The Effect of Florida's Test-Based Promotion Policy on Student Performance Prior to the Retention Decision

Marcus A. Winters, Ph.D.
Associate Professor
School of Education
Boston University
2 Silber Way
Boston, MA 02215
917-270-6154
Marcusw@bu.edu

Abstract:

Test-based promotion policies require students to demonstrate minimal proficiency on one or more standardized tests in order to be default promoted to the next grade. Though these policies have clear implications for students retained by them, they also provide incentives to both students and schools to make academic improvements in order to avoid the threat of retention. I utilize a difference-in-difference design to measure the impact of Florida's third grade test-based promotion policy on student test scores within the third grade. I find evidence that the policy led to a significant test score increase that was largest for students who entered the third grade with relatively low reading ability.

Acknowledgements: This research was funded by a grant from the Kellogg Foundation.

1. Introduction

Sixteen states and several urban school districts require students to score above a minimum threshold on standardized tests (always reading, and sometimes additionally math) in order to be default promoted to the next grade (Workman 2014). The large majority of these policies hold the third grade as the gateway grade for promotion, under the argument that third is the grade in which students "stop learning to read and start reading to learn" (Center for Public Education 2015).

Recent studies have found somewhat mixed results for the effect of being retained as a consequence of test-based promotion policies (Jacob and Lefgren 2004, 2009; Mariano and Martorell 2013; Roderick and Nagaoka 2005; Greene and Winters 2007, 2009; Winters and Greene 2012, Schwerdt, West, and Winters 2017; Ozek 2015). Most relevant to this paper, Schwerdt, West, and Winters (2017) found that treatment under Florida's third grade test-based promotion policy led to large short run test score improvements that faded over time, decreased the number of remedial courses that students took in high school, and raised high school grade point average, but the treatment did not impact the probability that students graduated from high school or enrolled in post-secondary education.

The effect of retention under test-based promotion policies warrants the considerable attention is has received in the literature. But this research likely does not provide a full picture of the impact that these policies have on student learning throughout the jurisdiction.

In addition to their impact on students retained by them, test-based promotion policies might also have broad impacts by providing both students and schools with a strong incentive to make academic improvements before and within the gateway grade in order to avoid retention.

Any such pre-retention effects would likely add to the overall effect of the policy on the performance of students retained by it and could apply to at least a portion of the far larger group of students who are not retained. If widely distributed across students, even a small pre-retention impact from a test-based promotion policy would likely lead to substantially larger overall educational gains than does the retention treatment itself, which is restricted to a limited number of students (Eide and Goldhaber 2005; Babcock and Bedard 2011; Schwerdt, West, and Winters 2017).

I measure the impact of Florida's third grade test-based promotion policy on student test score performance within the third grade. Florida's policy is particularly interesting to consider because it is often cited by those advocating for other states to adopt test-based promotion. Florida also experienced substantial test score improvements – both on the statewide exams and on the National Assessment of Educational Progress – in early elementary grades in this time period that have not yet been satisfactorily explained (Winters 2012).

I apply a difference-in-difference approach that measures whether there was an increase in third grade test scores relative to other grades within the school in the policy's first year. The motivation for this approach is that threat of retention under the policy impacted students in the third grade but not students in higher grades who were not in danger of being retained by it.

I look for an impact of Florida's test-based promotion policy in two datasets. I first evaluate the overall average effect of the policy on third grade test scores using statewide longitudinal school-by-grade level data. I also use longitudinal student-level data from a large

Florida school district, Hillsborough County (the eighth largest U.S. public school district[1]), which allows me to assess whether the treatment effect differed by prior student reading scores.

I find evidence that the introduction of Florida's test-based promotion policy led to a statistically significant and substantial increase in average third grade test scores. Though the reading test was the only one that triggered retention, the effect was similar in both math and reading. Using the data from Hillsborough, I find evidence that the policy had a large positive effect on the performance of students who entered the third grade with very low reading proficiency but had no significant effect on the performance of students who entered the third grade with relatively high reading scores. This result is consistent with the theory that the policy would provide schools with an incentive to focus on low-performing students, but it also suggests that any such reallocation did not reduce the performance of high-achieving students who were unlikely to be retained by the policy.

The remainder of the paper proceeds as follows. Section 2 describes the prior research evaluating the impact of test-based promotion policies on student performance within the gateway grade. Section 3 describes the policy and the data. Section 4 describes the estimation strategy. Section 5 reports results. And Section 6 provides a brief summary and concludes.

2. Previous Research

Though the literature on the effect of retention on later student performance is becoming somewhat thick, the research evaluating the effect of test-based promotion policies on student achievement prior to the retention decision is quite limited. In addition, this study contributes to an important body of research evaluating the distributional effects of accountability policies.

