| | |
|---|---|
| **From:** | Patricia Levesque (patricia@afloridapromise.org) <patricia@afloridapromise.org> |
| **Sent time:** | Friday, September 09, 2011 7:01:56 PM |
| **To:** | Padget, John; Desai, Dr. Akshayi; Martinez, Roberto; Feingold, Barbara; Chartrand, Gary; Shanahan, Kathleen; Bradshaw, Sally |
| **Cc:** | Mary Laura Bragg (MaryLaura@excelined.org) <MaryLaura@excelined.org>; Joanna Hassell (Joanna@afloridapromise.org) <Joanna@afloridapromise.org>; Robinson, Gerard; Abbott, Lynn |
| **Subject:** | FW: new research on florida's reading promotion policies |
| **Attachments:** | The Medium Run Effects of Retention.docx |

---

Dear members of the State Board of Ed,

I love when we can share good news with you. The Manhattan Institute is following up on the research that they did several years ago on Florida's 3rd grade reading promotion policies. Their prior research tracked the first cohort of retained students two years after their retention (to 5th grade). That research found that the retained students did significantly better than similar students who had been socially promoted and the gains grew in the 2nd year.

Their newest research tracks the 1st two cohorts of retained students through middle school. They found that the gains of the retained students over the promoted students still exists in middle school (even though the effects decrease slightly) and the benefits of the reading policy can be seen in the math and science gains as well.

Dr. Winters is trying to get his research reviewed in an academic journal so it may be published in the next year. Just wanted you to see first.

Patricia

p.s. This research certainly supports the state board's recent position to continue to emphasize reading in your budget recommendations with your $40 million requested increase for reading coaches.

-----Original Message-----

From: Winters, Marcus [mailto:mwinters@Manhattan-Institute.org]

Sent: Friday, September 09, 2011 2:46 PM

To: Patricia Levesque (patricia@excelined.org)

Cc: Mary Laura Bragg (MaryLaura@excelined.org)

Subject: RE:

Attached is the version of the retention paper that we submitted to a journal. Let me know if have any questions or comments.

marcus

# The Medium-Run Effects of Retention: Evidence from Florida's Test-Based Promotion Policy

Marcus A. Winters, Ph.D.*
Assistant Professor
Department of Leadership, Research and Foundations
College of Education
University of Colorado at Colorado Springs

Jay P. Greene, Ph.D.
Endowed Professor
Department of Education Reform
College of Education and Health Professions
University of Arkansas

*Corresponding author

Marcus A. Winters
Columbine Hall 3049
University of Colorado at Colorado Springs
1420 Austin Bluffs Pkwy
Colorado Springs, CO 80918
mwinters@uccs.edu
Phone: 719-255-3614

Jay P. Greene
201 Graduate Education Building
University of Arkansas
Fayetteville, AR 72701

Acknowledgements: We would like to thank the Florida K-20 Data Warehouse for providing the data necessary for the analysis. We'd also like to Martin West for valuable comments.

Abstract:

We use a regression discontinuity strategy to produce causal estimates for the effect of being retained under Florida's test-based promotion policy on multiple achievement outcomes for up to five years after the retention intervention. Retention has a statistically significant and substantial positive effect on student achievement in math, reading, and science in the years immediately following the treatment. But the effect of retention dissipates over time. Nonetheless, we find that the effect of retention on academic achievement is statistically significant and of a meaningful magnitude several years after the intervention.

Keywords: Social Promotion, Retention, Accountability, Standards, High-Stakes Testing, Regression Discontinuity

JEL: I20, I28

**Introduction**

Among the policies on top of the standards-based reform agenda are those that curtail "social promotion" -- the longstanding practice of promoting students to the next grade level for socialization reasons even if they have not demonstrated an adequate level of proficiency for academic promotion. Several states and school districts now employ policies requiring students in particular grades to demonstrate possession of some minimal skill on one or more standardized assessments before they are promoted to the next grade level. Most notably, such test-based promotion policies are operating in the Florida, Texas, New York City, and Chicago public school systems. In addition, Oklahoma, Arizona, and Indiana adopted similar programs in the last legislative session, and other states are reported to be seriously considering such policies for the next session.

Test-based promotion policies can substantially increase the percentage of students who are retained in grade. For instance, the percentage of third graders retained in Florida increased to about 12 percent in the first two years of the state's adoption of the policy from only about 3 percent the previous two years (Greene and Winters 2009).

School systems adopting policies that dramatically increase grade retention do so despite a large body of research finding that retention is harmful to student achievement (for literature reviews expressing this view of the research see Holmes 1989; and Jimerson 2001). However, researchers have recently raised serious doubt about the believability of that prior research (for recent examples example see Greene and Winters 2007; Allen et. al 2009; and Hughes, Chen, Thoemmes, and Kwok 2010).

Estimating the effect of retention on student achievement is difficult because students are not retained randomly. Making matters worse, when determining whether to hold a child back in the grade a teacher will often consider factors that are unobserved by the researcher, such as the child's maturity level. Estimation will be biased if the

researcher does not account for such unobserved differences between retained and promoted students that are also related to later student achievement.

Few prior studies of grade retention have utilized a meaningful comparison group or adequate statistical controls necessary for making causal claims. Of the twenty-two papers evaluating the effect of grade retention on achievement from 1990 through 2006 that were identified in a recent meta-analysis by Allen et. al (2009), only six could be defined as “high quality”, meaning that they included comparison groups with similar observed characteristics at baseline and adequate statistical controls. The meta-analysis discovered that studies of higher quality report more positive effects from grade retention than do lower quality evaluations. Nonetheless, even these higher quality papers do not tend to find that retention leads to substantial academic improvements.

Some recent work has dramatically improved upon prior research on the effects of retention by utilizing a regression discontinuity identification strategy (Greene and Winters 2007; Jacobs and Lefgren 2004; Jacob and Lefgren 2007; Roderick and Nagaoka 2005). Such papers deserve particular attention because, unlike even very sophisticated matching strategies also characterized as "high quality" designs by Allen et. al, under certain conditions regression discontinuity accounts for both observed *and unobserved* characteristics related to both the likelihood that a student is retained and his later academic achievement (Van der Klaauw 2002; Imbens and Lemieux 2008).

In addition to concerns about its quality, much of the prior research on the impact of retention is limited by the short-run nature of its findings. Prior studies of retention tend to follow students for only one or two years after the retention year. But it is the longer term impact of these programs that is of the greatest policy relevance. We might worry that any positive effect from retention might fade over time as students grow older. On the other hand, it is possible that the benefits from retention could actually grow over

time as the failure to acquire adequate skills places a student on a downward spiral relative to her peers.

Several papers have attempted to measure the extent to which the short-run impact of retention on student achievement is sustained in later years. But only two of the six studies characterized by Allen et. al as employing a high quality design evaluated student performance more than two years after the retention decision. Further, both of those studies (Rust and Wallace 1993 and Jimerson et. al 1997) utilized a matching design, which relies entirely on a set of observed characteristics and thus may not account for unobserved student factors related to retention. Though too recent to have been included in the recent meta-analysis, Hughes, Chen, Thoemmes, and Kowk (2010) use a propensity score matching technique in order to look at the effect of retention in the first grade on student achievement four years later. The findings from these recent evaluations using matching techniques for identification suggest that retention has a positive effect in the short run that declines over time, often to the point of statistical insignificance.

The current paper significantly extends the literature on the sustained effects of grade retention by expanding upon previous research evaluating Florida's test-based promotion policy (Winters and Greene 2007). We utilize a regression discontinuity design in order to provide causal estimates of the relationship between a student having been retained in the third grade due to the state's policy and achievement through the seventh grade. We map the academic performance of multiple student cohorts on both high- and low-stakes reading and math exams. We also provide estimates of the effect of retention on student achievement on an elementary science exam.

We find that students retained under Florida's policy made substantial academic gains in all subjects immediately following retention, and that these gains declined as the

students progressed through school. However, we find a statistically significant and meaningful difference in student proficiency several years after the retention year.

