

The Effect of Retention Under Mississippi's Test-Based Promotion Policy

Kirsten Slungaard Mumma and Marcus A. Winters



Boston University Wheelock College of Education & Human Development
Wheelock Educational Policy Center



The Effect of Retention Under Mississippi's Test-Based Promotion Policy

Kirsten Slungaard Mumma, PhD* Marcus A. Winters, PhD†

Abstract

We apply a regression discontinuity design to investigate the effect of retention under Mississippi's third grade test-based promotion policy on student outcomes through the sixth grade. Retention led to large improvements in ELA scores, though we find no significant impacts in math. The test score impacts are driven by Black and Hispanic students. Retention did not significantly impact attendance rate or the likelihood that a student is later classified as having a disability.

Acknowledgements: We extend our gratitude to the Mississippi Department of Education for their cooperation and support for this project. This project was made possible by a grant from *ExcelinEd*. Neither the Department of Education nor *ExcelinEd* had editorial control over the content in this manuscript. All remaining errors are our own.

*Kirsten Slungaard Mumma is a postdoctoral fellow at the Wheelock Educational Policy Center.

†Marcus A. Winters is an associate professor at the Boston University Wheelock College of Education & Human Development and faculty director of the Wheelock Educational Policy Center.

1 Introduction

Motivated by a need to improve early literacy, 17 states currently require students who score below a minimum threshold on a standardized test be retained in the third grade, where they also receive additional interventions (Cummings and Turner, 2020). The impacts of treatment under such “test-based promotion” policies in early grades tend to be positive, but vary somewhat by locality. Retention under Florida’s third-grade policy increased student test scores (Schwerdt et al., 2017; Winters and Greene, 2012; Greene and Winters, 2007, 2009), but also disciplinary incidences in the short-run (Özek, 2015). Treatment in the fourth grade under Louisiana’s policy had no impact on high school graduation probability (Eren et al., 2017). Evidence from Chicago suggests that treatment under the city’s fourth-grade policy had an initial positive effect (Jacob and Lefgren, 2009) that dissipated by the second year following retention (Roderick and Nagaoka, 2005). New York City’s fourth grade policy led to improvements in student test scores (Mariano and Martorell, 2013) and had no impact on behavioral outcomes (Martorell and Mariano, 2018).

We add to this line of research evidence on the impact of retention in the third grade under Mississippi’s test-based promotion policy. In 2013, Mississippi adopted test-based promotion as part of a comprehensive effort to improve early literacy outcomes that also included a program to support collaborations between public and private pre-K providers and targeted supports to teachers to build skills in the science of reading (Burk 2020). The state has since received considerable attention for its substantial improvements in student reading outcomes. Between 2013 and 2019, average fourth grade reading scores on the NAEP increased by 10 points in Mississippi, more than any other state, while the national average declined by a point.

We apply a fuzzy regression discontinuity (RD) design leveraging the change in the likelihood of retention at the passing threshold on the state’s third grade ELA test for the first cohort of students subjected to the policy. Retained students scored more than a standard deviation higher relative to their grade-level peers in the sixth grade than if they had been promoted, and this impact is driven by Black and Hispanic students. We do not

find effects in math. Retention did not impact student absences or the likelihood of being classified as having a disability in later years.

2 Data and Method

We analyze longitudinal administrative data containing standardized test scores, demographics, classification status, and number of days absent for each student from 2014-15 through 2018-19 provided by the Mississippi Department of Education. We thus can follow the typical student from the first cohort subjected to the policy who was retained in the third grade as far as the sixth grade. See Table A1 in the Online Appendix for relevant descriptive statistics.

Mississippi’s test-based promotion policy required third grade students to score Level 2 (second lowest level) or above on the state’s standardized reading test in order to be automatically promoted to the fourth grade. For the 2014-15 school year, the test was the Mississippi K-3 Assessment System (MKAS), and we use scale scores on this test for the first-stage in the method described below. The following year, the statewide assessment changed to the Mississippi Academic Assessment Program (MAAP). When considering test scores as outcomes, we use MAAP scores standardized by grade and year.

As is common for such policies, retention is not the only intervention delivered to treated students. In addition to repeating the grade, schools are required to provide retained students with 90-minutes of reading instruction and intensive interventions with progress monitoring and other supports. Thus, it should be kept in mind that the estimates in this paper reflect the overall treatment under the policy, not only retention.

