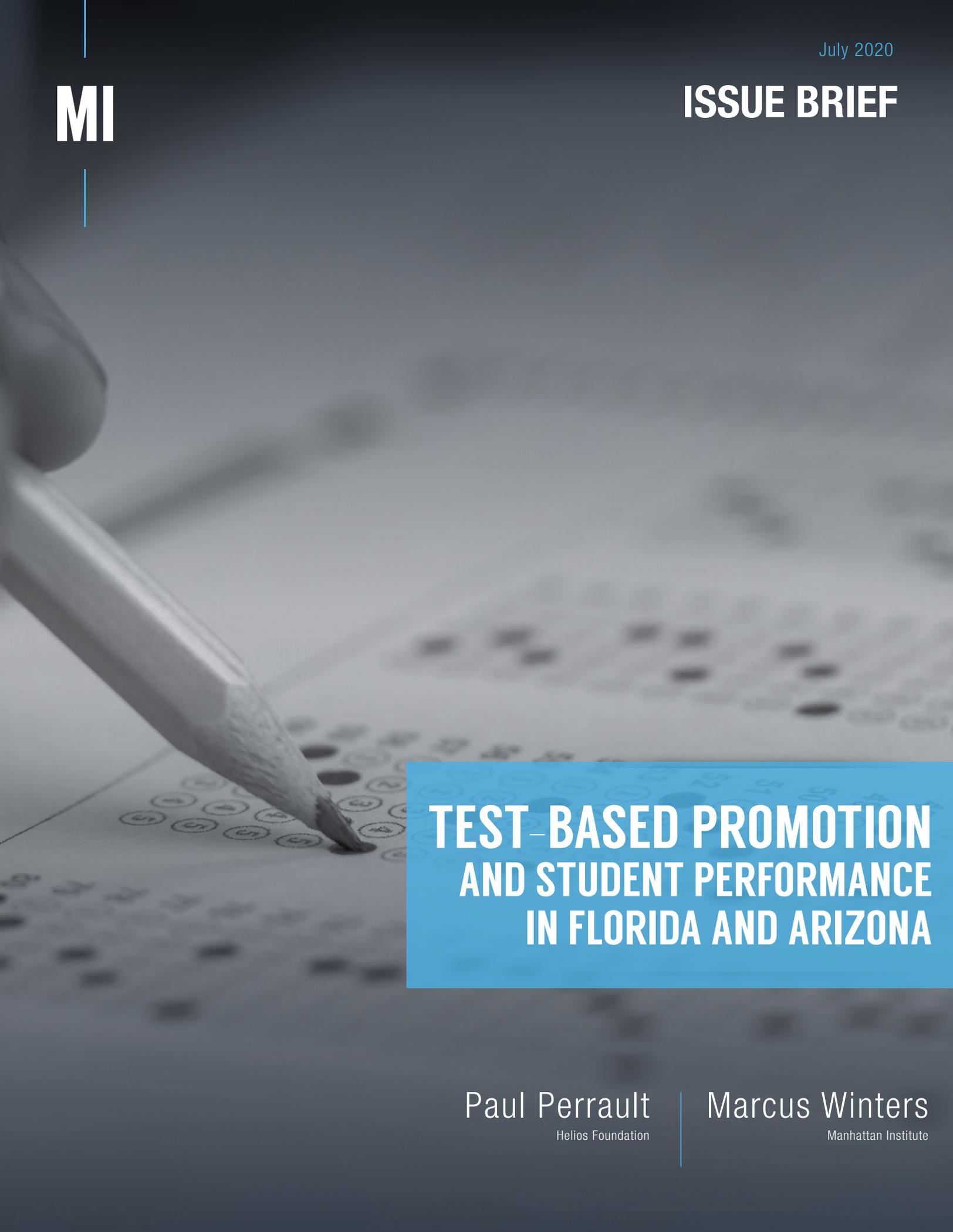


July 2020

MI

ISSUE BRIEF

A close-up, grayscale photograph of a pencil tip pointing at a test bubble sheet. The pencil is in the foreground, angled from the top left towards the center. The background is a blurred test sheet with rows of bubbles and some faint text. A blue rectangular box is overlaid on the right side of the image, containing the title text.

