2012). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *The Journal of Human Resources, 48*(1), 236-264.
- De Feyter, J. J., & Winsler, A. (2009). The early developmental competencies and school readiness of low-income, immigrant children: Influences of generation,

- race/ethnicity, and national origins. *Early Childhood Research Quarterly*, 24(4), 411-431. doi: 10.1016/j.ecresq.2009.07.004
- Duffy, M., Giordano, V. A., Farrell, J. B., Paneque, O. M., & Crump, G. B. (2009). No Child Left Behind: Values and research issues in high-stakes assessments. *Counseling and Values*, 53(1), 53–66. doi: 10.1002/j.2161-007X.2009.tb00113.x
- Fantz, A. (2015). *Prison time for some Atlanta school educators in cheating scandal*. Retrieved from <http://www.cnn.com/2015/04/14/us/georgia-atlanta-public-schools-cheating-scandal-verdicts/>
- Florida Department of Education. (2004). *Assessment & accountability briefing book*. Tallahassee, FL: Author.
- Florida Department of Education. (2002). Technical report: For operational test administrations of the 2000 Florida Comprehensive Assessment Test. Tallahassee, FL: Author.
- Florida Department of Education. (2011). *Archives*. Retrieved from <http://www.fldoe.org/eias/eiaspubs/archives.asp>
- Florida Department of Education. (2013). *Public school student progression; remedial instruction; reporting requirements*. Retrieved from [http://leg.state.fl.us/Statutes/index.cfm?App\\_mode=Display\\_Statute&Search\\_String=&URL=1000-1099/1008/Sections/1008.25.html](http://leg.state.fl.us/Statutes/index.cfm?App_mode=Display_Statute&Search_String=&URL=1000-1099/1008/Sections/1008.25.html)
- Florida Comprehensive Assessment Test (FCAT). (2011). *Achievement levels*. Retrieved from <http://fcats.fldoe.org/pdf/fcAchievementLevels.pdf>
- Frey, N. (2005). Retention, social promotion, and academic redshirting: What do we know and need to know? *Remedial and Special Education*, 26(6), 332–346. doi: 10.1177/07419325050260060401
- Garcia, E., & Jensen, B. (2009). Early educational opportunities for children of Hispanic origins. *Social Policy Report*, 23(2), 3-20.
- Greene, J. P., & Winters, M. A. (2007). Revisiting grade retention: An evaluation of Florida's test-based promotion policy. *Education Finance and Policy*, 2(4), 319–340. doi: 10.1162/edfp.2007.2.4.319
- Greene, J. P., & Winters, M. A. (2009). The effects of exemptions to Florida's test-based promotion policy: Who is retained? *Economics of Education Review*, 28(1), 135–142. doi:10.1016/j.econedurev.2008.02.002

- Haager, D., & Vaughn, S. (2013). The common core state standards and reading: Interpretations and implications for elementary students with learning disabilities. *Learning Disabilities Research and Practice, 28*(1), 5-16.
- Hedges, L. V., & Nowell, A. (1998). Black-White test score convergence since 1965. In C. Jenks & M. Phillips (Eds.), *The Black-White test score gap* (pp. 149-181). Washington, DC: Brookings Institution.
- Heubert, J. P., & Hauser, R. M. (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, DC: National Academy Press.
- Horn, C. (2003). High-stakes testing and students: Stopping or perpetuating a cycle of failure? *Theory Into Practice, 42*(1), 30–41. doi:10.1207/s15430421tip4201\_5
- Johnson, D. R., Thurlow, M. L., Stout, K. E., & Mavis, A. (2007). Cross-state study of high-stakes testing practices and diploma options. *Journal of Special Education Leadership, 20*(2), 53-65.
- Kim, Y. K., Curby, T. W., & Winsler, A. (2014). Child, family, and school characteristics related to English proficiency development among low-income, dual language learners. *Developmental Psychology, 50*(12), 2600-2613. doi: 10.1037/a0038050
- Lacour, M., & Tissington, L. D. (2011). The effects of poverty on academic achievement. *Education Research and Reviews, 6*(7), 522-527.
- Lee, J. C., & Klugman, J. (2013). Latino school concentration and academic performance among Latino children. *Social Science Quarterly, 94*(4), 977-1015.
- McCoy, A. R., & Reynolds, A. J. (1999). Grade retention and school performance: An extended investigation. *Journal of School Psychology, 37*(3), 273-298. doi: [http://dx.doi.org/10.1016/S0022-4405\(99\)00012-6](http://dx.doi.org/10.1016/S0022-4405(99)00012-6)
- Meisels, S. J., & Liaw, F. R. (1993). Failure in grade: Do retained students catch up? *The Journal of Educational Research, 87*(2), 69–77. doi:10.1080/00220671.1993.9941169
- Miami-Dade County Public Schools. (2014). Retrieved from [http://www.dadeschools.net/StudentEnroll/Calendars/enroll\\_stats\\_aor.asp](http://www.dadeschools.net/StudentEnroll/Calendars/enroll_stats_aor.asp)
- National Center for Education Statistics. (2009). The condition of education. (NCES 2009-081). Retrieved from <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009081>
- Penfield, R. D. (2010). Test-based grade retention: Does it stand up to professional