We refer to our results as "medium-run" effects of grade retention because Florida's policy has not been in existence for enough years to allow for an estimate of its effect on arguably the most important long-term academic outcome: the likelihood that a student will graduate from high school. Though no work that we know of has yet evaluated the effect of retention in elementary grades on high school graduation, using a regression discontinuity design Jacob and Lefgren (2007) find that retention in the sixth grade is not related to graduation, while retention in the eighth grade leads to a higher probability of dropping out. Nor are we able to estimate the effect on other long-term life outcomes such as market wages. A recent study by Babcock and Bedard (2011) finds evidence that an increase in early grade retention within a state is related to increases in mean male hourly wages. Nonetheless, with many thousands of students currently subjected to such policies and other states considering adopting similar programs, a consideration of the impact of Florida's program thus far is of substantial policy interest.

**Data**

We utilize a rich dataset made available from the Florida Department of Education's K-20 Data Warehouse. The dataset contains test score and demographic information for the universe of test-taking students in Florida public schools in grades three through eight from 2002-03 through 2008-09. The dataset also includes a unique student identifier that allows us to track individual student performance over time.

We impose no restrictions on the data in addition to those used to develop the sample for the regression discontinuity analysis. The sample restrictions and descriptive statistics for the treatment and comparison groups are provided in a later section.

## Florida's Test-Based Promotion Policy

Florida's policy to end social promotion was among a series of reforms adopted under the governorship of Jeb Bush. Students who entered the third grade in the fall of 2002 were the first subjected to the mandate. The law has applied to all subsequent cohorts of third grade students in the state.

Florida's policy requires that third grade students earn scores at or above Level 2 (the second lowest of five levels) on the state's high-stakes reading exam, the Florida Comprehensive Assessment Test (FCAT), in order to be default promoted to the next grade level. The benchmark score necessary to reach Level 2 has remained consistent over time. Performing below the test score threshold can only be said to influence "default promotion" because students can receive one of several exemptions and be promoted despite their low performance. In fact, nearly half of the students with test scores below the threshold in the policy's first year were promoted (Greene and Winters 2009).

Students retained according to the policy are also subjected to additional interventions during the retained year. They are required to attend a summer reading camp. Also, the schools are required to assign them to a "high quality teacher" as determined by performance data and above-satisfactory performance appraisals. Schools must also develop academic improvement plans for these students that address their specific needs during the retention year. Unfortunately, our analysis is not able to disentangle the effects of retention from that of these other interventions, and thus our estimates should be considered that of an average treatment effect of these interventions.

## Student Cohorts and the Special Case of Cohort 1

The fact that Florida's policy has been in effect for seven years allows us to evaluate the influence of the policy for multiple cohorts and over a sustained period of time. We estimate the effect of retention on these cohorts both individually and as a group.

Table 1 tracks the movement of the cohorts under consideration through grade levels over time. "T" indicates the group of students who were retained (treated) and C indicates the group of students who were promoted at the end of the third grade (control). The number next to the letter indicates the student's cohort. For instance, Cohort 1 first entered the third grade in 2002-03, and Cohort 2 first entered the third grade in 2003-04.

[TABLE 1 ABOUT HERE]

As the table shows, the furthest our dataset allows us to follow a cohort – Cohort 1 – is the eighth grade, because it is the last grade in which both retained and promoted groups are observed. The youngest cohort evaluated is Cohort 5, which is the last that our dataset allows us to observe fourth grade achievement for both the promoted and the retained group.

Cohort 1, the first subjected to the retention policy, represents a special case. As Table 1 shows, all cohorts subsequent to Cohort 1 share a grade level with a group of students in another cohort. For instance, the table shows that when retained students in Cohort 3 (T3) are in the fourth grade they share classrooms with promoted students from Cohort 4 (C4). Because Cohort 1 was the first to be subjected to the policy, students from that group who were promoted to the fourth grade do not share classrooms with a large group of students who were retained in the third grade from a previous cohort. An implication of this phenomenon is that the quality of peers sharing classrooms with promoted students from Cohort 1 in subsequent years is different than it is for students who were retained from that group.

Essentially, students who were promoted to the fourth grade at the end of 2002-03 no longer shared classrooms with a large number of very low performing students who were instead retained in the third grade. About 14 percent of this third grade class was retained in the third grade. Subsequent cohorts subject to the retention policy also were removed from many of their low-performing classmates, but these students were replaced in the cohort by students who were retained at the end of the previous year.

Promoted students in Cohort 1 (C1) – whether in our sample or not – attend later grades with a higher average quality peers than do students in subsequent cohorts. Prior research suggests that peer quality influences student learning gains during the school year (see for instance Hanushek, Kain, Markman, and Rivkin 2003). Since the promoted group from Cohort 1 has higher quality peers on average in later grades than does the retained group from Cohort 1, the estimated influence of retention on test scores at the end of later grades are likely biased downward.

Though perhaps not generalizable to the overall effect of retention policies, estimates of the effect of retention within Cohort 1 are relevant for policy. The experience of the first cohort subject to a test-based retention policy similar to Florida's that is adopted by another school system would have an experience similar to our Cohort 1. However, all subsequent cohorts in Florida (and anywhere a similar policy is adopted) might be expected to have a different experience than the first cohort subject to the policy.

We address the special case of Cohort 1 in multiple ways. First, in analyses that include multiple cohorts, we report the results of models that include and exclude students from Cohort 1.[1] Also, we estimate the effect of retention on each individual cohort over

[1] Notice that exclusion of Cohort 1 students during estimation does not remove their influence on peer quality for students in Cohort 2 because the students still did in fact share classrooms and thus played a role in the education production process determining Cohort 2's achievement that is under consideration.

time, and keep in mind the differing peer effects when comparing estimates resulting from Cohort 1 to those from other cohorts.

**Identification Strategy**

We employ a regression discontinuity identification strategy in order to provide causal estimates of the effect of retention under Florida's policy on later student achievement. Regression discontinuity is applicable in cases where assignment of a treatment is either entirely or partially a function of whether an individual falls above or below a continuously measured benchmark. When certain assumptions are satisfied, the procedure closely approximates a randomized experiment (Van der Klaauw 2002; Imbens and Lemieux 2008).

We take advantage of Florida's use of a discrete benchmark on the third grade reading test to help determine whether a child is promoted. Though their reading proficiencies are essentially equivalent, students with reading scores that fell just below the threshold for promotion were far more likely to be retained under Florida's policy than were students whose scores were just above the threshold.

Our analyses only include observations for students whose test score in the third grade fell within a small neighborhood around the passing cutoff -- a score of 1046 on the test's developmental scale. We choose as the neighborhood scores that were within 18 points below the cutoff or 23 points above the cutoff. We choose these points on the distribution because it is recommended that the analysis include no fewer than four points on the distribution on either side of the cutoff and, since the scale on the third grade assessment only allows for certain values, these are the four nearest observation points on either side of the cutoff. We choose this most narrow band possible because the smaller neighborhood improves the likelihood that the above and below groups are equivalent on

both observed and unobserved characteristics, and also because our statewide dataset allows for a large number of student observations within even this very narrow bandwidth.

Table 2 reports the number and percentage of students earning each observed initial third grade test score who were retained or promoted the following year. The table makes apparent that for each cohort a student with a score below the threshold for default promotion (1046 on the FCAT developmental scale) was far more likely to be retained than were students whose scores fell just above this cutoff.

[TABLE 2 ABOUT HERE]

The final column of Table 2 also shows all third grade students who were retained or promoted in particular years for whom we have test score data. It is clear that the percentage of third graders being retained has decreased substantially over the life of the policy. Much of that decline has to do with a general increase in student performance on the reading test across the state. However, it is worth noting that such general academic improvements in the state do not hinder our estimation. Though the distribution of third grade student test scores has shifted, there are still many students with academic achievement very near the threshold for promotion. Further, the table also shows that the percentage of students who are retained from our restricted group of students just above and below the threshold has remained relatively consistent, particularly for cohorts 2, 3, and 4.

That students and schools know about the policy and the exact cutoff score for default promotion is worrisome because the regression discontinuity procedure will not produce causal estimates if the subjects manipulate whether they fall just above or below the threshold (Schochet, et. al 2010). However, though we expect that both students and schools likely respond to the policy by attempting to push past the threshold for

promotion, we argue that this is not a particular problem for our estimation. There is no reason to believe that the student's standardized reading score is not a valid and reliable measure of the student's true proficiency level. Further, as seen by the table, we do not observe a heavy clustering of students just above the threshold for promotion, which would be expected in the case where the forcing variable has been manipulated.