**TEST-BASED PROMOTION
AND STUDENT PERFORMANCE
IN FLORIDA AND ARIZONA**

Paul Perrault
Helios Foundation

Marcus Winters
Manhattan Institute

Contents

Introduction	3
Previous Research	3
Settings	4
Data	5
Methodology	5
Limitations	6
Results	6
Conclusion	7
Appendix.....	9
Endnotes.....	11



Introduction

Sixteen states and several urban school districts require students to score above a minimum threshold on standardized tests (always reading, and sometimes math) in order to be default-promoted past one or more “gateway grades” to the next grade. The large majority of these policies apply to students being promoted from the third to the fourth grade, on the theory that, by the third grade, students “stop learning to read and start reading to learn.”¹

Most studies of test-based promotion policies focus on measuring the effect of retention (being left back) on later student outcomes, and the evidence is fairly mixed.² However, test-based promotion policies do not only affect the students who are retained. Presumably, they also affect students and schools as they try to improve reading performance in order to avoid being retained or having to retain students. The threat of retention could plausibly have either positive or negative effects on students within the gateway grade. On the one hand, the pressure to score above a particular threshold on a standardized test might backfire by overwhelming both students and schools. On the other hand, test-based promotion policies might incentivize students and schools to make academic improvements within the targeted grade in order to avoid retention.

The effect of test-based promotion policies on student performance prior to the retention decision has not been studied enough. Filling this hole in the literature is important in order to understand the full impact that these policies have on students within a given jurisdiction. Even if relatively few students are actually retained under these policies, many students are in danger of scoring below the benchmark when they enter the gateway grade in the fall, and thus might be motivated by the policy to do better during the year than they would have otherwise. Therefore, even a small effect on students within the gateway grade could have a larger overall effect on student learning in a school system.³

We apply a difference-in-difference design to statewide longitudinal school-by-grade data from two states (Florida and Arizona) and to longitudinal student-level data from a large public school district (Hillsborough County, Florida) to investigate the effect of introducing a third-grade test-based promotion requirement on third-grade test scores. We measure whether there was an increase in third-grade test scores that occurred in the policy’s first year relative to scores in other grades within the school that were not directly targeted by the policy. We find evidence that enacting the policy led to a statistically significant and meaningful increase in average third-grade test scores in both states. The magnitude of the effect is very similar across the two states, despite their being enacted nearly a decade apart, as are the differences in the policy details.

We then use student-level data from public schools in Hillsborough County, where the school district independently tested second-grade students (even though statewide testing begins in the third grade), to evaluate whether the effects of the policy differed based on the student’s reading proficiency at the end of the second grade. We find that the effect of adopting the policy on third-grade test scores was similar for students regardless of their reading ability at the end of the second grade.

Previous Research

Our analysis adds to a handful of previous studies evaluating the impact that test-based promotion policies have on student performance prior to the retention decision. An evaluation of Chicago’s test-based promotion policy found that it led to

substantially improved student performance in both math and reading for students in grades three, six, and eight.⁴ In the third grade, there was a positive treatment effect in math across the distribution of previous performance—but it was largest for students most at risk of retention. But while third-grade students, on average, made substantial improvements on the reading exam, higher-performing students with little to no risk of falling below the policy threshold experienced substantial test-score declines. Another study found that the positive effect from Chicago’s policy was not apparent on tests other than those used to determine promotion.⁵ An evaluation of the fifth-grade test-based promotion policy in New York City compared the performance of entering fifth-graders in the first year of the policy to that of a matched comparison group comprising entering fifth-graders from the previous year. It 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 no effect in math.⁶

Our study also contributes to the more general literature on the effects of accountability policies on student performance. Though retention under test-based promotion policies is intended not as a punishment but rather as an educational intervention, these policies are rightly classed within a family of accountability policies that attach an undesirable consequence to the failure to meet a performance benchmark. Previous studies have found that such policies tend to increase average student achievement⁷ but that the effect often varies according to the student’s previous performance relative to the benchmark.⁸ Test-based promotion policies are perhaps especially interesting to consider within an accountability framework. They could engage the participation of students in a way that other accountability policies that only directly affect schools (for example, grading schools from A to F) do not.

Settings

Florida

Florida’s test-based promotion policy requires students to demonstrate basic reading proficiency before they are promoted to the fourth grade. Students scoring at 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% of students scoring below the threshold in the first year of the policy were nonetheless promoted.⁹ Still, the policy considerably increased the use of grade retention in the state: 2.8% of third-grade students were retained in the year prior to the policy’s implementation, compared with 13.5% in the first year of the policy.