---

[1] Digest of Education Statistics 2015, Table 215.10

Test-based promotion can be thought of as among the class of accountability policies that attach rewards or consequences to scoring above a particular threshold on a test. Research finds that such policies can increase average student achievement (see for example Dee and Jacob 2011), but that the effect often varies according to the student's prior performance relative to the benchmark (Neal and Schanzenback 2010; Reback 2008; Jacob 2005). Policies triggered by scoring below a particular threshold give schools an incentive to focus on students with scores near the cutoff. On one hand, schools might determine that very low performing students have little chance of passing the test. On the other hand, schools could pull attention away from high performing students who are in little danger of scoring below the policy threshold.

It is at least somewhat odd to think about test-based promotion policies within an accountability framework because retention is intended to be an educational intervention rather than a punishment. In another sense, however, test-based promotion policies are particularly interesting to consider within an accountability framework because unlike school-centered policies they provide clear incentives both for schools but perhaps especially for students. We might expect that these policies could thus engage the participation of students in a way that other school-based accountability policies (for example, grading schools from A to F) do not. However, to date few studies have specifically evaluated the impact that test-based promotion policies have on student performance prior to the retention decision.

Using a hierarchical linear model, Roderick, Jacob, and Byrk (2002) find that Chicago's test-based promotion policy for students in grades three, six, and eight led to substantially improved student performance in both math and reading. In the third grade they found a positive treatment effect in math across the distribution of prior performance that was largest for students most at risk for retention. But while third grade students made substantial improvements on the reading exam on average, higher performing students with little to no risk of falling below the policy threshold experienced substantial test score declines. Jacob (2005) additionally found that the positive effect from Chicago's policy were not apparent on tests other than those used to determine promotion.

In an evaluation of the fifth grade test-based promotion policy in New York City, Mariano et al. (2009) compared the performance of entering fifth graders in the first year of the policy to that of a matched comparison group of entering fifth graders the previous year. They found a substantial positive effect within the gateway grade on the English Language Arts exam for students with considerable risk of falling below the policy threshold without intervention. But they found no effect on student math performance.

3. Setting and Data

3.1. Florida's Test-Based Retention Policy

Florida's test-based promotion policy requires students to demonstrate basic reading proficiency before they are promoted to the fourth grade. Students scoring Level 1 (the lowest of five levels) on the reading portion of the Florida Comprehensive Assessment Test (FCAT) are flagged for retention. Third grade students in the 2002-03 school year were the first subjected to the policy.

Students scoring below the threshold could receive one of several exemptions and be promoted. About 46 percent of students scoring below the threshold in the first year of the policy were nonetheless promoted (Greene and Winters 2009). Still, the policy considerably increased the use of grade retention in the state: 2.8 percent of third grade students were retained following the year prior to the policy's implementation, compared to 13.5 percent in the first year of the policy (Schwerdt, West, and Winters 2017).

3.2 Standardized Testing in Florida

In Florida, as in many states, statewide standardized tests are administered in the spring to all public school students in grades three through ten. A fairly unique feature in Florida is that during the time period of this study students were required to take both a criterion referenced high-stakes test and a norm-referenced low-stakes test in both math and reading.

The FCAT is a high-stakes test tied to the state's learning standards. The test was introduced in 1998, but was first expanded to all grades three through ten in 2001. It is the FCAT reading test on which students had to score above a threshold in order to avoid retention.

Unfortunately for the purposes of this paper, the state also used the FCAT for its school accountability system in a way that is problematic for estimating a treatment effect from the test-based promotion policy. Initially the state's accountability system used the proportion of students scoring proficient or above to assess school quality. Beginning in 2003, the treatment year for the present analysis, the state began using student growth on the FCAT reading test for school accountability purposes in grades four through eight. That policy change is problematic for my estimation procedure which, as described in Section 4, compares changes in the trend of the performance of third grade students during the initial treatment year to that of students in fourth through eighth grades. The simultaneous increase in the incentive to improve FCAT reading scores in grades four through eight would tend to bias the estimated treatment effect of the test-based promotion policy on student performance within the third grade downward.[2]

Fortunately, at this time the state also administered a commercial norm-referenced exam, the Stanford-9, in both math and reading to all students in grades three through ten statewide.[3] Results on the Stanford-9 exam played no role in either student or school accountability and were used for informational purposes and thus were not likely to have been influenced by factors such as teaching-to-the-test or other manipulations.[4] I thus rely on data for student scores on the Stanford-9 to estimate the treatment effect.

3.3 Data

I evaluate the impact of Florida's test-based promotion policy within two datasets. I first evaluate the impact of the policy statewide using longitudinal school-by-grade level test scores and demographic characteristics downloaded from the Florida Department of Education website. Though student-level data is often preferred, it is not essential in this case. The primary benefit of student-level data would be increased precision, but the estimates reported in the next section are precise enough to detect relatively small effects.