Table 2 also shows that there are many students with scores below the threshold who were promoted, and some students with scores above the threshold who were nonetheless retained. Because retention is not strictly determined by where the student's score falls relative to the threshold for default promotion we utilize the so-called "fuzzy" regression discontinuity strategy. Essentially, this strategy treats an indicator for whether the student's score is below the threshold as an instrumental variable for retention in a two-stage least squares approach.

There are two stages in the estimation. The first stage utilizes observed characteristics about the student, including her test score and an indicator for whether her score falls above or below the threshold for retention, in order to predict the likelihood that she is retained in the third grade. Only observations of students in the third grade whose score falls within the previously defined neighborhood of the cutoff score are utilized in this first stage regression. Formally:

(1) $retained_{is3} = \beta_0 + \beta_1 X_{is3} + \beta_2 Below_{is3} + \gamma_s + \epsilon_{is3}$

Where $retained_{is}$ equals one if student i enrolled in school s was retained at the end of his third grade year[2]; X is a series of observed characteristics about the student including race/ethnicity, gender, an indicator for whether he is eligible for free or reduced price lunch, an indicator for whether he has been identified as having a disability, and his scores on the third grade high- and low-stakes reading and math tests (that is, four tests in

[2] A student is classified as having been retained if he is again observed in the third grade in the next year.

total), including the FCAT reading score that is linked to the retention policy (the forcing variable); Below, the instrumental variable, is an indicator that equals one if the student's FCAT reading score in the third grade was below the threshold for default promotion set by the policy and zero if the score was above the threshold; γ is a fixed-effect for the student's school; ε is a stochastic term clustered by school; and $\beta_0$ through $\beta_2$ are parameters to be estimated.

Equation (1) is estimated via OLS, which results in a linear probability model. Though "retained" is a dichotomous dependent variable, we utilize the LPM in this case rather than a Probit model because inclusion of more than a thousand school fixed effects is difficult with Probit, which is estimated via maximum likelihood.

We estimate (1) independently for students in each third grade cohort from 2002-03 through 2006-07. The model only includes students whose third grade test score was within the previously defined small neighborhood above or below the cutoff for default promotion.

The coefficient estimates from (1) are used as an estimate of the probability the child was retained conditional on observed characteristics, which we write as $\widehat{ret}$ and capture for each student. We then utilize $\widehat{ret}$ in a second-stage OLS regression evaluating the student's test score in a subsequent year. Formally:

(2) $Y_{isg} = \propto_0 + \propto_1 X_{isg} + \propto_2 Y_{is3} + \propto_3 \widehat{ret}_i + \delta_s + \mu_{isg}$

Where Y is the student's test score; δ represents a school fixed effect; μ is a stochastic term clustered by school; and $\alpha_0$ through $\alpha_3$ are parameters to be estimated. If we believe the identification assumptions (tested in the next section) then the estimate of $\alpha_3$ can be interpreted as the causal influence of retention on student academic achievement. Notice that since the probability that the child was retained in the third grade does not change

over time, $\widehat{ret}$ is the same value for the student regardless of the grade level test score under consideration.

The index g represents the student's grade in the year under consideration. Note that the student's test score that is used as a regressor on the right hand side of (2) always represents the student's test score in the initial third grade year.[3] We specify the student's prior achievement in this way for two reasons. First, the regression discontinuity procedure requires that the regression model controls for the point system used to make the student subject to treatment, and this "forcing variable" in our case is the student's score on the third-grade FCAT reading exam (Schochet et. al 2010). Secondly, we are interested in evaluating the effect of third grade retention on student achievement in grades several years following the retention year. If we estimated a common value-added model in which we accounted for the student's test score in the prior year, the results would be difficult to interpret because the student's previous test score is also partly determined by the treatment. For example, when evaluating the effect of third grade retention on the student's test scores in the sixth grade, we would not want to account for the student's fifth grade test score because it was also influenced by the student's retention in (or promotion from) the third grade. Thus, our procedure is to identify treatment and control groups who had very similar test scores at the end of their shared third grade year and then evaluate whether and to what extent the level of their test score in subsequent grades differs.

Equation (2) is estimated for students at the same grade level, not the same year. The performance of retained students is compared to that of other students with whom they attended the third grade for the first time. Since students in the treatment group were retained in the third grade, most of them are a grade level behind their original third grade

[3] Though retained students are in the third grade twice, we utilize as the control in (2) their initial score in that grade, which contributed to their retention.

classmates in each subsequent year. There is somewhat of a debate in the prior research on grade retention over whether comparisons between retained and promoted students should be made within-grade or within-year. We follow the within-grade approach and compare the performance of retained and promoted students when they attended the same grade level, though for the treated group the attendance in that grade comes after an additional year of instruction.

Like some other prior researchers we argue that the within-grade comparison is the most policy relevant because it most aligns with what schools are interested in – the student's performance relative to his same-grade peers (see for instance Alexander, Entwisle, and Dauber 1994). In addition, we point out that if we think of schooling in the long-term context, the additional year of schooling is *precisely* the intervention under consideration. The ultimate question facing those interested in retention policies is whether retained students acquire a greater level of proficiency by the time that they leave school than they would without the intervention in the elementary grade. One potential benefit from retention is the additional year of schooling. If both graduate after four years in high school, then a retained student who completes high school will have received one more year of schooling than a student who was not retained. The ultimate relevant comparison between those students is their final test score in the twelfth grade, not their test score nine years after the third grade. Consequently, at each point in their academic careers, it is the within-grade comparison that is of most interest.

We estimate various forms of (2). Each model is restricted to students observed in the particular grade level under consideration. When estimating models that combine cohorts into a single regression, we only include cohorts for which we observe both the treated and control group in a particular grade. For example, recalling Table 1, the combined estimation of the effect of retention on student performance in grade seven is

only able to include students from Cohorts 1 and 2; while the combined estimation of the effect of retention on student performance in grade four includes all five cohorts. We also report the results of models restricted to individual cohorts in order to follow their progress as they enter later grades.

We estimate models that alter the dependent variable to include the student's score on one of five standardized tests. We analyze math and reading scores from two different assessments: the FCAT, which is used for different facets of the state's accountability system, and the NRT, which is a low-stakes test – a version of the Stanford 10– that was administered to all Florida public school students in grades three through ten through the 2007-08 school year before it was discontinued for budgetary reasons. Finally, we also evaluate student achievement on the state's fifth grade science exam, which is the only grade in which the exam is given that we observe for both retained and promoted students.[4] In order to ease interpretation, we standardize scores on all exams by grade and year to have a mean of zero and standard deviation of one.

**Testing the Identification Assumptions**

The first set of assumptions to consider are those linked to any instrumental variable approach. Use of the Below variable as an instrument assumes that it is correlated with retention and otherwise exogeneous to later student achievement. Discontinuity measures such as Below are widely accepted in the literature as meeting the second criteria because there is no reason to believe that whether or not a student's score falls just above or below an arbitrary threshold is related to later achievement independent of the test score itself.

[4] The test is also administered in the eighth grade. Though we observe both promoted and retained students from Cohort 1 in the eighth grade, we do not include this analysis because of the special nature of Cohort 1 described in a previous section of this paper.

The descriptive statistics provided in Table 2 show that Below also satisfies the first assumption necessary for an instrument – it is clear from the table that those with scores below the threshold are far more likely to be retained than those with scores just above the threshold. In addition, Table 3 verifies that descriptive evidence by reporting the estimates resulting from estimating equation (1) for each cohort individually. For each cohort there is a statistically significant relationship between whether the student's score falls below the threshold for default promotion and the likelihood that the student is retained. Thus, it appears that Below satisfies both qualities of an instrumental variable.[5]

[TABLE 3 ABOUT HERE]

However, the test for using Below in a regression discontinuity framework is more substantial than that for an instrument generally. Because regression discontinuity is meant to approximate a random assignment procedure, a common requirement for strong identification is that the baseline characteristics of students with scores above and below the threshold are statistically identical. We test that identifying assumption by evaluating whether there is a statistically significant difference in the demographic characteristics of students whose scores fall above and below the threshold for default promotion.