2.1 Within-Age vs Within-Grade Comparisons

Comparing the later outcomes of students retained at a point in time to students in their cohort who were promoted is complicated by the fact that the two groups are enrolled in different grade levels during later years. One could choose a “within-age” approach by

comparing the groups after a particular amount of time or a “within-grade” approach by comparing them when enrolled in a particular grade. Prior studies have applied both approaches. Interpretation under both strategies requires additional assumptions and neither is strictly preferable when investigating the impact of retention. Within-age comparisons can be confounded by differences in learning and behavioral trajectories across grades, while within-grade comparisons are complicated by age differences when the students complete a particular grade (Schwerdt et al., 2017).

The preferred comparison also depends on how one thinks of the treatment. Policy-induced retention is a somewhat unique intervention in that the additional year of schooling itself is arguably an important component of the treatment. For example, if we were to consider the effect of policy-induced retention in the third grade on long-run student academic performance, it is arguably more policy-relevant to compare the difference in student proficiency at the point at which they graduate rather than nine years following their initial entry into the third grade.

When evaluating standardized test scores as the outcome, within-age comparisons are further complicated by the fact that retained and promoted students take different grade-level tests at any given point in time. In some cases, authors can address this challenge by utilizing test scores reported on a vertically aligned scale (for example, Schwerdt et al. (2017)).

Unfortunately, within-age comparisons of student test scores are not possible in Mississippi because scores on the state’s standardized tests are comparable within grades over time but not across grades. Non-test score outcomes are not impacted by this measurement issue and so we apply both the within-age and within-grade approaches when considering them.

2.2 Identification Strategy

The primary challenge with estimating the causal effect of retention induced by Mississippi’s test-based promotion policy is that there are likely unobserved characteristics that are cor-

related both with a student’s later outcomes and the probability that they were treated under the policy. We address this challenge by employing a fuzzy RD design that exploits the discontinuous relationship at the policy threshold between a student’s third-grade ELA score and the probability they are retained.

We estimate a two-stage least-squares regression within a sample restricted to include only third grade students who scored within 20 scale points of the Level 2 threshold on the MKAS ELA test. The first stage uses a vector of observed baseline characteristics (X), the difference between the student’s score and the passing threshold (dif), and an indicator for whether the student’s score on the initial third grade ELA test fell below the policy threshold ($Below$) to predict the likelihood a student repeated the third grade the following year ($Retained$). The second stage then uses these controls but replaces $Below$ with the predicted retention from the first stage ($\hat{Retained}_i$) to predict the respective outcome in the 6th grade (y_{i6}). Formally:

$$Retained_i = \alpha_0 + \alpha_1 dif_i + \alpha_2 X_i + \alpha_{FS} Below_i + \mu_i \tag{1}$$

$$y_{ig} = \beta_0 + \beta_1 dif_i + \beta_2 X_i + \beta_{IV} \hat{Retained}_i + \epsilon_{ig} \tag{2}$$

When investigating non-test-score outcomes, we also estimate within-age regressions that instead measure the dependent variable as of 2018-19, when the typical retained student is in the sixth grade and the typical promoted student is in the seventh grade.

The coefficient β_{IV} represents the Local Average Treatment Effect (LATE) for the impact of retention due to the policy on the respective outcome. The key identifying assumptions are that 1) Conditional on the covariates there is a significant relationship between where a student’s score fell relative to the policy threshold and the probability they were retained, and 2) The only reason that the conditional relationship between scoring below the threshold and retention probability exists is that scoring below the threshold triggers the policy.

Figure 1 speaks to the first key assumption by illustrating the relationship between third grade ELA scores and retention probability. Notice that most students who scored below the policy threshold were not retained because they received one of the several exemptions under the policy. Indeed, within this cohort half of the students who scored below the passing threshold on the first attempt passed on their second or third attempt. Nevertheless, we do observe a discontinuous jump in the probability of being retained on either side of the cut score. This relationship is also reflected in the results from the first-stage regressions reported below.