Even if student-level data were available, I would not be able to evaluate whether the effect of the policy differed by prior student performance because statewide testing begins in the

[2] Consistent with this interpretation, there was a large increase in FCAT reading scores in grades four through eight in 2003 that is not apparent on the FCAT math test or the Stanford-9. In addition, though possibly related to the differences in stakes, the prior trends assumption required for the empirical design holds well for the Stanford-9 but not for the FCAT. Results on the FCAT exam change substantially when accounting for prior trend. Though accounting for trends is useful and does not necessarily prohibit a causal interpretation, the fact that we observe only a single year of prior trend makes these results difficult to interpret.

[3] The state referred to the exam as the FCAT-Norm Referenced Tests (FCAT-NRT). However, it was simply the Stanford-9.

[4] The state discontinued administration of a norm-referenced test in 2008, well after the time period analyzed in this paper.

third grade. However, during the time-period under consideration several school districts additionally administered the Stanford-9 exam in the second grade. I acquired the longitudinal student-level administrative test score and demographic data from the Hillsborough County school district (Tampa and surrounding area), which during the time period of the analysis independently administered the Stanford-9 to all second grade students.[5]

3.4 Estimation Sample

The estimation sample includes data from each observed grade within schools that offer the third grade and that posted valid test scores and demographic characteristics used in the respective analysis. Most schools in the sample include grades three, four and five. However, there are also some K-8 schools, and in those cases I include all of the higher grades. Excluding grades higher than five does not meaningfully change the results.

The analysis includes observations regarding characteristics and from the spring administrations of the Stanford-9 math and reading tests in 2001, 2002, and 2003. In Hillsborough, I am able to observe scores in grades two through eight on the Stanford-9 beginning in 2000, and so the need for a prior year score to evaluate heterogeneous effects does not require the loss of a year from the sample.

Tables 1 and 2 report descriptive statistics from the reading estimation samples for students in the third grade and all higher included grades for the statewide and the Hillsborough County analyses, respectively.

---

[5] I also acquired data from Lee and Broward Counties, which also administered a second grade test. However, I do not report results from these districts because the data did not allow for an analysis that would produce clearly interpretable causal policy effects. For example, useable data from Broward County begins in 2001, and thus for the evaluation of the heterogeneous effects of the policy by prior student proficiency I observed only a single pre-policy year and thus cannot evaluate or control for differences in trends.

[TABLE 1 ABOUT HERE]

[TABLE 2 ABOUT HERE]

4. Methodology

I employ a difference-in-difference (DiD) design to estimate treatment effects. The first difference is across years. The second difference is across grade levels. Of primary interest is whether there was a significant change in the trajectory of third grade test scores compared to the trajectory in other grades in the first year the policy was implemented. The motivation of this approach is that the test-based promotion policy would provide an incentive within the third grade and earlier grades but not in later grades where students faced no danger of retention under the policy.

The model for estimation takes the form:

$$(1)\ y_{gst} = \alpha + \gamma_g + \delta_s + \varphi_t + \sum_g \sum_t \beta_{gt}(\gamma_g \times \varphi_t) + \theta X_{gst} + \sum_g \lambda_g(\gamma_g \times t) + \epsilon_{gst}$$

Where y_{gst} is the average test score within grade g of school s in year t, X is a vector of time-variant observed characteristics; γ, δ, and φ are fixed effects for the grade, school, and year; t is a linear time trend that varies by grade level and ϵ is a stochastic term clustered by school. Though not reported, results are very similar in models that exclude the prior trend variables. The coefficient of interest is $\beta_{3,2003}$, which represents the causal effect of the policy on the average student test scores within the third grade.[6]

---

[6] Causal interpretation of β relies on the assumption that there was no other change that influenced third grade scores in 2003 relative to scores in other grades. In the fall of 2001, then-Governor Jeb Bush signed an executive order calling for the introduction of the Just Read program, which included the addition of reading coaches in many schools and over the years would include a variety of initiatives meant to improve early grade reading. The state also began to participate in the federal Reading First initiative in the 2002-03 school year. However, the implementation of these programs was not uniform and was quite limited at the time period analyzed in this

In the Hillsborough sample I then expand the analysis to evaluate whether the treatment effect differed by prior student test score. The model adds indicator variables for whether the student's prior score fell within one of six thresholds on the reading exam and interactions between the treatment year and each of six groupings of prior performance categories. The model also allows the linear time trend to vary by grade and group. Formally, I expand upon (1) to estimate:

$$(2)\ y_{igst} = \alpha + \gamma_g + \delta_s + \varphi_t + \eta_l + \sum_g \sum_t \sum_l \beta_{gtl}(\gamma_g \times \varphi_t \times \eta_l) + \theta X_{igst} + \sum_g \sum_l \lambda_{gl}(\gamma_g \times \eta_l \times t) + \epsilon_{igst}$$

where η are fixed effects for the group l, which indicates one of six categories for the previous year's reading score.