- standards for fair and appropriate test use? *Educational Researcher*, 39(2), 110–119. doi:10.3102/0013189X10363007
- Plake, B. S. (2002). Evaluating the technical quality of educational tests used for high-stakes decisions. *Measurement and Evaluation in Counseling and Development*, 35(3), 144-152. doi: 10.3102/0013189X10363007
- Reynolds, A. J. (1992). Grade retention and school adjustment: An explanatory analysis. *Educational Evaluation and Policy Analysis*, 14(2), 101-21. doi: 10.3102/01623737014002101
- Roderick, M., & Nagaoka, J. (2005). Retention under Chicago's high-stakes testing program: Helpful, harmful, or harmless? *Educational Evaluation and Policy Analysis*, 27(4), 309-340. doi: 10.3102/01623737027004309
- Siegel, H. (2004). High stakes testing, educational aims and ideals, and responsible assessment. *Theory and Research in Education*, 2(3), 219–233. doi:10.1177/1477878504046515
- Stanford Achievement Test Series, Tenth Edition (SAT-10). (2011). Guides teaching and learning toward high achievement standards. Retrieved from <http://www.pearsonassessments.com/learningassessments/products/100000415/stanford-achievement-test-series-tenth-edition.html>
- Stewart, P. (2011). *Florida Department of Education: Read to learn*. Retrieved from <http://info.fl DOE.org/docushare/dsweb/Get/Document-6738/dps-2013-56b.pdf>
- Texas Education Agency. (2013). Student Success Initiative manual: Grade advancement requirements. Retrieved from <http://www.tea.state.tx.us/student.assessment/ssi/>
- United Opt Out. (2014). *United opt out: The movement to end corporate education reform*. Retrieved from <http://unitedoptout.com/>
- US Census. (2013). *State and county quick facts: Miami-Dade county, Florida*. Retrieved from <http://quickfacts.census.gov/qfd/states/12/12086.html>
- Winsler, A., Hutchison, L., De Feyter, J. J., Manfra, L., Bleiker, C., Hartman, S., & Levitt, J. (2012). Child, family, and childcare predictors of delayed school entry and kindergarten retention among linguistically- and ethnically-diverse children. *Developmental Psychology*, 48 (5), 1299-1314. doi: 10.1037/a0026985
- Winsler, A., Karkhanis, D. G., Kim, Y. K., & Levitt, J. (2013). Being black, male, and gifted in Miami: Prevalence and predictors in elementary school gifted education programs. *Urban Review*, 45, 416-447. doi: 10.1007/s11256-013-0259-0

Winsler, A., Kim, Y. K., & Richard, E. R. (2014). Socio-emotional skills, behavior problems, and Spanish competence predict the acquisition of English among English language learners in poverty. *Developmental Psychology*. doi: 10.1037/a0037161

Winsler, A., Tran, H., Hartman, S. C., Madigan, A. L., Manfra, L., & Bleiker, C. (2008). School readiness gains made by ethnically diverse children in poverty attending center-based childcare and public school pre-kindergarten programs. *Early Childhood Research Quarterly*, 23(3), 314–329. doi:10.1016/j.ecresq.2008.02.003

## **BIOGRAPHY**

Tanya Tavassolie received her Bachelor of Arts degree from Franklin and Marshall College in 2011 in Biological Foundations of Behavior: Neuroscience. She was employed as a research assistant at University of Maryland, Child Development Lab for two years. She is currently a graduate student working under Dr. Adam Winsler at George Mason University. This thesis is the first of many papers that she hopes to publish during her time at George Mason University.