[FIGURE 1 ABOUT HERE]

We investigate the plausibility that the second key assumption holds by evaluating whether the observed covariates are balanced on either side of the threshold, conditional on *dif*. Consistent with this expectation, Table A2 in the Online Appendix shows no discontinuities in student characteristics around the cut score.

3 Results

Table 1 reports our results for each outcome from the full sample and for samples restricted by a student's race/ethnicity. The table includes estimates from the relevant first stage and reduced form in addition to the causal instrumental-variable estimate for the effect of retention. See Tables A2 and A3 in the Online Appendix for estimates from models that use alternative specifications for the forcing variable and bandwidths. The magnitude and direction of the estimates are robust to multiple specifications, though some models are estimated less precisely.

[TABLE 1 ABOUT HERE]

Column (1) reports results for ELA test scores. For the full sample, students retained under the policy scored about 1.15 standard deviations higher on the ELA test in the sixth

grade than they would if they had instead been promoted. This result is estimated imprecisely and is significant only at the 10% level. The results from the analyses by race/ethnicity suggest that the overall effect is primarily driven by impacts on Black and Hispanic students.

The remaining columns report results for other outcomes of interest. We find no significant impact of retention on student math scores, absence rate, or the likelihood that a student was classified as having a disability.

The variation by student race/ethnicity in the first stage estimates is interesting to consider from a policy perspective. Scoring below the threshold on the third grade ELA test increased the likelihood that a Black student, by far the largest subgroup, was retained by only 3.7 percentage points. For Hispanic students, scoring below the threshold increased the likelihood of retention in the third grade by about 19 percentage points. Notably, both Black and Hispanic students benefited substantially in ELA if they were retained.

4 Implications and Future Directions

Our results are generally promising for the effects of test-based retention as implemented in Mississippi. We find large positive impacts from retention on student ELA achievement, which is the policy's primary goal. That we fail to find impacts on SPED classification status or absences suggests that retention did not have lasting negative impacts on the students' experiences in school, as some fear.

There are, however, some notable differences between our findings for Mississippi and prior evidence from other localities that are worthy of future consideration. Most importantly, though like most other states Mississippi's policy is targeted towards improving student reading proficiency, the fact that we do not find impacts of repeating a grade on student math scores is at least somewhat concerning and is inconsistent with findings from other localities.

Further, differences in the implementation of the retention treatment between Mississippi and other states is a notable distinction with implications for future research. For

example, Schwerdt et al. (2017) report that for the first cohort of third grade students subjected to Florida's policy, scoring below the threshold increased the probability a student was retained by 37.3 percentage points. In contrast, we find that scoring below the threshold increased the probability of retention by only 5.8 percentage points in Mississippi. Much of this large difference is due to the fact that Florida set a very high standard for students to obtain an exemption from the treatment by passing an alternative test than did Mississippi. Since our estimates can only be interpreted as LATEs, we are not able to assess whether students who obtained an exemption would have benefited if they were instead retained.

From a policy perspective, there is a clear need to update the analysis in this paper in future years in order to investigate the potential for the ELA impact to fade over time and also to consider other outcomes such as educational attainment. We also look forward to future research investigating impacts on later student cohorts. In particular, an interesting feature of Mississippi's policy is that the state raised the passing threshold on the test beginning with the 2018-19 cohort of third grade students. It will be interesting to consider in the future whether this change had implications for the impact of retention under this policy.