5, Results

5.1 Visual Evidence

Figure 1 compares the trajectory of average math and reading Stanford-9 scores in the third grade to the trajectory in grades four through eight. Between 2000-01 and 2001-02, prior to the policy, the average scores in grade three grew at a similar trajectory as the later grades. There was a noticeable increase in test scores between 2001-02 and 2002-03, after adoption of the treatment, in grade three that is not apparent in the other grades, which suggests the potential for a treatment effect.

paper. In the 2002-03 school year, the treatment year of the current analysis, 42 schools in Hillsborough County participated in Reading First and no schools received reading coaches. To account for any bias due to adoption of these policies, the analyses below include controls for school participation in these programs. Excluding these variables has no meaningful impact on the estimated effect of the test-based promotion policy.

[FIGURE 1 ABOUT HERE]

5.2 Regression Estimates

Table 3 reports regression results. Looking first at the results from the statewide analysis, the policy led to a statistically significant test score gain on the Stanford-9 exam in both math and reading. The effect translates to an increase of about 0.11 standard deviations on the school-level distribution or 0.06 standard deviations on the student test score distribution in each subject.

[TABLE 3 ABOUT HERE]

In Hillsborough, the analysis finds a significant positive average treatment effect in both subjects that is similar in magnitude to the estimated effect statewide. Table 3 also reports results from models evaluating whether the treatment effect varied by prior student reading score. The table is structured so that the *Grade 3 * Cohort 0203* variable represents the effect for students with prior Stanford-9 reading scores in the bottom decile. The remaining coefficients in the table represent the difference between the effect for those in the bottom decile and that for the respective group. The table reports the results of F-tests for whether the sum of the coefficient on *Grade 3 * Cohort 0203* and the respective interaction term is different from zero. However, since the results for each percentile grouping are difficult to interpret from the table, Figures 2 and 3 illustrate the estimate and 95 percent confidence interval for each group in reading and math, respectively.

[FIGURE 2 ABOUT HERE]

[FIGURE 3 ABOUT HERE]

The policy had a substantial positive impact on the math and reading scores of students who entered the third grade with reading test scores in the bottom quartile. For example, students who entered the third grade in the bottom decile of second grade reading test scores experienced test score improvements due to the policy that amount to about 0.17 standard deviations in reading and 0.09 standard deviations in math.

In math, the magnitude of the effect is relatively similar for each subgroup, though the estimate narrowly misses the cutoff for significance at the five percent level for those with entering scores between the 50^{th} and 75^{th} percentiles and those with entering scores between the 76^{th} and 9^{th} percentiles. In reading, the effect of the policy was large for relatively low-performing students at entry and declines in prior student achievement. There again is no significant effect for students with prior reading scores in the two highest achieving groups considered.

6. Summary and Conclusion

I find evidence that the introduction of Florida's test-based promotion policy led to meaningful average test score improvements within the third grade statewide. I also find evidence that the effect of the policy differed depending on the student's entering level of reading proficiency. The policy had a large positive effect on the lowest performing students but no significant effect on students who entered the third grade with relatively high reading scores.

My results have important consequences both for the academic literature and for policymakers. In particular, my findings suggest that the effect of test-based promotion policies expands beyond the impact of retention in an important way that must be considered when evaluating their overall effectiveness. Florida's policy led to meaningful test score improvements for a much larger group of students than those who were eventually retained by it. Further, though they performed substantially better in the third grade then they would have in absence of the policy, many students who entered the third grade with very low reading ability nonetheless scored below the policy threshold and were retained. Thus, prior analyses of the effect of retention under test-based promotion policies on later student performance likely understate the overall impact of the policy for retained students because they omit the policy-induced gains within the third grade prior to the retention decision.

My finding that Florida's policy led to large test score improvements for previously low-performing students but did not benefit students who entered the gateway grade with relatively high test scores is consistent with the theory that a threshold-based policy would lead to a reallocation of resources toward students who are in danger of being retained under the policy. Importantly, however, this result also suggests that any such reallocation in this case did not lead to significant harms for other students. This finding differs from the results from Chicago, where Rodrick, Jacob, and Bryk (2002) found evidence that the policy reduced the achievement of high-performing students.

There are at least two additional important differences between the estimates from Florida and those from New York City that are worth considering. First, while in those districts the policy did not have a positive effect on students entering with very low test scores, in Florida I find that even those with prior reading scores in the bottom decile made significant gains. Further, the positive effect in Florida appears on a test (the Stanford-9) and in a subject (math) that was not used for the retention decision or for any other formal accountability purpose, which contrasts with Jacob's (2005) findings in Chicago.