Consistent with the identifying assumption, the results of this comparison for all cohorts, shown in Table 4, show few significant differences in the observed characteristics of the two groups. The main exception is that students with scores just below the threshold appear to be slightly more likely to have a disability than students just above the threshold. The few other exceptions where the groups are statistically different are not meaningful in size, and tend to disadvantage the students with scores below the threshold.

[5] We also estimated models that changed the functional form of the forcing variable by adding polynomials of it as regressors, but this produced no meaningful difference in any results. We also tested models that incorporated an interaction between the forcing variable and Below, but this also did not produce a meaningful change. We thus utilize the more parsimonious model in our analysis, which treats the relationship between the forcing variable and retention as linear.

[TABLE 4 ABOUT HERE]

A final factor to consider regarding the validity of the comparison has to do with attrition from the sample over time. The issue of attrition is particularly important in our present analysis, which follows a group of students over a long period of time. Attrition from the sample would be particularly alarming if it is disproportionate to either students who were retained by the policy or promoted. Students could leave our sample if they moved to a school outside of the Florida public school system, dropped out, or for some reason did not have an observable test score in a later year.

Table 5 maps the attrition for the treatment and control group for each of the five cohorts. Notice that for each cohort we observe 100 percent of the retained students the year after the initial third grade year. That is because we can only identify if a student has been retained if he is observed in the subsequent year. At each point, it appears that attrition is relatively similar for both the group of retained and promoted students for each cohort. If we were to characterize a difference, it appears that there was slightly less attrition for students who were retained than from students who were promoted after the first third grade year.

[TABLE 5 ABOUT HERE]

**Results**

We first consider the results of models that combine our cohorts into a single estimation. The tables make clear that, because we do not observe all cohorts in each grade level, not all cohorts are utilized in each estimation. Since at this point we are focused on models that combine cohorts, we only report the results of models that utilize at least two cohorts for estimation. Finally, because of the special case of Cohort 1

described previously, we report the results from models that either include or exclude that first cohort subjected to the policy.

Table 6A reports the results from our estimate of the impact of retention by grade on reading test scores when Cohort 1 is included. The first set of results reports estimates when a school fixed effect is not included, and the second set of columns adds a school fixed effect into the estimation. For these models that utilize multiple cohorts, we prefer the estimates that incorporate a school fixed effect, though results are similar across the specifications.

On both the FCAT and the NRT exam we find evidence that students make substantial improvements in reading proficiency in the years immediately following retention: When retained students are in the fifth grade they perform about a fifth of a standard deviation better than did socially promoted students when they were in the fifth grade. It is also clear from the regressions that the influence of retention on student reading scores tends to decline each year. Nonetheless, by the seventh grade -- five years after the retained student's initial third grade year -- retention is related to about a 0.102 standard deviation increase in reading scores. Results are similar whether or not a school fixed effect is included. Table 6B also shows that results are similar when we exclude Cohort 1 from the analysis.

[TABLE 6A ABOUT HERE]

[TABLE 6B ABOUT HERE]

Table 7 reports the results of regressions that are restricted to students of particular gender or race/ethnicity. In reading, the results suggest that the effect of retention does not meaningfully differ for students in these subgroups.

[TABLE 7 ABOUT HERE]

Our findings from math are similar to those reported for reading. Table 8A shows that retention is related to an early substantial increase in math tests scores that tends to decline each year. However, we again see a statistically significant and relatively substantial influence from retention of about a fifth of a standard deviation when students are in the seventh grade. Results are similar on the high-stakes FCAT exam as well as the low stakes NRT. Table 8B again shows that these estimates are robust to excluding Cohort 1 students from the sample

[TABLE 8A ABOUT HERE]

[TABLE 8B ABOUT HERE]

Table 9 reports the results of regressions looking at whether the relationship between retention and math achievement differed by student gender or race/ethnicity. We again see little evidence that the influence of retention differs meaningfully according to such demographic characteristics.

[TABLE 9 ABOUT HERE]

We now consider how retention has influenced the achievement of students from each cohort separately. Table 10 reports the results of regression models evaluating reading achievement on the FCAT for each cohort by grade. We do not incorporate a school fixed effect in these models because in most cases the more limited sample did not allow for enough observations of students within particular schools to be estimated confidently -- in most cases we observe only an average of four students per school when the sample is restricted by cohort and grade.

The results reported in Table 10 again suggest that retention had a large effect on students early on that tends to fade over time. However, for each cohort except Cohort 1, the effect of retention remains statistically significant in later years, and is often quite

substantial. Retained students from Cohort 2 outperform their socially promoted peers on the FCAT reading test in the seventh grade by about 0.183 standard deviations.

[TABLE 10 ABOUT HERE]

The results are similar in math. Table 11 shows that retained students from Cohort 2 perform about 0.174 standard deviations better than their socially promoted peers in the seventh grade. We again see that the effects from retention tend to decline over time, but remain statistically significant in all cases, this time including Cohort 1.

[TABLE 11 ABOUT HERE]

The results from both the math and reading versions of the NRT, reported in Tables 12 and 13, respectively, are similar to those from the FCAT exams.

[TABLE 12 ABOUT HERE]

[TABLE 13 ABOUT HERE]

Finally, Table 14 reports the results of retention on the state's fifth grade science exam. The results show that the positive effect of retention, at least in fifth grade, are apparent in the student's performance in subjects other than the core subjects of math and reading. For each cohort we find that that retention increased student science achievement in the fifth grade by between a fifth and a quarter of a standard deviation.

[TABLE 14 ABOUT HERE]

**Conclusion**

This paper has evaluated the sustained effects of retention under Florida's test-based promotion policy on student achievement. We find evidence that retention has a large positive effect on student achievement in the years closely following retention, but that the magnitude of this effect declines over time. However, the effect of retention in

the third grade is still statistically distinguishable and of a meaningful magnitude as late as the seventh grade – five years after the retention decision.

The results of this paper add considerably to our knowledge about expanding policies to end social promotion. Though the results of this study are generally positive for the use of test-based retention policies, future research on these programs is necessary. In particular, we are interested in the long term outcomes of early grade retention in the form of the likelihood that a student graduates from high school as well as life outcomes such as labor market earnings. Further, given that it is relatively expensive to educate a child for an extra year, additional research is necessary in order to determine whether the retention intervention is a cost effective strategy for increasing student achievement in the long term.

References

Alexander, K., Entwisle, D., & Dauber, S., (1994). *On the success of failure: A reassessment of the effects of retention in the primary grades*. New York: Cambridge University Press.

Allen, C. S., Q. Chen, V. L. Wilson, & J. N. Hughes (2009). Quality of research design moderates effects of grade retention on achievement: A meta-analytic, multilevel analysis. *Educational Evaluation and Policy Analysis*, 31: 480-499.

Babcock, P., Bedard, K. (2011). The wages of failure: New evidence on school retention and long-run outcomes. *Education Finance and Policy*, 6(3), 293-322.

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.

Greene, J. P., & Winters, M. A. (2009). The effects of exemptions to Flordia's test-based promotion policy: Who is retained? Who benefits academically? *Economics of Education Review*, 28, 135-142.

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, 18, 527-544.

Holmes, C. T. (1989) Grade-level retention effects: A meta-analysis of research studies. In L. A. Shepard & M. L. Smith (Eds.) *Flunking grades: Research and policies on retention* (pp. 16-33). London: Falmer.

Hughes, J. N., Q. Chen, F. Thoemmes, & O. Kwok (2010). An investigation of the relationship between retention in first grade and performance on high stakes tests in third grade. *Educational Evaluation and Policy Analysis*, 32, 166-182.

Imbens, G. W. & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice." *Journal of Econometrics*, 142(2), 615-635.

Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics*, 86, 26-244.

Jacob, B. A., & Lefgren, L. (2007). The effect of grade retention on high school completion. NBER Working Paper W13514.

Jimerson, S. R. (2001). A synthesis of grade retention research: Looking backward and moving forward. *California School Psychologist*, 6, 47-59.

Jimerson, S., Carlson, E., Robert, M. Egeland, B., & Sroufe, L. A. (1997). A prospective, longidudinal study of the correlates and consequences of early grade retention. *Journal of School Psychology*, 35, 3-25.