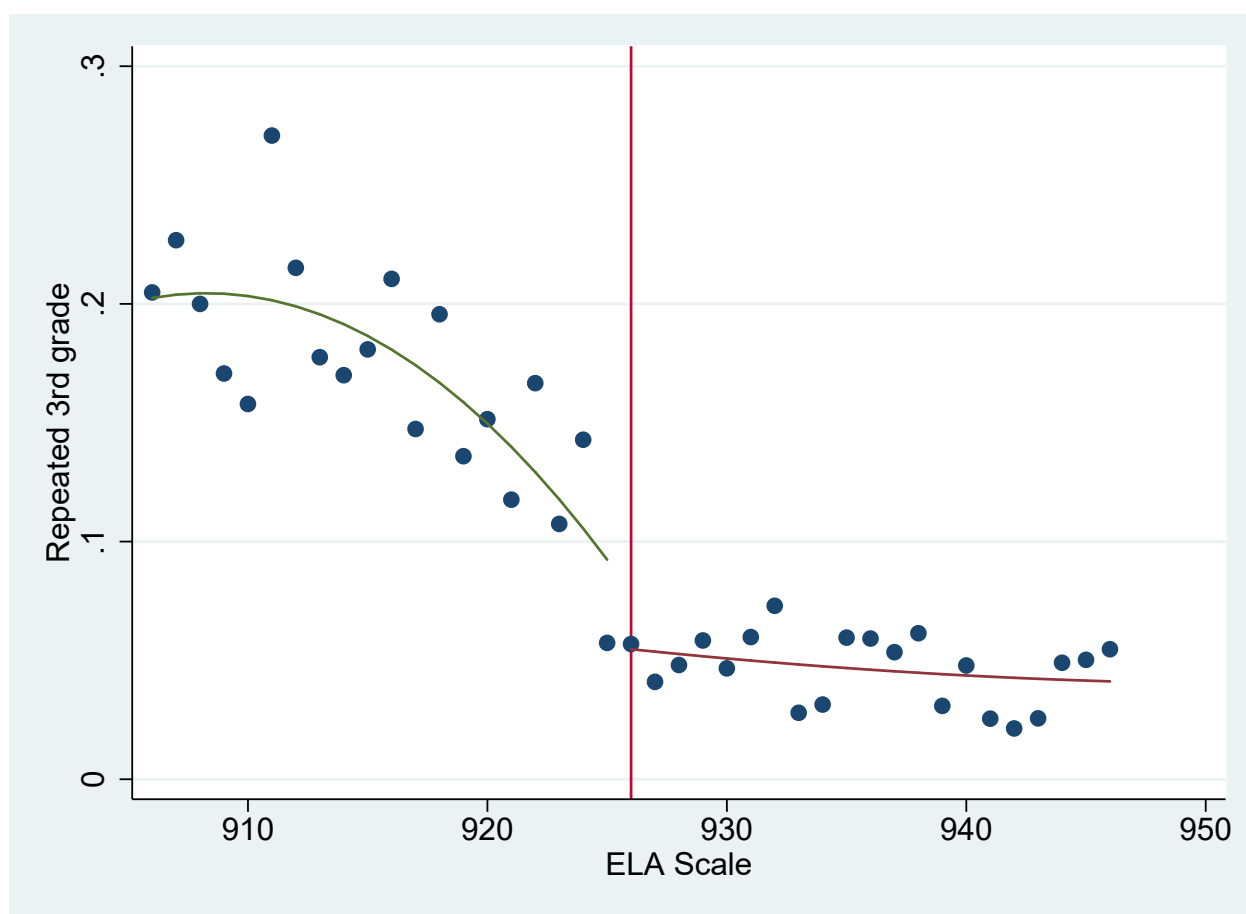
References

- Cummings, A. and Turner, M. (2020). Covid-19 and third-grade reading policies.
- Eren, O., Depew, B., and Barnes, S. (2017). Test-based promotion policies, dropping out, and juvenile crime. *Journal of Public Economics*, 153:9–31.
- Greene, J. P. and 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. and Winters, M. A. (2009). The effects of exemptions to Florida’s test-based promotion policy: Who is retained? Who benefits academically? *Economics of Education Review*, 28(1):135–142.
- Jacob, B. A. and Lefgren, L. (2009). The effect of grade retention on high school completion. *American Economic Journal: Applied Economics*, 1(3):33–58.
- Mariano, L. T. and Martorell, P. (2013). The academic effects of summer instruction and retention in New York City. *Educational Evaluation and Policy Analysis*, 35(1):96–117.
- Martorell, P. and Mariano, L. T. (2018). The causal effects of grade retention on behavioral outcomes. *Journal of Research on Educational Effectiveness*, 11(2):192–216.
- Özek, U. (2015). Hold back to move forward? Early grade retention and student misbehavior. *Education Finance and Policy*, 10(3):350–377.
- Roderick, M. and 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., West, M. R., and Winters, M. A. (2017). The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida. *Journal of Public Economics*, 152:154–169.

