WORKING PAPER 2026-4
FALL 2026

Student Responses to the Credible Possibility of Grade Retention

Marcus A. Winters, Ph.D., Boston University

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

WEPC
WHEELOCK EDUCATIONAL POLICY CENTER

ABSTRACT

Test-based promotion policies may affect achievement not only through grade retention, but also through behavioral responses to the possibility of retention. Using a sharp regression discontinuity design, I estimate the effect of failing early administrations of Indiana's third-grade literacy exam that signal retention risk but do not determine grade placement. Failing these exams leads to immediate improvements in achievement, including gains observed within four to six weeks on distinct summative assessments. The results provide causal evidence that salient retention threat can generate rapid learning gains, even when the probability of eventual retention is low.

ACKNOWLEDGMENTS

I thank the Indiana Department of Education for providing access to their data and offering valuable feedback throughout the research process. I received helpful comments from Cory Koedel, Kevin Lang, Joshua Goodman, Eric Taylor, Jesse Bruhn, and John Fallon, as well as seminar participants at Boston University's Applied Microeconomics Seminar and the Northeastern Economics of Education Workshop. The research-practice partnership under which this research was conducted benefits considerably from generous financial support from the William T. Grant Foundation and administrative support from Sidrah Baloch and Meagan Comb. All remaining errors are my own. The opinions expressed are those of the authors and do not represent views of the Indiana Department of Education.

INTRODUCTION

Test-based promotion policies are widely used to ensure that students have sufficient literacy skills to access instruction before they are promoted past a key gateway grade. Under these policies, students must demonstrate proficiency on a standardized assessment to advance to the next grade. A substantial empirical literature examines the effects of grade retention under such policies and generally finds that retention can raise test scores in the short run, though longer-run effects are mixed and vary across settings (Hwang and Koedel, 2025; Schwerdt et al., 2017; Jacob and Lefgren, 2004; Eren et al., 2017; Mariano and Martorell, 2013; Slungaard Mumma and Winters, 2023; Berne et al., 2026).

Much less is known, however, about how students, parents, and schools respond to the possibility of retention. The prospect that a student could repeat a grade if they do not acquire the skills necessary to pass the gateway test may influence behavior by increasing effort, reallocating instructional resources, or intensifying parental engagement. Under these policies, students have multiple opportunities to pass the gateway test, creating periods in which students who pass know they will be promoted while those who fail face a salient retention threat if they do not develop sufficient skills in time to pass their final attempt. Despite their prevalence in policy design, there is little causal evidence on whether such perceived stakes affect student achievement.[1]

This paper isolates the incentive margin of a test-based promotion policy. I study Indiana's third-grade literacy policy, which requires students to pass a statewide reading assessment (IREAD) before the summer following third grade in order to be promoted to the fourth grade. Students who pass earlier administrations of the test in the second grade or the spring of third grade will be promoted, while those who fail an early administration receive a clear signal that they are at risk of retention. I use a sharp regression discontinuity (RD) design to measure the effect of failing each of these two distinct administrations of the IREAD that do not directly determine retention on student third grade scores on the state's summative standardized English Language Arts (ELA) and math tests. This design isolates behavioral responses to perceived stakes rather than instructional or peer effects associated with repeating a grade.

The results show that failing these earlier IREAD administrations leads to measurable improvements in student learning. Failing the second grade administration of IREAD increased third grade ELA scores by about 0.063σ. Failing the spring of third grade administration, the penultimate attempt before triggering the policy, increased scores in both ELA (0.039σ) and math (0.020σ) on the state's summative standardized test administered just four to six weeks later. To help put the magnitudes into context, typical student reading growth between third and fourth grade is about 0.36σ (Hill et al., 2008).

The short time-horizon over which the effect of failing the spring of third grade administration is a unique feature of the analysis that highlights the responsiveness of student achievement to salient performance thresholds. This result suggests considerable potential for educational interventions to rapidly improve learning for low-performing students. The closest analogue in the literature is Aucejo et al. (2022), who show that a policy of student retesting as part of the formal teacher evaluation system induced short-window increases in teacher effort that improved student performance within a similar time-frame.

Failing the second grade administration of the IREAD produced benefits, though students scoring at the threshold likely would have eventually passed the gateway test without intervention. The findings show that perceived stakes, not just high objective treatment risk, can generate behaviors that produce learning gains. The pattern is consistent with reference-dependent behavior in which being "below the cut" frames outcomes as losses and amplifies effort (Kahneman and Tversky, 1979; Kőszegi and Rabin, 2006), and also with salience/attention models where vivid thresholds draw disproportionate focus (Bordalo et al., 2012).

The results have implications for the design of early literacy policies and accountability systems more broadly. While retention decisions carry substantial instructional and social consequences, the findings here suggest that earlier, salient signals—combined with timely instructional support—may generate achievement gains for a broader set of students, even in the absence of actual retention.