Unfortunately, there is not a clear way to deconstruct the reasons for the different findings in Florida and these other districts. However, there are some notable differences in the designs of these policies that are worth future consideration. For instance, the test score standard that students were required to reach under New York City's policy was far lower than the standard in either Chicago or Florida. Also, both Chicago and New York City required students to score above a certain threshold on both a math and reading exam, while Florida's policy required students to demonstrate minimal proficiency only in reading.

One limitation of the analysis is that I am not able to evaluate whether students maintained the benefits caused by the threat of retention in later years or if the effect faded over time. Unfortunately, such an analysis is impossible because many of the students (about 13.5 percent of the first cohort) were retained under the policy the following year. When evaluating later test scores it is not obvious how one would disentangle the effect of retention from that of the response within the third grade in an attempt to avoid retention.

It is similarly difficult to convincingly estimate the causal effect of the threat of retention on the performance of later entering third grade cohorts. Winters (2012) presents descriptive evidence that the test scores of entering third grade students in Florida grew substantially over time after adoption of the policy. Causal analysis of these data is complicated by the large increase in retention due to the policy fundamentally altering the student bodies in each grade over time. Further, the state continued its focus on improving early grade literacy during this time period. The results of those interventions would also be apparent on third grade scores.

Prior research found that retention under Florida's test-based promotion policy led to increased academic performance and did not reduce the likelihood that students graduated from high school (Schwerdt, West, and Winters 2017). This paper adds evidence that the policy also had a broad positive impact on the vast majority of students within the third grade, a time that research suggests is important for determining the student's later academic development. This combination of results is promising for the use of test-based promotion policies based on Florida's model.

References

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

Center for Public Education (2015). Learning to Read, Reading to Learn: Why third grade is a pivotal year for mastering literacy.

Dee, T. S. & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management,* 30(3), 418-446.

Eide, Eric R. and Dan D. Goldhaber (2005). Grade retention: What are the costs and benefits? *Journal of Education Finance*, 31(2):195–214.

Greene, J. P. and M. A. Winters (2007). Revisiting grade retention: An evaluation of Florida's test-based promotion policy. *Education Finance and Policy*, Vol. 2, Number 4, Fall.

Greene, J. P. and M. A. Winters (2009). The effects of exemptions to Florida's test-based promotion policy: Who is retained? Who benefits academically? *Economics of Education Review*, 28(1), pp: 135-142.

Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago public schools. *Journal of Public Economics*, 89: 761-796.

Jacob, B. A., & Lefgren, L. (2004). Remedial Education and Student Achievement: A Regression Discontinuity Analysis. The Review of Economics and Statistics, 86(1): 226-244.

Jacob, B.A., & Lefgren, L. (2009). The Effect of Grade Retention on High School Completion. American Economic Journal: Applied Economics 1(3): 33-58.

Mariano et al. (2009). Ending Social Promotion Without Leaving Children Behind: The Case of New York City. Santa Monica, CA: RAND Corporation.

Mariano, L., & Martorell, P. (2013). The academic effects of summer instruction and retention in New York City. Educational Evaluation and Policy Analysis, 35(1): 96–117.

Neal, D. & D. W. Schanzenback (2010). Left behind by design: Proficiency counts and test-based accountability. *Review of Economics and Statistics*, 92(2): 263-283.

Ozek, U. (2015). Hold back to move forward? Early grade retention and student misbehavior. Education Finance and Policy, 10(3), 350-377.

Reback, R. (2008). Teaching to the rating: School accountability and the distribution of student achievement. *Journal of Public Economics*, 92: 1394-1415.

Roderick, M., Jacob, B. A., & Bryk, A. S. (2002). The impact of high-stakes testing in Chicago on student achievement in promotional gate grades. Educational Evaluation and Policy Analysis, 24, 333-357.

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.

Schwerdt, G., M. R. West, and M. A. Winters (2017). The Effects of Test-based Retention on Student Outcomes over Time: Regression Discontinuity Evidence from Florida. *Journal of Public Economics*, 152, pp: 154-169.

Winters, M. A. (2012). Florida defeats the skeptics. *Education Next*, 12(4).

Winters, M. A. and J. P. Greene (2012). The medium-run effects of Florida's test-based promotion policy. *Education Finance and Policy*, 7(3), pp: 305 - 330.

Workman, E. (2014). Third grade reading policies. Technical Report, Denver, CO: Education Commission of the States.