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

Tables and Figures

Figure (1) Relationship Between 3rd Grade ELA Score and Retention Probability



Notes: This figure illustrates the relationship between scores on the 3rd grade administration of the MKAS ELA test in 2014-15 and the probability a student was observed in the 3rd grade the following year. Dots represent average outcome for students who obtained a particular score on the test. Vertical line illustrates the passing threshold – observations to the left of the line did not meet the policy’s promotion requirement. Lines going through the dots represent flexible polynomial fits. Figure illustrates only scores that fell within 20 points of the passing threshold.

Table (1) Regression Results

	<i>Grade 6</i>				<i>Year 4</i>	
	(1) ELA	(2) Math	(3) Absences	(4) Sped	(5) Absences	(6) Sped
	<i>Full Sample</i>					
First Stage	0.058*** (0.016)				0.067*** (0.017)	
Reduced Form	0.067* (0.035)	-0.025 (0.042)	-0.115 (0.546)	-0.016 (0.016)	0.505 (0.579)	-0.024 (0.018)
Retained (IV)	1.153* (0.657)	-0.436 (0.743)	-1.983 (9.433)	-0.284 (0.292)	7.512 (8.820)	-0.361 (0.280)
Average Outcome	-.84	-.73	9.61	.20	10.41	.21
N	4729	4719	4729	4729	4562	4562
	<i>Black</i>					
First Stage	0.037** (0.019)				0.045** (0.020)	
Reduced Form	0.106*** (0.040)	-0.017 (0.047)	-0.021 (0.646)	-0.018 (0.017)	0.580 (0.664)	-0.024 (0.019)
Retained (IV)	2.842 (1.731)	-0.464 (1.295)	-0.574 (17.321)	-0.485 (0.525)	12.925 (15.921)	-0.544 (0.499)
Average Outcome	-.90	-.83	9.17	.15	9.92	.15
N	3382	3376	3382	3382	3253	3253
	<i>Hispanic</i>					
First Stage	0.186** (0.077)				0.230*** (0.084)	
Reduced Form	0.379** (0.186)	0.071 (0.225)	-0.704 (2.147)	-0.076 (0.060)	2.722 (2.522)	-0.081 (0.054)
Retained (IV)	2.042 (1.250)	0.380 (1.194)	-3.793 (11.492)	-0.408 (0.368)	11.832 (10.683)	-0.350 (0.269)
Average Outcome	-.57	-.33	8.1	.12	8.85	.12
N	206	205	206	206	204	204
	<i>White</i>					
First Stage	0.098*** (0.036)				0.103*** (0.039)	
Reduced Form	-0.066 (0.077)	-0.093 (0.099)	0.074 (1.190)	-0.001 (0.045)	0.429 (1.329)	-0.007 (0.047)
Retained (IV)	-0.678 (0.869)	-0.940 (1.084)	0.760 (12.135)	-0.014 (0.455)	4.156 (12.964)	-0.066 (0.455)
Average Outcome	-.69	-.54	11.33	.39	12.21	.39
N	1042	1039	1042	1042	998	998

Note: Table reports first-stage, reduced-form, and IV estimates for the full sample and samples restricted by race/ethnicity. Dependent variable for first-stage regressions is an indicator for whether the student was retained in the third grade; dependent variables for reduced-form and IV regressions are listed at top of each column. Columns (1) - (4) report within-grade comparisons; Columns (5) and (6) report within-age comparisons. All samples are restricted to only students who scored within 20 points of the Level 2 threshold on the third grade MKAS ELA test in 2014-15. All regressions control for original third grade ELA score, gender, and special education and Limited English proficiency status as of the initial third grade year. Full sample regressions also control for student race/ethnicity. See Online Appendix for results from models that apply alternative bandwidths and specification for the forcing variable. Robust standard errors reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Suggested Citation: Slungaard Mumma, K., & Winters, M. A. (2023). *The Effect of Retention Under Mississippi's Test-Based Promotion Policy*. (Working Paper 2023-1). Wheelock Educational Policy Center. Available at wheelockpolicycenter.org.

OUR MISSION

The Wheelock Educational Policy Center (WEPC) conducts and disseminates rigorous, policy-relevant education research in partnership with local, state, and federal policymakers and stakeholders to improve pk-20 educational opportunities and holistic outcomes for underserved students.

www.wheelockpolicycenter.org
wheelockpolicy@bu.edu



Boston University Wheelock College of Education & Human Development
Wheelock Educational Policy Center