Roderick, M., & Nagaoka, J. (2005). Retention under Chicago's high-stakes testing program: Helpful, harmful, or harmless? *Educational Evaluation and Policy Analysis*, 27, 309-340.

Rust, J., & Wallace, K. (1993). Effects of grade level retention for four years. *Journal of Instructional Psychology*, 20, 162-166.

Schochet, P., Cook, T., Deke, J., Imbens, G., Lockwood, J.R., Porter, J., Smith, J. (2010). Standards for Regression Discontinuity Designs. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_rd.pdf.

Van der Klaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment: A regression-discontinuity approach. *International Economic Review*, 43(4), 1249-1287.

**Table 1**
**Description of Cohorts**

| | Grade 3 | | Grade 4 | | Grade 5 | | Grade 6 | | Grade 7 | | Grade 8 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| **2002-03** | T1 C1 | | | | | | | | | | |
| | | | | | | | | | | | |
| **2003-04** | T1 T2 C2 | | C1 | | | | | | | | |
| | | | | | | | | | | | |
| **2004-05** | T2 T3 C3 | | T1 C2 | | C1 | | | | | | |
| | | | | | | | | | | | |
| **2005-06** | T3 T4 C4 | | T2 C3 | | T1 C2 | | C1 | | | | |
| | | | | | | | | | | | |
| **2006-07** | T4 T5 C5 | | T3 C4 | | T2 C3 | | T1 C2 | | C1 | | |
| | | | | | | | | | | | |
| **2007-08** | T5 | | T4 C5 | | T3 C4 | | T2 C3 | | T1 C2 | | C1 |
| | | | | | | | | | | | |
| **2008-09** | | | T5 | | T4 C5 | | T3 C4 | | T2 C3 | | T1 C2 |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

C – Control group, not retained
T – Treatment group, retained
Numbers indicate cohort

**Table 2**
**Number and Percent Retained For Students With Particular Third-Grade Reading Scores in Sample**

| | | | *Below Threshold* | | | | | *Above Threshold* | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Initial 3rd Grade Reading Score** | **1027** | **1033** | **1039** | **1045** | | **1051** | **1057** | **1063** | **1069** | | **All 3rd Grade** |
| | | | | | | | | | | | | |
| | Number Promoted | 433 | 443 | 482 | 504 | | 851 | 859 | 920 | 982 | | 151,106 |
| *Cohort 1* | Number Retained | 346 | 300 | 341 | 333 | | 40 | 54 | 33 | 37 | | 24,334 |
| | Percent Retained | 44% | 40% | 41% | 40% | | 4% | 6% | 3% | 4% | | 14% |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | Number Promoted | 508 | 535 | 542 | 603 | | 805 | 833 | 829 | 847 | | 173,836 |
| *Cohort 2* | Number Retained | 220 | 190 | 219 | 168 | | 21 | 38 | 30 | 22 | | 18,664 |
| | Percent Retained | 30% | 26% | 29% | 22% | | 3% | 4% | 3% | 3% | | 10% |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | Number Promoted | 472 | 528 | 540 | 578 | | 836 | 779 | 862 | 843 | | 172,842 |
| *Cohort 3* | Number Retained | 253 | 242 | 216 | 207 | | 30 | 33 | 31 | 27 | | 15,942 |
| | Percent Retained | 30% | 26% | 29% | 22% | | 3% | 4% | 3% | 3% | | 8% |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | Number Promoted | 379 | 364 | 422 | 389 | | 650 | 695 | 664 | 675 | | 178,655 |
| *Cohort 4* | Number Retained | 173 | 176 | 187 | 183 | | 22 | 28 | 26 | 25 | | 11,421 |
| | Percent Retained | 31% | 33% | 31% | 32% | | 3% | 4% | 4% | 4% | | 6% |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | Number Promoted | 495 | 550 | 553 | 586 | | 700 | 733 | 793 | 838 | | 169,163 |
| *Cohort 5* | Number Retained | 146 | 151 | 135 | 137 | | 14 | 31 | 24 | 19 | | 11,604 |
| | Percent Retained | 23% | 22% | 20% | 19% | | 2% | 4% | 3% | 2% | | 6% |

**Table 3**
**Results of First Stage Regression**

| | Cohort 1 | Cohort 2 | Cohort 3 | Cohort 4 | Cohort 5 |
|---|---|---|---|---|---|
| | | | | | |
| Below | 0.347*** | 0.199*** | 0.221*** | 0.276*** | 0.183*** |
| | [0.0231] | [0.0217] | [0.0225] | [0.0283] | [0.0204] |
| | | | | | |
| FCAT Math | -0.000155*** | -0.000132*** | -0.000116*** | -9.89e-05** | -0.000105*** |
| | [3.40e-05] | [3.77e-05] | [3.73e-05] | [4.90e-05] | [3.40e-05] |
| | | | | | |
| FCAT Reading | -0.00059 | -0.00138* | -0.00150* | -0.000296 | -0.000244 |
| | [0.000804] | [0.000778] | [0.000837] | [0.000986] | [0.000716] |
| | | | | | |
| NRT Reading | -0.00397*** | -0.146*** | -0.00214*** | -0.00222*** | -0.00179*** |
| | [0.000289] | [0.0130] | [0.000332] | [0.000382] | [0.000247] |
| | | | | | |
| NRT Math | -0.000745*** | -0.0248** | -0.00108*** | -0.000863*** | -0.000785*** |
| | [0.000289] | [0.0116] | [0.000273] | [0.000333] | [0.000235] |
| | | | | | |
| Male | 0.0173* | 0.0146 | 0.0095 | 0.000299 | 0.0134 |
| | [0.0105] | [0.0105] | [0.0115] | [0.0147] | [0.0103] |
| | | | | | |
| Asian | 0.0135 | -0.0165 | -0.0202 | -0.0423 | -0.0492 |
| | [0.0478] | [0.0375] | [0.0483] | [0.0626] | [0.0439] |
| | | | | | |
| African-American | -0.0219 | -0.00724 | -0.0153 | -0.00413 | -0.0329* |
| | [0.0161] | [0.0170] | [0.0175] | [0.0225] | [0.0175] |
| | | | | | |
| Hispanic | 0.0038 | 0.00359 | 0.00633 | 0.00654 | -0.0305* |
| | [0.0172] | [0.0167] | [0.0165] | [0.0226] | [0.0172] |
| | | | | | |
| Indian | -0.0709 | 0.134 | -0.038 | 0.0186 | -0.140** |
| | [0.0986] | [0.104] | [0.0827] | [0.0910] | [0.0623] |
| | | | | | |
| Multiple Race | -0.0087 | -0.0294 | -0.0218 | 0.00784 | 0.0167 |
| | [0.0360] | [0.0328] | [0.0367] | [0.0450] | [0.0330] |
| | | | | | |
| Lunch – Applied, Not Eligible | -0.0314 | 0.0457 | -0.0393 | -0.0608 | 0.0409 |
| | [0.0341] | [0.0428] | [0.0344] | [0.0456] | [0.0348] |
| | | | | | |
| Eligible for Free Lunch | 0.00747 | 0.00296 | 0.00787 | -0.0232 | 0.0297** |
| | [0.0160] | [0.0150] | [0.0156] | [0.0192] | [0.0141] |
| | | | | | |
| Eligible for Reduced Lunch | 0.0172 | 0.0245 | -0.00925 | -0.0182 | 0.00491 |
| | [0.0207] | [0.0205] | [0.0215] | [0.0285] | [0.0181] |
| | | | | | |
| IEP | -0.0500*** | -0.0531*** | -0.0753*** | -0.0831*** | -0.0385*** |
| | [0.0126] | [0.0115] | [0.0122] | [0.0156] | [0.0114] |
| | | | | | |
| Constant | 3.681*** | 1.550* | 3.686*** | 2.346** | 1.957** |

| | | | | | |
|---|---|---|---|---|---|
| | [0.867] | [0.819] | [0.889] | [1.063] | [0.765] |
| | | | | | |
| School Fixed Effect | √ | √ | √ | √ | √ |
| | | | | | |
| Observations | 7,093 | 6,511 | 6,609 | 4,921 | 6,128 |
| R-squared | 0.455 | 0.406 | 0.417 | 0.475 | 0.405 |

Dependent variable in all models is an indicator for whether the student repeated the third grade the following year. Model is estimated via OLS. Standard errors clustered by school are reported in brackets.