Figure 1
Stanford-9 Test Score Trends

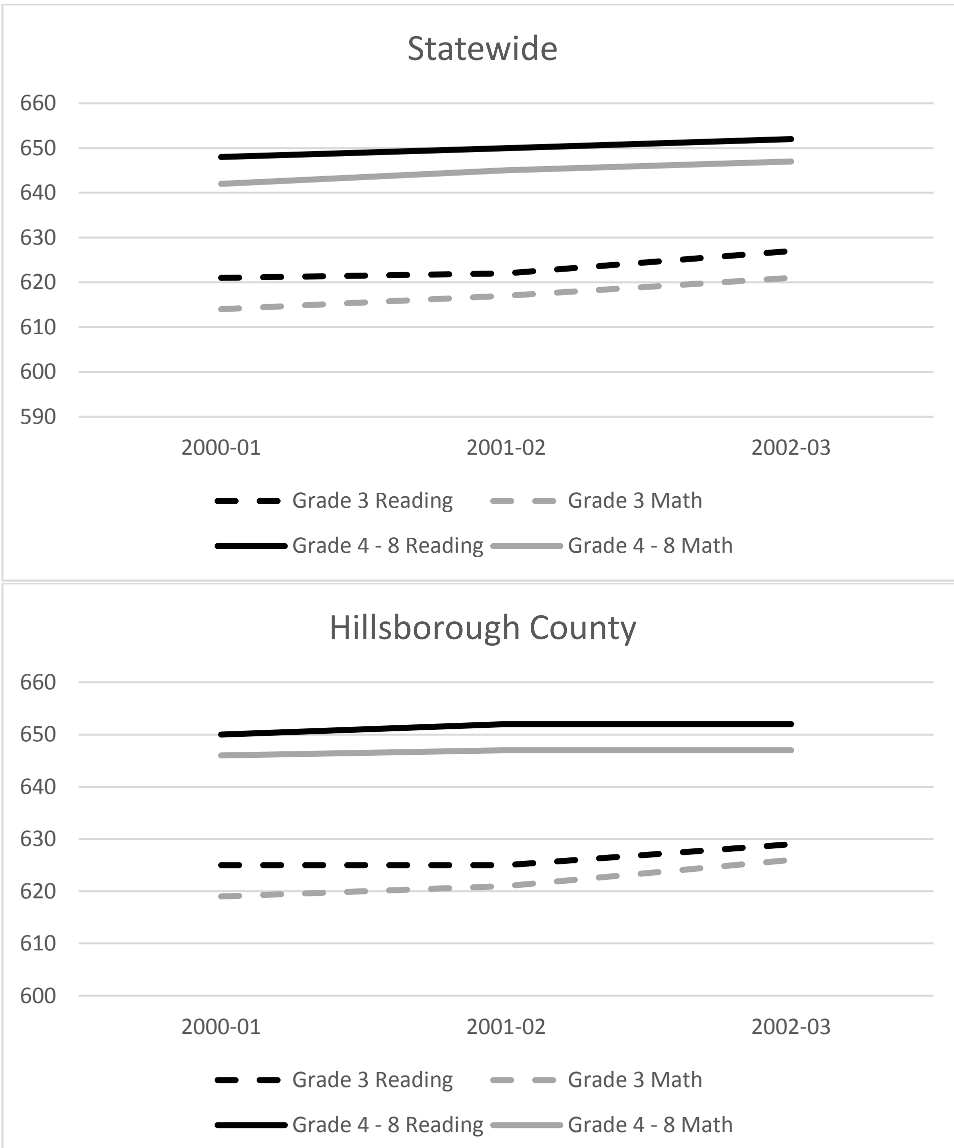
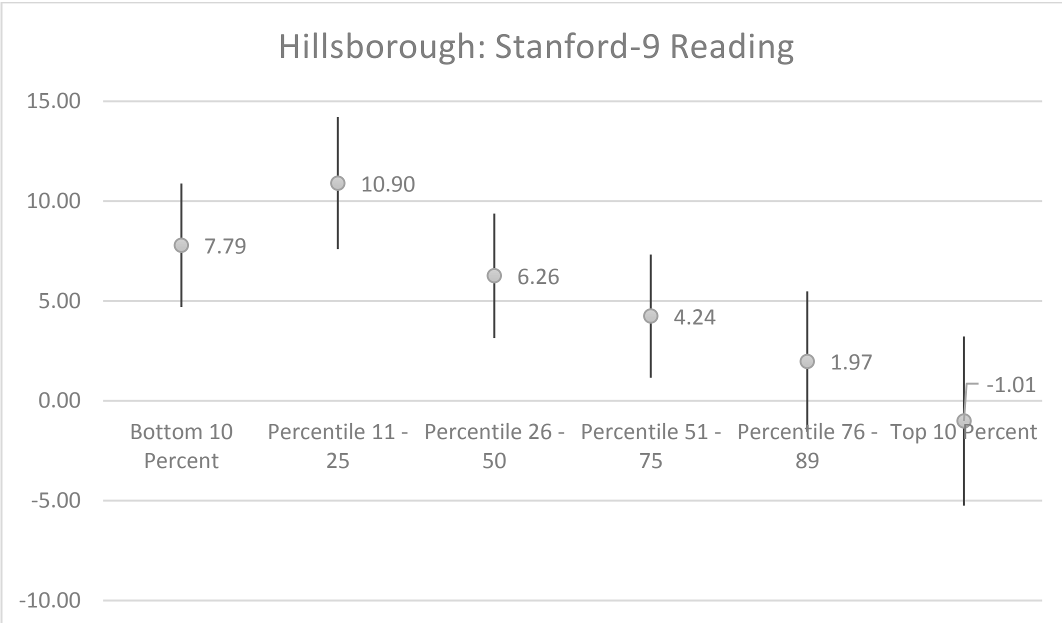
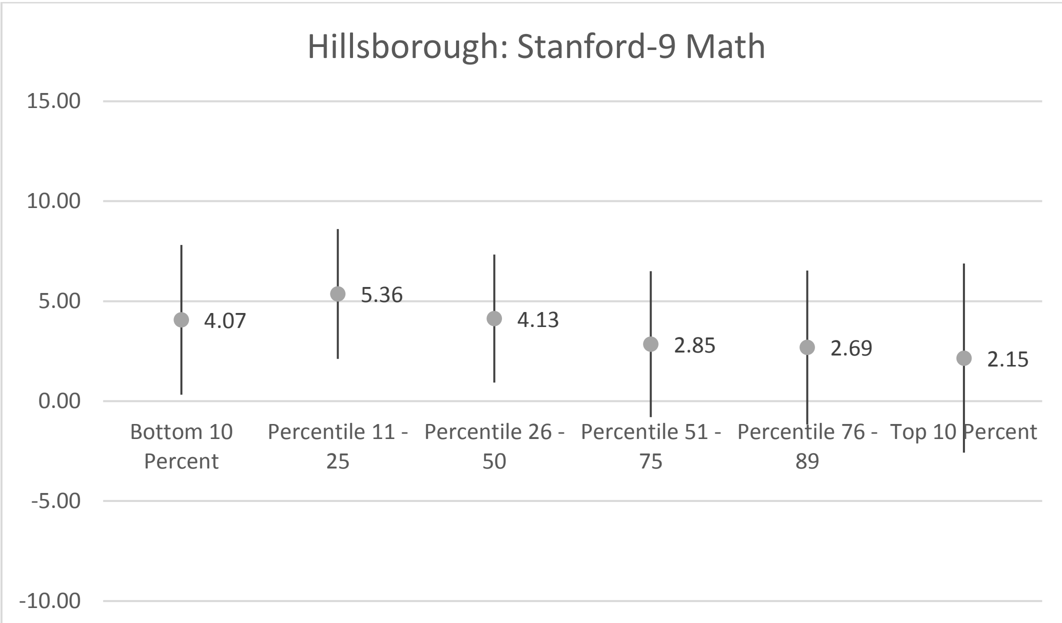