Florida used its high-stakes criterion-referenced test, the FCAT, for accountability purposes in a way that is problematic for estimating the treatment effect from the test-based promotion policy.¹⁰ Fortunately, at this time the state also administered a commercial norm-referenced exam, the Stanford-9, in math and reading to all students in grades three through 10. Results on the Stanford-9 exam were used for informational purposes only and played no role in student or school accountability. For this reason, they were not likely to have been influenced by factors such as teaching-to-the-test or other manipulations. We thus rely on data for student scores on the Stanford-9 to estimate the treatment effect.

Arizona

The Arizona legislature adopted the Move On When Reading (MOWR) policy in 2010. The policy requires third-grade students to demonstrate a minimal level of reading proficiency by scoring above the threshold for “Falls Far Below” on the state’s annual standardized test in order to be default-promoted to the fourth grade. Students in the third grade in the 2013–14 school year were the first subjected to the policy.

MOWR was modeled in part on Florida’s test-based promotion policy. However, a potentially important difference is that

Arizona's policy set the performance standard that students were required to reach on the state's reading test at a much lower level than did Florida's policy, and the policy actually retained a much smaller percentage of the student population. Approximately 3% of Arizona's third-grade students scored in the Falls Far Below category in the policy's first year and thus were targeted for retention, and less than 1% of third-grade students were retained under the policy.¹¹

It is plausible that Arizona's significantly lower standard would reduce the potential effect of the policy. However, a recent study, based on interviews with teachers and administrators and observations of several third-grade classrooms in five Arizona school districts, found that districts and schools made intentional efforts to avoid student retention under the policy in ways that could lead to improvements in student performance before the retention decision.¹² Teachers described an increased awareness of the importance of building literacy foundations, and administrators reported that they allocated additional financial and curricular resources toward early-grade instruction. Districts reported that students as well as parents felt pressure from the policy to improve performance.

Data

We evaluate the impact of the test-based promotion policies statewide in Florida and Arizona using longitudinal school-by-grade-level test scores and demographic characteristics. For each statewide analysis, we use data from the first year that the policy was adopted (2002–03 in Florida and 2013–14 in Arizona) and the two prior years. Data from Florida are publicly available and were downloaded from the Florida Department of Education website. In Arizona, we acquired aggregated data from a data request to the Arizona Department of Education.¹³

To estimate effects in Hillsborough County, Florida, we use longitudinal student-level data for the universe of students in grades two through five from school years 1999–2000 through 2002–03. Student-level data are beneficial not only because of the increased precision but because Hillsborough is one of several districts in the state that administered the Stanford-9 exam district-wide in the second grade.

The availability of second-grade scores enables us to examine whether the effect of introducing the test-based promotion policy differed according to the student's reading ability when entering the third grade.

The estimation samples include all grades (or in Hillsborough, all students within grades) in schools that have valid test-score data in grades three, four, and five in an observed year. Results are similar if we also include grades six through eight, which are found in several K–8 schools.

Methodology

Our goal is to determine whether, in the first year that the policy was implemented, there was a significant change in the trajectory of third-grade test scores compared with the trajectory in other grades in the school. The intuition underlying this approach is that the test-based promotion policy would provide an incentive within the third grade (and perhaps earlier grades) but not in later grades, where students faced no danger of retention under the policy. Specifically, we employ a difference-in-difference design, where the first difference is across grades and the second difference is over time.

The unit of observation for the statewide analysis is a grade within a school. The dependent variable is the average math or reading score for students within that grade and year. The primary regression analysis includes fixed effects for each school, grade, and year, as well as our variable of interest, which equals 1 if the observation is of the third grade during the first treated year, and equals 0 otherwise (i.e., a treatment indicator). The coefficient on the treatment variable represents the differential change in third-grade test scores relative to fourth- and fifth-grade scores in the first year that the policy was in effect. We weight the regression according to the number of students who took the test within the school in a given year, and we cluster

the standard errors by school.

The data from Hillsborough County enable us to evaluate whether the treatment effect differs by the student's reading-test score in the prior year. We first designated six performance thresholds within the scores for the second-grade exam. Then we interact the treatment indicator variable (indicating that the observation is of a third-grade student during the treated year) with variables that indicate how the student performed on the prior year's reading exam.¹⁴ The coefficients on these interaction variables represent the effect of the treatment for students whose second-grade reading score was within a particular performance category.