* $p < 0.10$
** $p < 0.05$
*** $p < 0.01$

**Table 4**
**Demographics of Students Above and Below Test Score Threshold**

| | | White | African-American | Hispanic | Lunch – Applied, Not Eligible | Eligible for Free Lunch | Eligible for Reduced Lunch | IEP |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | | |
| *Cohort 1* | Above Threshold | 36.6% | 36.0% | 26.0% | 3.1% | 55.1% | 10.7% | 31.2% |
| | Below Threshold | 34.0% | 35.6% | 26.6% | 2.5% | 56.0% | 11.0% | 32.6% |
| | | | | | | | | |
| *Cohort 2* | Above Threshold | **33.1%** | 37.6% | **25.1%** | 2.6% | 59.0% | 10.1% | **34.5%** |
| | Below Threshold | **30.2%** | 36.6% | **29.0%** | 2.7% | 61.3% | 10.6% | **37.2%** |
| | | | | | | | | |
| *Cohort 3* | Above Threshold | 37.2% | 33.5% | 30.3% | 3.3% | **60.1%** | **10.5%** | 36.3% |
| | Below Threshold | 31.4% | 35.4% | 29.0% | 2.7% | **62.5%** | **8.7%** | 37.8% |
| | | | | | | | | |
| *Cohort 4* | Above Threshold | 27.5% | 38.7% | 28.6% | 2.6% | 63.5% | 10.1% | **39.2%** |
| | Below Threshold | 28.4% | 38.7% | 28.7% | 2.2% | 64.5% | 10.0% | **42.3%** |
| | | | | | | | | |
| *Cohort 5* | Above Threshold | 28.1% | 35.7% | 31.0% | 3.0% | 61.0% | 11.7% | **32.1%** |
| | Below Threshold | 27.4% | 37.5% | 30.8% | 3.0% | 63.2% | 10.9% | **35.4%** |
| | | | | | | | | |
| | | | | | | | | |
| Bold = Significant at $p < 0.05$ | | | | | | | | |

**Table 5**
**Attrition by Cohort and Promotion Status**

| Year | 2002-03 | 2003-04 | 2004-05 | 2005-06 | 2006-07 | 2007-08 | 2008-09 |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| ***Cohort 1*** | | | | | | | |
| Promoted | 5,669 | 5,150 | 5,052 | 4,829 | 4,665 | 4,595 | 4,415 |
| Percent of Beginning | | 91% | 89% | 85% | 82% | 81% | 78% |
| | | | | | | | |
| Retained | 1,558 | 1,558 | 1,485 | 1,437 | 1,368 | 1,317 | 1,282 |
| Percent of Beginning | | 100% | 95% | 92% | 88% | 85% | 82% |
| | | | | | | | |
| ***Cohort 2*** | | | | | | | |
| Promoted | | 5,621 | 5,233 | 5,100 | 4,883 | 4,761 | 4,648 |
| Percent of Beginning | | | 93% | 91% | 87% | 85% | 83% |
| | | | | | | | |
| Retained | | 974 | 974 | 921 | 889 | 849 | 817 |
| Percent of Beginning | | | 100% | 95% | 91% | 87% | 84% |
| | | | | | | | |
| | | | | | | | |
| ***Cohort 3*** | | | | | | | |
| Promoted | | | 5,642 | 5,206 | 5,034 | 4,834 | 4,689 |
| Percent of Beginning | | | | 92% | 89% | 86% | 83% |
| | | | | | | | |
| Retained | | | 1,096 | 1,096 | 1,044 | 1,010 | 971 |
| Percent of Beginning | | | | 100% | 95% | 92% | 89% |
| | | | | | | | |
| ***Cohort 4*** | | | | | | | |
| Promoted | | | | 4,436 | 4,041 | 3,910 | 3,780 |
| Percent of Beginning | | | | | 91% | 88% | 85% |
| | | | | | | | |
| Retained | | | | 879 | 879 | 830 | 802 |
| Percent of Beginning | | | | | 100% | 94% | 91% |
| | | | | | | | |
| ***Cohort 5*** | | | | | | | |
| Promoted | | | | | 5,507 | 5,105 | 4,976 |
| Percent of Beginning | | | | | | 93% | 90% |
| | | | | | | | |
| Retained | | | | | 718 | 718 | 670 |
| Percent of Beginning | | | | | | 100% | 93% |

**Table 6A**
**Regression Results By Grade, Reading**

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | **FCAT** | | | | | |
| | | | | | | | | | |
| | Grade 4 | Grade 5 | Grade 6 | Grade 7 | | Grade 4 | Grade 5 | Grade 6 | Grade 7 |
| | | | | | | | | | |
| Retention (predicted) | 0.371*** | 0.256*** | 0.199*** | 0.117*** | | 0.333*** | 0.201*** | 0.176*** | 0.102*** |
| | [0.0182] | [0.0230] | [0.0260] | [0.0328] | | [0.0202] | [0.0263] | [0.0280] | [0.0347] |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | **NRT** | | | | | |
| | | | | | | | | | |
| Retention (predicted) | 0.256*** | 0.234*** | 0.0995*** | | | 0.205*** | 0.177*** | 0.0826** | |
| | [0.0212] | [0.0243] | [0.0309] | | | [0.0238] | [0.0287] | [0.0345] | |
| | | | | | | | | | |
| Student Controls | √ | √ | √ | √ | | √ | √ | √ | √ |
| Original Third Grade Test Scores | √ | √ | √ | √ | | √ | √ | √ | √ |
| Treated Cohort 1 | √ | √ | √ | √ | | √ | √ | √ | √ |
| Treated Cohort 2 | √ | √ | √ | √ | | √ | √ | √ | √ |
| Treated Cohort 3 | √ | √ | √ | | | √ | √ | √ | |
| Treated Cohort 4 | √ | √ | | | | √ | √ | | |
| Treated Cohort 5 | √ | | | | | √ | | | |
| School Fixed Effect | | | | | | √ | √ | √ | √ |

Dependent variable is the student's score on the FCAT or NRT reading test. Test scores are standardized by grade and year to have a mean zero and standard deviation of one. Student Controls include gender, race/ethnicity, free or reduced priced lunch eligibility status, and an indicator for whether the student is disabled. Standard errors clustered by school reported in brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

**Table 6B**
**Regression Results By Grade Excluding Cohort 1, Reading**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | **FCAT** | | |
| | | | | | | | |
| | Grade 4 | Grade 5 | Grade 6 | | Grade 4 | Grade 5 | Grade 6 |
| | | | | | | | |
| Retention (predicted) | 0.340*** | 0.236*** | 0.223*** | | 0.292*** | 0.168*** | 0.190*** |
| | [0.0218] | [0.0286] | [0.0335] | | [0.0251] | [0.0341] | [0.0371] |
| | | | | | | | |
| | | | | | | | |
| | | | | | **NRT** | | |
| | | | | | | | |
| Retention (predicted) | 0.268*** | 0.231*** | | | 0.190*** | 0.131*** | |
| | [0.0262] | [0.0327] | | | [0.0312] | [0.0437] | |
| | | | | | | | |
| Student Controls | √ | √ | √ | | √ | √ | √ |
| Original Third Grade Test Scores | √ | √ | √ | | √ | √ | √ |
| Treated Cohort 1 | | | | | | | |
| Treated Cohort 2 | √ | √ | √ | | √ | √ | √ |
| Treated Cohort 3 | √ | √ | √ | | √ | √ | √ |
| Treated Cohort 4 | √ | √ | | | √ | √ | |
| Treated Cohort 5 | √ | | | | √ | | |
| School Fixed Effect | | | | | √ | √ | √ |

Dependent variable is the student's score on the FCAT or NRT reading test. Test scores are standardized by grade and year to have a mean zero and standard deviation of one. Student Controls include gender, race/ethnicity, free or reduced priced lunch eligibility status, and an indicator for whether the student is disabled. Standard errors clustered by school reported in brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