Figure 2
Effect of Policy by Percentile Grouping: Reading



Note: Figure illustrates the coefficient and 95 percent confidence interval for each percentile grouping using the regression results reported on Table 3.

Figure 3
Effect of Policy by Percentile Grouping: Math



Note: Figure illustrates the coefficient and 95 percent confidence interval for each percentile grouping using the regression results reported on Table 3.

Table 1
Descriptive Statistics for Statewide Analysis

	2000-01		2001-02		2002-03	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
			Third Grade			
Average SAT-9 Reading	621.1	18.2	622.5	18.3	626.9	18.2
Average SAT-9 Math	614.0	17.0	616.8	17.2	621.3	16.7
Reading First	0.000	0.000	0.000	0.000	0.175	0.380
Special Education	0.157	0.062	0.156	0.063	0.158	0.065
Free or Reduced Priced Lunch	0.551	0.247	0.542	0.248	0.550	0.252
English Language Learner	0.085	0.114	0.088	0.114	0.080	0.107
			Grades Four Through Eight			
Average SAT-9 Reading	648.1	19.3	649.8	19.3	651.9	19.6
Average SAT-9 Math	642.4	19.9	645.2	19.9	646.9	19.9
Reading First	0.000	0.000	0.000	0.000	0.168	0.374
Special Education	0.159	0.068	0.157	0.067	0.162	0.083
Free or Reduced Priced Lunch	0.546	0.247	0.536	0.250	0.545	0.253
English Language Learner	0.084	0.114	0.086	0.114	0.076	0.105
Grade 4	0.463	0.499	0.459	0.498	0.455	0.498
Grade 5	0.458	0.498	0.454	0.498	0.453	0.498
Grade 6	0.051	0.221	0.056	0.230	0.054	0.226
Grade 7	0.015	0.120	0.017	0.128	0.019	0.137
Grade 8	0.013	0.114	0.015	0.121	0.018	0.135