Limitations

Lack of data for more than two years of pretreatment outcomes in any of the jurisdictions is the most important limitation of our analysis because it severely limits our ability to test an essential underlying assumption used to interpret the estimates. Interpreting the results as the causal effect of adopting test-based promotion on average third-grade scores assumes that the trajectory of average fourth- and fifth-grade scores after the introduction of the policy would have been the trajectory for average third-grade scores if the policy were not enacted.¹⁵ That assumption is intuitively plausible—there is no particular reason to believe that the trajectory of third-grade scores across either state prior to the adoption of the policy systematically differed from the trajectory of fourth- and fifth-grade scores. Nonetheless, the inability to observe these trends over a longer period means that we cannot confirm whether those grades were on similar trajectories before the introduction of the policy. Though our estimates remain policy-relevant and are at least suggestive of trends in outcomes across grades, readers should be cautious about interpreting the estimate as the causal effect of the policy.

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

Further, we are not able to measure the 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.¹⁶ 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. Finally, the state continued its focus on improving early-grade literacy during this time period, which may have had an effect on third-grade scores.

Results

Effect of Implementing Test-Based Promotion on Average Third-Grade Test Scores

We first consider the results from estimating the effect of the introduction of a third-grade test-based promotion policy on average math and reading scores. **Figure 1** reports the main statewide effect estimates using school-by-grade-level data from Arizona and Florida, as well as the student-level estimates using data from Hillsborough County. In each case, we find a statistically significant increase in third-grade scores in math and reading/ELA in the first year of the test-based policy, relative to the students in higher elementary grades who were not subjected to the policy.¹⁷

FIGURE 1

Estimated Effect of Implementing Test-Based Promotion Policy on Average Third-Grade Reading Scores

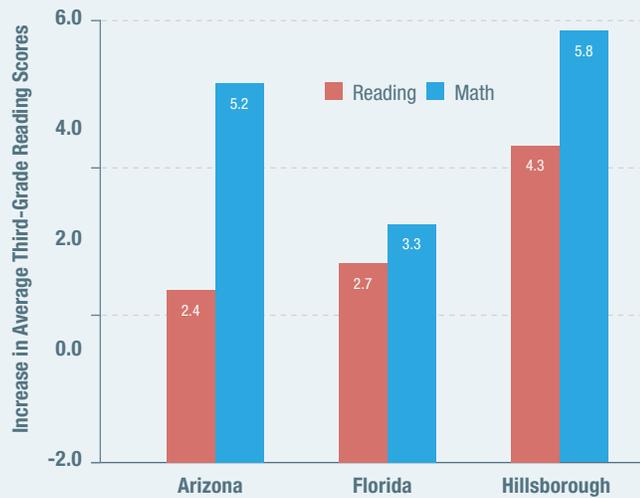
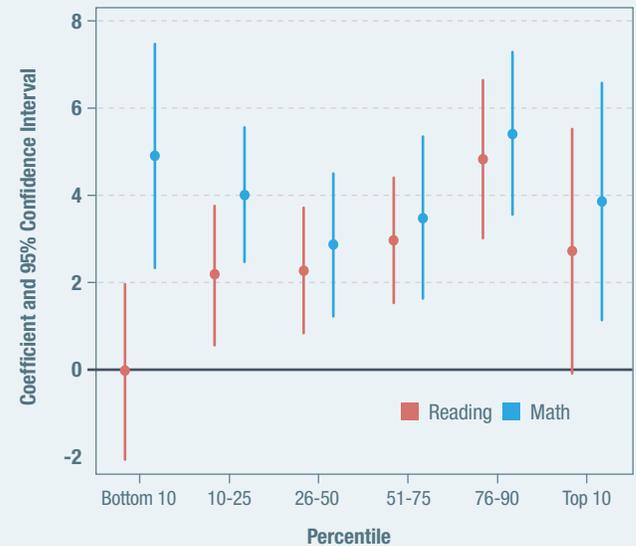


FIGURE 2

Impact of Test-Based Promotion Policy on Average Second-Grade Reading Score



Source (Figures 1 and 2): Author’s calculations based on data from the Arizona Department of Education and the Florida Department of Education.

The coefficients in the table represent the additional increase in average third-grade reading scores that occurred in the policy’s first year in scale-score points on the state test. After they are converted to comparable standard deviation units, the magnitude of the estimated treatment effect is similar across jurisdictions.¹⁸ The magnitude of the effect is “medium,” according to a recently proposed taxonomy.¹⁹ That the estimated treatment effect is similar in Hillsborough County and statewide in Florida is somewhat expected. However, it is reassuring that the effects in the Florida jurisdictions are similar to the estimated impact in Arizona, despite the policies having been implemented more than a decade apart.