**Table 7**
**Regression Results by Grade and Student Demographics, Reading**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | **Male** | | | | |
| | | | | | | | | |
| | Grade 4 | Grade 5 | Grade 6 | Grade 7 | | Grade 4 | Grade 5 | Grade 6 |
| | | | | | | | | |
| Retention (predicted) | 0.371*** | 0.255*** | 0.179*** | 0.0781* | | 0.337*** | 0.269*** | 0.227*** |
| | [0.0263] | [0.0319] | [0.0357] | [0.0470] | | [0.0312] | [0.0397] | [0.0478] |
| | | | | | | | | |
| | | | | | | | | |
| | | | | **Female** | | | | |
| | | | | | | | | |
| Retention (predicted) | 0.369*** | 0.253*** | 0.220*** | 0.160*** | | 0.339*** | 0.189*** | 0.213*** |
| | [0.0247] | [0.0303] | [0.0340] | [0.0429] | | [0.0294] | [0.0375] | [0.0448] |
| | | | | | | | | |
| | | | | | | | | |
| | | | | **African-American** | | | | |
| | | | | | | | | |
| Retention (predicted) | 0.395*** | 0.257*** | 0.217*** | 0.169*** | | 0.368*** | 0.238*** | 0.276*** |
| | [0.0304] | [0.0372] | [0.0413] | [0.0533] | | [0.0367] | [0.0475] | [0.0542] |
| | | | | | | | | |
| | | | | | | | | |
| | | | | **Hispanic** | | | | |
| | | | | | | | | |
| Retention (predicted) | 0.372*** | 0.303*** | 0.198*** | 0.0861 | | 0.334*** | 0.279*** | 0.211*** |
| | [0.0364] | [0.0427] | [0.0506] | [0.0635] | | [0.0445] | [0.0539] | [0.0664] |
| | | | | | | | | |
| | | | | | | | | |
| | | | | **White** | | | | |
| | | | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Retention (predicted) | 0.364*** | 0.235*** | 0.189*** | 0.108** | | 0.327*** | 0.212*** | 0.190*** |
| | [0.0313] | [0.0408] | [0.0442] | [0.0543] | | [0.0379] | [0.0502] | [0.0568] |
| | | | | | | | | |
| Student Controls | √ | √ | √ | √ | | √ | √ | √ |
| Original Third Grade Test Scores | √ | √ | √ | √ | | √ | √ | √ |
| Treated Cohort 1 | √ | √ | √ | √ | | | | |
| Treated Cohort 2 | √ | √ | √ | √ | | √ | √ | √ |
| Treated Cohort 3 | √ | √ | √ | | | √ | √ | √ |
| Treated Cohort 4 | √ | √ | | | | √ | √ | |
| Treated Cohort 5 | √ | | | | | √ | | |

Dependent variable is the student's score on the FCAT reading test. Test scores are standardized by grade and year to have a mean zero and standard deviation of one. Student Controls include gender (excluded from regressions evaluating male or female students), race/ethnicity (excluded from regressions evaluating students of particular race/ethnicity), free or reduced priced lunch eligibility status, and an indicator for whether the student is disabled. Standard errors clustered by school reported in brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

**Table 8A**
**Regression Results By Grade, Math**

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | **FCAT** | | | | | |
| | | | | | | | | | |
| | Grade 4 | Grade 5 | Grade 6 | Grade 7 | | Grade 4 | Grade 5 | Grade 6 | Grade 7 |
| | | | | | | | | | |
| Retention (predicted) | 0.450*** | 0.319*** | 0.219*** | 0.165*** | | 0.414*** | 0.327*** | 0.224*** | 0.204*** |
| | [0.0196] | [0.0229] | [0.0280] | [0.0329] | | [0.0220] | [0.0252] | [0.0288] | [0.0355] |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | **NRT** | | | | | |
| | | | | | | | | | |
| Retention (predicted) | 0.290*** | 0.259*** | 0.180*** | | | 0.244*** | 0.256*** | 0.190*** | |
| | [0.0228] | [0.0236] | [0.0264] | | | [0.0265] | [0.0269] | [0.0291] | |
| | | | | | | | | | |
| Student Controls | √ | √ | √ | √ | | √ | √ | √ | √ |
| Original Third Grade Test Scores | √ | √ | √ | √ | | √ | √ | √ | √ |
| Treated Cohort 1 | √ | √ | √ | √ | | √ | √ | √ | √ |
| Treated Cohort 2 | √ | √ | √ | √ | | √ | √ | √ | √ |
| Treated Cohort 3 | √ | √ | √ | | | √ | √ | √ | |
| Treated Cohort 4 | √ | √ | | | | √ | √ | | |
| Treated Cohort 5 | √ | | | | | √ | | | |
| School Fixed Effect | | | | | | √ | √ | √ | √ |

Dependent variable is the student's score on the FCAT or NRT math test. Test scores are standardized by grade and year to have a mean zero and standard deviation of one. Student Controls include gender, race/ethnicity, free or reduced priced lunch eligibility status, and an indicator for whether the student is disabled. Standard errors clustered by school reported in brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

**Table 8B**

**Regression Results By Grade Excluding Cohort 1, Math**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | **FCAT** | | |
| | | | | | | | |
| | Grade 4 | Grade 5 | Grade 6 | | Grade 4 | Grade 5 | Grade 6 |
| | | | | | | | |
| Retention (predicted) | 0.423*** | 0.277*** | 0.208*** | | 0.382*** | 0.278*** | 0.179*** |
| | [0.0229] | [0.0274] | [0.0366] | | [0.0267] | [0.0328] | [0.0390] |
| | | | | | | | |
| | | | | | | | |
| | | | | | **NRT** | | |
| | | | | | | | |
| Retention (predicted) | 0.262*** | 0.222*** | | | 0.160*** | 0.161*** | |
| | [0.0290] | [0.0306] | | | [0.0373] | [0.0386] | |
| | | | | | | | |
| Student Controls | √ | √ | √ | | √ | √ | √ |
| Original Third Grade Test Scores | √ | √ | √ | | √ | √ | √ |
| Treated Cohort 1 | | | | | | | |
| Treated Cohort 2 | √ | √ | √ | | √ | √ | √ |
| Treated Cohort 3 | √ | √ | √ | | √ | √ | √ |
| Treated Cohort 4 | √ | √ | | | √ | √ | |
| Treated Cohort 5 | √ | | | | √ | | |
| School Fixed Effect | | | | | √ | √ | √ |

Dependent variable is the student's score on the FCAT or NRT math test. Test scores are standardized by grade and year to have a mean zero and standard deviation of one. Student Controls include gender, race/ethnicity, free or reduced priced lunch eligibility status, and an indicator for whether the student is disabled. Standard errors clustered by school reported in brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

**Table 9**
**Regression Results by Grade and Student Demographics, Math**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | **Male** | | | | |
| | | | | | | | | |
| | Grade 4 | Grade 5 | Grade 6 | Grade 7 | | Grade 4 | Grade 5 | Grade 6 |
| | | | | | | | | |
| Retention (predicted) | 0.412*** | 0.305*** | 0.188*** | 0.103** | | 0.384*** | 0.258*** | 0.211*** |
| | [0.0257] | [0.0300] | [0.0402] | [0.0475] | | [0.0307] | [0.0363] | [0.0497] |
| | | | | | | | | |
| | | | | | | | | |
| | | | | **Female** | | | | |
| | | | | | | | | |
| Retention (predicted) | 0.494*** | 0.333*** | 0.257*** | 0.232*** | | 0.465*** | 0.292*** | 0.200*** |
| | [0.0281] | [0.0327] | [0.0366] | [0.0451] | | [0.0326] | [0.0393] | [0.0482] |
| | | | | | | | | |
| | | | | | | | | |
| | | | | **African-American** | | | | |
| | | | | | | | | |
| Retention (predicted) | 0.458*** | 0.307*** | 0.198*** | 0.149** | | 0.440*** | 0.259*** | 0.186*** |
| | [0.0322] | [0.0387] | [0.0469] | [0.0609] | | [0.0385] | [0.0477] | [0.0589] |
| | | | | | | | | |
| | | | | | | | | |
| | | | | **Hispanic** | | | | |
| | | | | | | | | |
| Retention (predicted) | 0.497*** | 0.352*** | 0.269*** | 0.158** | | 0.450*** | 0.297*** | 0.254*** |
| | [0.0407] | [0.0403] | [0.0593] | [0.0674] | | [0.0474] | [0.0487] | [0.0743] |
| | | | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | **White** | | | | |
| | | | | | | | | |
| Retention (predicted) | 0.426*** | 0.319*** | 0.205*** | 0.183*** | | 0.407*** | 0.292*** | 0.177*** |
| | [0.0310] | [0.0382] | [0.0442] | [0.0530] | | [0.0365] | [0.0464] | [0.0598] |