Table 2
Descriptive Statistics for Hillsborough County Analysis

	2000-01		2001-02		2002-03	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
	Third Grade					
Average SAT-9 Reading	630.1	44.0	630.7	44.8	634.2	43.2
Average SAT-9 Math	623.7	38.8	626.2	39.3	630.6	39.0
Average SAT-9 Prior Reading	603.4	42.3	603.0	42.1	603.9	41.4
English Language Learner	0.115	0.319	0.122	0.327	0.132	0.338
Eligible for Free or Reduced Priced Lunch	0.502	0.500	0.509	0.500	0.499	0.500
Male	0.489	0.500	0.510	0.500	0.499	0.500
African-American	0.241	0.427	0.248	0.432	0.235	0.424
Hispanic	0.162	0.368	0.170	0.375	0.182	0.386
Asian	0.017	0.130	0.017	0.129	0.017	0.130
Indian	0.004	0.059	0.003	0.059	0.003	0.058
Multiple Race	0.041	0.199	0.044	0.205	0.050	0.219
White	0.536	0.499	0.518	0.500	0.511	0.500
	Grades Four through Eight					
Average SAT-9 Reading	651.2	41.4	653.5	40.6	652.7	40.1
Average SAT-9 Math	646.7	38.5	649.0	38.3	648.7	38.1
Average SAT-9 Prior Reading	633.9	43.6	637.1	43.7	637.4	43.7
English Language Learner	0.170	0.376	0.174	0.379	0.189	0.392
Eligible for Free or Reduced Priced Lunch	0.522	0.500	0.524	0.499	0.526	0.499
Male	0.503	0.500	0.500	0.500	0.509	0.500
African-American	0.241	0.428	0.242	0.428	0.242	0.428
Hispanic	0.207	0.405	0.214	0.410	0.228	0.420
Asian	0.020	0.141	0.022	0.146	0.021	0.142
Indian	0.003	0.056	0.004	0.061	0.003	0.057
Multiple Race	0.023	0.150	0.033	0.178	0.040	0.197
White	0.506	0.500	0.486	0.500	0.465	0.499
Grade 4	0.480	0.500	0.472	0.499	0.477	0.499
Grade 5	0.470	0.499	0.479	0.500	0.487	0.500
Grade 6	0.015	0.121	0.020	0.139	0.014	0.116
Grade 7	0.018	0.132	0.014	0.118	0.012	0.109
Grade 8	0.018	0.133	0.015	0.120	0.010	0.100

Note: Table uses characteristics of students in the analysis of reading scores that includes prior student performance percentile.

Table 3
Regression Results: Effect of Policy on Math and Reading Scores

	Statewide		Hillsborough County			
	Reading	Math	Reading		Math	
Grade 3 * Cohort 0203	2.638***	2.509***	3.663***	7.789***	2.781**	4.067**
	[0.430]	[0.419]	[1.221]	[1.576]	[1.220]	[1.913]
(Grade 3 * Cohort 0203) * 10 - 25 Percentile				3.114*		1.292
				[1.688]		[1.658]
(Grade 3 * Cohort 0203) * 26-50 Percentile				-1.526		0.0612
				[1.592]		[1.632]
(Grade 3 * Cohort 0203) * 51-75 Percentile				-3.548**		-1.214
				[1.571]		[1.862]
(Grade 3 * Cohort 0203) * 76-90 Percentile				-5.822***		-1.380
				[1.792]		[1.966]
(Grade 3 * Cohort 0203) * 91+ Percentile				-8.798***		-1.915
				[2.163]		[2.412]
School Fixed Effect	✓	✓	✓	✓	✓	✓
Grade Fixed Effect	✓	✓	✓	✓	✓	✓
Year Fixed Effect	✓	✓	✓	✓	✓	✓
Linear Time Trend by Grade	✓	✓	✓		✓	
Linear Time Trend by Grade & Percentile Group				✓		✓
Observations	15,907	15,906	120,398	100,554	120,409	100,560
R-squared	0.943	0.937	0.308	0.655	0.306	0.535
F-Test (p-values):						
Grade 3 * Cohort 0203 + ……. = 0						
(Grade 3 * Cohort 0203) * 10-25 Percentile				0.000		0.000
(Grade 3 * Cohort 0203) * 26-50 Percentile				0.000		0.005
(Grade 3 * Cohort 0203) * 51-75 Percentile				0.001		0.061
(Grade 3 * Cohort 0203) * 76-90 Percentile				0.168		0.082
(Grade 3 * Cohort 0203) * 91+ Percentile				0.589		0.322

Note: Statewide regressions include school-by-grade level data. Hillsborough County regressions utilize student-level data. Model estimated via OLS. All reported models also include controls for observed student demographic characteristics identified on Tables 2 and 3. A standard deviation for the test in each subject is about 45 points. Standard errors clustered by third grade school reported in brackets. *p<0.10 **p<0.05 ***p<0.01