Figure 2 illustrates the results from models that use the Hillsborough County data to measure the impact of the test-based promotion policy by the student’s second-grade reading score. The figure illustrates the coefficient and 95% confidence interval for each performance group in reading and math, respectively.

Overall, the results suggest that the implementation of Florida’s test-based promotion policy, at least in Hillsborough, had similar impacts on third-grade students regardless of their reading score in the previous year. In most cases, the estimated effects by percentile category are similar, and the differences are not statistically significant.

Conclusion

Discussions about whether to introduce test-based promotion policies often focus only on the students who are actually retained. While the impact of retention on those students is a first-order concern, far-reaching policies such as test-based promotion likely produce a broad set of effects beyond the main treatment, which should not be ignored. Our goal with this report is to give policymakers a broader scope when considering the impacts of existing and future test-based promotion policies, including the effects on students before they even take the decisive test.

We find evidence that introducing third-grade test-based promotion policies in Florida and Arizona led to statistically significant and meaningful average test-score improvements within the third grade before the policy retained any students. We supplement the statewide analyses with an evaluation of longitudinal student-level data from Hillsborough County, in which we find that the effect of the policy on third-grade students was similar for those who scored well or poorly on the second-grade reading test.

In addition to test-based promotion policies, our findings are relevant for the more general set of policies that incentivize better student outcomes by linking a consequence to the failure to meet a particular performance standard. Our estimates are quite consistent with previous research in this area.

Our findings effectively rule out the concern that test-based promotion policies have unintended negative impacts for students in the gateway grade. For instance, some have suggested that raising the stakes of performance via the threat of retention would backfire and actually reduce student outcomes. Our results show that such concern is unfounded. Indeed, rather than declines, in both Florida and Arizona, adoption of test-based promotion led to substantial improvements for students in gateway grades.

The evidence on the outcomes for students who are retained under a test-based promotion policy is currently mixed, although it is notable that several earlier studies have found benefits from retention under Florida's policy.

Our results, however, suggest that earlier studies, which focus entirely on retained students, substantially understate the benefits of test-based promotion policies on student achievement. The test-score improvements that we find within the third grade for students in Arizona and Florida apply to a much larger group of students than those who were eventually retained by the policies. Indeed, our results show that the threat of retention improves student academic achievement, thus reducing the need for retention.

Appendix

Placebo Tests

Our approach looks for a particularly large increase in third-grade test scores, relative to test scores in other grades, in the first year that the respective state implemented its test-based promotion policy.

Figure 3 presents results from a placebo test by grade level. It is possible that by comparing the third grade to the combination of students in fourth and fifth grades, the main analysis masks unexpected gains in the non-affected grades. For this test, we add to our primary regression an interaction between the treatment year and the fourth grade (Placebo Grade 4). Thus, we separately estimate the differential gain during the initial year of the policy in both the third (treated) grade and the fourth grade against the comparison group of the fifth grade. Though there are significant differences in fourth- and fifth-grade scores during the treated year, the gain in the third grade is significantly and substantially larger than in either of the comparison grades. Models that control for the grade-specific trend find no difference in the treatment-year gain between the untreated fourth and fifth grades, with the one exception of the reading test in Hillsborough. These results are consistent with applying a causal interpretation to the primary estimates.

FIGURE 3

Placebo Test, Grade

	Arizona		Florida		Hillsborough	
	Reading					
Treatment	2.211*** (0.326)	4.077*** (0.595)	1.799*** (0.260)	1.611*** (0.254)	3.465*** (0.603)	4.148*** (1.063)
Placebo Grade 4	-0.314 (0.298)	0.251 (0.648)	-1.871*** (0.188)	-0.765* (0.419)	-1.695*** (0.516)	2.877*** (0.836)
	Mathematics					
Treatment	3.862*** (0.454)	2.673*** (0.784)	4.013*** (0.220)	2.712*** (0.383)	7.254*** (0.715)	2.538* (1.314)
Placebo Grade 4	-2.611*** (0.433)	0.121 (0.871)	1.515*** (0.187)	0.227 (0.413)	2.829*** (0.560)	1.494 (0.927)
Observations	745,209	745,209	1,692,209	1,692,209	97,076	97,076
R-squared	0.866	0.866	0.940	0.940	0.612	0.612
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Trend by Grade	No	Yes	No	Yes	No	Yes
Student Demographics	No	No	No	No	Yes	Yes
Lag Test Scores	No	No	No	No	Yes	Yes