Dependent variable is the student's score on the FCAT math test. Test scores are standardized by grade and year to have a mean zero and standard deviation of one. Student Controls include gender (excluded from regressions evaluating male or female students), race/ethnicity (excluded from regressions evaluating students of particular race/ethnicity), free or reduced priced lunch eligibility status, and an indicator for whether the student is disabled. Standard errors clustered by school reported in brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

**Table 10**
**Regression Results By Grade and Cohort, FCAT Reading**

| | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|
| | | | | | |
| | | *Cohort 1* | | | |
| | | | | | |
| Retention (predicted) | 0.324*** | 0.168*** | 0.0942** | 0.0052 | 0.0207 |
| | [0.0329] | [0.0348] | [0.0384] | [0.0422] | [0.0442] |
| | | | | | |
| | | *Cohort 2* | | | |
| | | | | | |
| Retention (predicted) | 0.311*** | 0.224*** | 0.231*** | 0.183*** | |
| | [0.0404] | [0.0447] | [0.0459] | [0.0491] | |
| | | | | | |
| | | *Cohort 3* | | | |
| | | | | | |
| Retention (predicted) | 0.292*** | 0.238*** | 0.210*** | | |
| | [0.0399] | [0.0499] | [0.0470] | | |
| | | | | | |
| | | *Cohort 4* | | | |
| | | | | | |
| Retention (predicted) | 0.385*** | 0.327*** | | | |
| | [0.0447] | [0.0490] | | | |
| | | | | | |
| | | *Cohort 5* | | | |
| | | | | | |
| Retention (predicted) | 0.409*** | | | | |
| | [0.0490] | | | | |
| | | | | | |
| Student Controls | √ | √ | √ | √ | √ |
| Original Third Grade Test Scores | √ | √ | √ | √ | √ |
| School Fixed Effect | | | | | |

Dependent variable is the student's score on the FCAT reading test. Test scores are standardized by grade and year to have a mean zero and standard deviation of one. Student Controls include gender, race/ethnicity, free or reduced priced lunch eligibility status, and an indicator for whether the student is disabled. Standard errors clustered by school reported in brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

**Table 11**
**Regression Results By Grade and Cohort, FCAT Math**

| | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|
| | | | | | |
| | | *Cohort 1* | | | |
| | | | | | |
| Retention (predicted) | 0.413*** | 0.316*** | 0.150*** | 0.0872* | 0.0888** |
| | [0.0373] | [0.0393] | [0.0432] | [0.0446] | [0.0450] |
| | | | | | |
| | | *Cohort 2* | | | |
| | | | | | |
| Retention (predicted) | 0.360*** | 0.268*** | 0.199*** | 0.174*** | |
| | [0.0431] | [0.0470] | [0.0526] | [0.0525] | |
| | | | | | |
| | | *Cohort 3* | | | |
| | | | | | |
| Retention (predicted) | 0.368*** | 0.276*** | 0.217*** | | |
| | [0.0455] | [0.0461] | [0.0483] | | |
| | | | | | |
| | | *Cohort 4* | | | |
| | | | | | |
| Retention (predicted) | 0.416*** | 0.328*** | | | |
| | [0.0438] | [0.0506] | | | |
| | | | | | |
| | | *Cohort 5* | | | |
| | | | | | |
| Retention (predicted) | 0.523*** | | | | |
| | [0.0479] | | | | |
| | | | | | |
| Student Controls | √ | √ | √ | √ | √ |
| Original Third Grade Test Scores | √ | √ | √ | √ | √ |
| School Fixed Effect | | | | | |

Dependent variable is the student's score on the FCAT math test. Test scores are standardized by grade and year to have a mean zero and standard deviation of one. Student Controls include gender, race/ethnicity, free or reduced priced lunch eligibility status, and an indicator for whether the student is disabled. Standard errors clustered by school reported in brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

**Table 12**
**Regression Results By Grade and Cohort, NRT Reading**

| | Grade 4 | Grade 5 | Grade 6 | Grade 7 |
|---|---|---|---|---|
| | | | | |
| | | *Cohort 1* | | |
| | | | | |
| Retention (predicted) | 0.153*** | 0.195*** | 0.0437 | 0.0201 |
| | [0.0362] | [0.0367] | [0.0393] | [0.0436] |
| | | | | |
| | | *Cohort 2* | | |
| | | | | |
| Retention (predicted) | 0.238*** | 0.248*** | 0.108** | |
| | [0.0404] | [0.0446] | [0.0480] | |
| | | | | |
| | | *Cohort 3* | | |
| | | | | |
| Retention (predicted) | 0.277*** | 0.230*** | | |
| | [0.0432] | [0.0475] | | |
| | | | | |
| | | *Cohort 4* | | |
| | | | | |
| Retention (predicted) | 0.337*** | | | |
| | [0.0486] | | | |
| | | | | |
| Student Controls | √ | √ | √ | √ |
| Original Third Grade Test Scores | √ | √ | √ | √ |
| School Fixed Effect | | | | |

Dependent variable is the student's score on the NRT reading test. Test scores are standardized by grade and year to have a mean zero and standard deviation of one. Student Controls include gender, race/ethnicity, free or reduced priced lunch eligibility status, and an indicator for whether the student is disabled. Standard errors clustered by school reported in brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

**Table 13**
**Regression Results By Grade and Cohort, NRT Math**

| | Grade 4 | Grade 5 | Grade 6 | Grade 7 |
|---|---|---|---|---|
| | | | | |
| | | *Cohort 1* | | |
| | | | | |
| Retention (predicted) | 0.254*** | 0.257*** | 0.155*** | 0.0865** |
| | [0.0377] | [0.0379] | [0.0357] | [0.0417] |
| | | | | |
| | | *Cohort 2* | | |
| | | | | |
| Retention (predicted) | 0.245*** | 0.257*** | 0.148*** | |
| | [0.0392] | [0.0439] | [0.0435] | |
| | | | | |
| | | *Cohort 3* | | |
| | | | | |
| Retention (predicted) | 0.277*** | 0.199*** | | |
| | [0.0416] | [0.0425] | | |
| | | | | |
| | | *Cohort 4* | | |
| | | | | |
| Retention (predicted) | 0.281*** | | | |
| | [0.0404] | | | |
| | | | | |
| Student Controls | √ | √ | √ | √ |
| Original Third Grade Test Scores | √ | √ | √ | √ |
| School Fixed Effect | | | | |

Dependent variable is the student's score on the NRT math test. Test scores are standardized by grade and year to have a mean zero and standard deviation of one. Student Controls include gender, race/ethnicity, free or reduced priced lunch eligibility status, and an indicator for whether the student is disabled. Standard errors clustered by school reported in brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

**Table 14**
**Regression Results By Cohort, FCAT Fifth Grade Science Exam**

| | All Cohorts | All But Cohort 1 | Cohort 1 | Cohort 2 | Cohort 3 | Cohort 4 |
|---|---|---|---|---|---|---|
| | | | | | | |
| Retention (predicted) | 0.268*** | 0.216*** | 0.261*** | 0.229*** | 0.242*** | 0.256*** |
| | [0.0222] | [0.0262] | [0.0416] | [0.0460] | [0.0511] | [0.0519] |
| | | | | | | |
| Student Controls | √ | √ | √ | √ | √ | √ |
| Original Third Grade Test Scores | √ | √ | √ | √ | √ | √ |

Dependent variable is the student's score on the FCAT fifth grade science test. Test scores are standardized by grade and year to have a mean zero and standard deviation of one. Student Controls include gender, race/ethnicity, free or reduced priced lunch eligibility status, and an indicator for whether the student is disabled. Standard errors clustered by school reported in brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$