* = p < 0.10 ** = p < 0.05 *** = p < 0.01

Figure 4 reports the results from a placebo test evaluating differences in third-grade performance over time. In this model, we add to our primary regression an interaction between a variable indicating third grade and a variable indicating the year before policy introduction. The results from this test suggest that the increase in the third-grade scores that occurred in the first treated year is significantly and substantially larger than third-grade test scores in either of the observed years prior to the treatment.

FIGURE 4

Placebo Test, Year

	Arizona		Florida		Hillsborough	
	Reading					
Treatment	1.838*** (0.363)	1.586*** (0.405)	2.743*** (0.217)	1.615*** (0.254)	4.858*** (0.668)	3.247*** (0.712)
Placebo Year	-1.056*** (0.342)	-1.182*** (0.377)	0.062 (0.210)	-0.503** (0.236)	1.076* (0.570)	0.264 (0.577)
	Mathematics					
Treatment	6.029*** (0.494)	4.263*** (0.549)	3.471*** (0.229)	4.453*** (0.270)	7.239*** (0.751)	8.898*** (0.798)
Placebo Year	1.709*** (0.446)	0.825* (0.492)	0.435** (0.214)	0.928*** (0.238)	2.714*** (0.668)	3.551*** (0.660)
Observations (Weighted)	745,209	745,209	1,692,209	1,692,209	97,076	97,076
R-squared	0.866	0.866	0.939	0.940	0.612	0.612
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Trend by Grade	No	Yes	No	Yes	No	Yes
Student Demographics	No	No	No	No	Yes	Yes
Lag Test Scores	No	No	No	No	Yes	Yes

* = p < 0.10 ** = p < 0.05 *** = p < 0.01



Endnotes

- ¹ Sally Banks Zakariya, “Learning to Read, Reading to Learn: Why Third Grade Is a Pivotal Year for Mastering Literacy,” *Center for Public Education* (2015): 1–15.
- ² Brian A. Jacob and Lars Lefgren, “Remedial Education and Student Achievement: A Regression-Discontinuity Analysis,” *Review of Economics and Statistics* 86, no. 1 (February 2004): 226–44; Brian A. Jacob and Lars Lefgren, “The Effect of Grade Retention on High School Completion,” *American Economic Journal: Applied Economics* 1, no. 3 (July 2009): 33–58; Jennifer Sloan McCombs, Sheila Nataraj Kirby, and Louis T. Mariano, *Ending Social Promotion Without Leaving Children Behind: The Case of New York City* (Santa Monica, CA: Rand Corporation, 2009); Melissa Roderick and Jenny Nagaoka, “Retention Under Chicago’s High-Stakes Testing Program: Helpful, Harmful, or Harmless?” *Educational Evaluation and Policy Analysis* 27, no. 4 (Winter 2005): 309–40; Marcus A. Winters and Jay P. Greene, “The Medium-Run Effects of Florida’s Test-Based Promotion Policy,” *Education Finance and Policy* 7, no. 3 (Summer 2012): 305–30; Jay P. Greene and Marcus A. Winters, “Revisiting Grade Retention: An Evaluation of Florida’s Test-Based Promotion Policy,” *Education Finance and Policy* 2, no. 4 (Fall 2007): 319–40; Jay P. Greene and Marcus A. Winters, “The Effects of Exemptions to Florida’s Test-Based Promotion Policy: Who Is Retained? Who Benefits Academically?” *Economics of Education Review* 28, no. 1 (February 2009): 135–42; Guido Schwerdt, Martin R. West, and Marcus A. Winters, “The Effects of Test-Based Retention on Student Outcomes over Time: Regression Discontinuity Evidence from Florida,” *Journal of Public Economics* 152 (August 2017): 154–69; Umut Özek, “Hold Back to Move Forward? Early Grade Retention and Student Misbehavior,” *Education Finance and Policy* 10, no. 3 (2015): 350–77.
- ³ Eric R. Eide and Dan D. Goldhaber, “Grade Retention: What Are the Costs and Benefits?” *Journal of Education Finance* 31, No. 2 (Fall 2005): 195–214; Philip Babcock and Kelly Bedard, “The Wages of Failure: New Evidence on School Retention and Long-Run Outcomes,” *Education Finance and Policy* 6, no. 3 (Summer 2011): 293–322; Schwerdt, West, and Winters, “The Effects of Test-Based Retention on Student Outcomes over Time.”
- ⁴ Melissa Roderick, Brian A. Jacob, and Anthony S. Bryk, “The Impact of High-Stakes Testing in Chicago on Student Achievement in Promotional Gate Grades,” *Educational Evaluation and Policy Analysis* 24, no. 4 (December 2002): 333–57.
- ⁵ Brian A. Jacob, “Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools,” *Journal of Public Economics* 89, nos. 5–6 (June 2005): 761–96.
- ⁶ McCombs, Kirby, and Mariano, *Ending Social Promotion Without Leaving Children Behind*.
- ⁷ See, e.g., Thomas S. Dee and Brian Jacob, “The Impact of No Child Left Behind on Student Achievement,” *Journal of Policy Analysis and Management* 30, no. 3 (May 2011): 418–46.
- ⁸ Derek Neal and Diane Whitmore Schanzenbach, “Left Behind by Design: Proficiency Counts and Test-Based Accountability,” *Review of Economics and Statistics* 92, no. 2 (May 2010): 263–83; Randall Reback, “Teaching to the Rating: School Accountability and the Distribution of Student Achievement,” *Journal of Public Economics* 92, nos. 5–6 (June 2008): 1394–1415; Jacob, “Accountability, Incentives and Behavior.”
- ⁹ Greene and Winters, “The Effects of Exemptions to Florida’s Test-Based Promotion Policy.”
- ¹⁰ Beginning in 2003, the treatment year for the analysis, the state began using student growth on the FCAT in reading for school accountability purposes in grades four through eight. That policy change is problematic for our approach, which compares changes in the trend of the performance of third-grade students during the initial treatment year with 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.
- ¹¹ Lenay Dunn, Patricia Castillo, and Scott Vince, “Move On When Reading: Implementation of a Third-Grade Retention Policy in Arizona,” WestEd and Helios Education Foundation, 2016.
- ¹² *Ibid.*
- ¹³ Though student-level data are often preferred, they are not essential. In the statewide analyses, the primary benefit of student-level data is increased precision. But, as will be apparent when we report the results, the treatment effects for both states are estimated precisely enough to

detect even relatively small effects.

- ¹⁴ Large differences in the pretreatment levels for the dependent variable in each school system can be concerning if there is reason to believe that the level differences could affect trends (Ariella Kahn-Lang and Kevin Lang, “The Promise And Pitfalls of Differences-In-Differences: Reflections on 16 and Pregnant and Other Applications,” NBER working paper no. 24857: July 2018). In the case of our analysis, pretreatment differences are explained by the fact that we are comparing students in different grade levels, who naturally have scores that are on different levels on the test scales. Indeed, the fourth-graders within a school were, for the most part, third-graders in the school the year before. Though there might be differences in student cohorts within any given school, there is no reason to suspect that there are major differences in within-school cohorts from one year to the next across a state or large school district.
- ¹⁵ Causal interpretation relies on the assumption that there was no other change that influenced third-grade scores in the treated year relative to scores in other grades. In the fall of 2001, Florida’s 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 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 the adoption of these policies, the analyses below include a dummy variable controlling for whether the school was participating in Reading First that year. Excluding these variables has no meaningful impact on the estimated effect of the test-based promotion policy.
- ¹⁶ Marcus A. Winters, “Florida Defeats the Skeptics: Test Scores Show Genuine Progress in The Sunshine State,” *Education Next* 12, no. 4 (Fall 2012): 64–68.
- ¹⁷ Including a linear time trend by grade does not substantially change the interpretation of the estimates. The results are also robust to the inclusion of school-by-year fixed effects.
- ¹⁸ Converted to standard deviation units at the school level, the models estimate that the policy increased the average ELA test score in the third grade by about 0.135 standard deviations in Florida and about 0.091 standard deviations in Arizona. Converted into standard deviations according to the student-level distribution, the estimated effect within Arizona and Florida, respectively, was 0.04 and 0.05 in reading and 0.07 and 0.08 in math.
- ¹⁹ Matthew A. Kraft, “Interpreting Effect Sizes of Education Interventions,” *Educational Researcher* 49, no. 4 (April 2020): 241–53.

MANHATTAN
INSTITUTE