SETTING AND DATA

Indiana's policy requires students to demonstrate a minimal level of literacy on the state's gateway reading test—the IREAD—before promotion to the fourth grade. Students have multiple attempts to pass the test. Students with disabilities and English learners take the IREAD but are exempt from the policy and thus I exclude them from the analyses. Eligible students who do not pass by the end of summer following the third grade are designated for retention. Tables A.1 and A.2 report summary statistics by IREAD administration.

Since 2011–12, all third grade students take the spring IREAD in early- to mid-March. IREAD is an untimed, computer-adaptive test. Both the test's conceptual design and its passing score has remained constant. Schools have immediate access to preliminary results, and final student scores are available after quality control checks. Schools typically notify parents within two weeks if their child officially failed the spring IREAD.

Students who pass the spring IREAD are promoted to the fourth grade, are not required to attend summer school, and do not retake subsequent administrations of the IREAD. All eligible students who fail the spring IREAD must attend summer school, after which they retake the test. Students who passed the summer IREAD should have been promoted to the fourth grade, and those who failed should have repeated the third grade.[2]

In the 2021–22 school year, the state began allowing schools to voluntarily administer the IREAD in the second grade. Students who pass the IREAD in the second grade do not take subsequent administrations of the test. Of the state's nearly 1,500 elementary schools, 372 participated in the spring of 2022, and 771 participated in spring of 2023. Because identification within my RD design relies on comparisons among students within schools that voluntarily administer the second-grade IREAD, school-level participation decisions do not affect the internal validity of the estimates.

Importantly for our purposes, the IREAD is distinct from the state's annual summative tests taken by all students in grades three through eight in ELA and mathematics. The state administers summative tests annually between mid-April and early-May, about four to six weeks after third grade students take the spring IREAD. The IREAD assessment is a statewide, standards-aligned test designed to measure whether students have attained foundational reading skills by

the end of third grade, focusing on basic decoding, vocabulary, and comprehension (Indiana Department of Education, 2024). By contrast, the statewide English Language Arts test assesses a broader set of literacy skills—such as higher-order text comprehension, communication, and writing—aligned to the full set of Indiana Academic Standards for ELA across multiple content domains (Indiana Department of Education, 2021). These differences in test design and skill coverage mean that IREAD and the ELA assessment capture related but not identical aspects of students' literacy development.

My primary analyses examine third grade spring summative ELA and math scores for all Indiana public school students. With the exception of 2020 when the state did not administer standardized tests, I analyze the effect of failing the spring of third grade test on summative third grade test scores from 2012 through 2024, and the effect of failing the second grade IREAD administration on third grade test scores in 2023 and 2024. Students took the ISTEP+ until the state changed to the ILEARN in 2019. I standardize third grade scores on the ILEARN and ISTEP+ by year.

The test-based promotion policy was one of several reforms Indiana adopted during this period to improve early literacy. These additional initiatives would affect the validity of the regression discontinuity analyses only if they differentially influenced outcomes for students scoring just above versus just below the IREAD passing threshold. While such reforms may have shifted the overall distribution of reading achievement in the state, the passing threshold remained fixed over time. As a result, the analysis consistently identifies effects for students with comparable literacy skills at the cutoff.

My analysis is motivated by the idea that parents, students, and schools might respond to the possibility of treatment even before the treatment goes into effect. However, students who scored near the passing threshold may face little practical risk of failing the test the following year, even without further intervention. Figure 1 illustrates that the likelihood a student failed the summer IREAD declines sharply with prior IREAD score. About 10% of students who scored 25 points below the threshold on IREAD in Grade 2, and about 35% of those scoring 25 points below the threshold on the spring Grade 3 IREAD, also failed the final summer retest.

I also investigate the potential that the treatment effect could be related to the quality of the student's school. For example, we might expect that higher quality schools might be more effective at providing targeted supports to push students flagged for possible retention under the policy above the passing threshold. I use test scores in grades three through five from 2007–2011—prior to the state's adoption of the test-based promotion policy—to measure district and school value-added impacts on student test scores. I estimate simple value-added models using school/district fixed effects and the student's lagged test score to predict test scores. Formally:

$$y_{idt} = \alpha + \phi y_{idt-1} + \theta_d + \mu \quad (1)$$

The estimate for $\hat{\theta}_d$ gives the average test score gain made by students attending each district relative to other districts in the state. I also run the same model and capture school fixed effects, $\hat{\theta}_s$. I use these captured fixed-effects as a measure of persistent district/school quality, which requires the assumption that district/school value-added from 2007–2011 is a reliable proxy for its quality from 2012–2024.

IDENTIFICATION STRATEGY

My goal is to estimate the causal effect of failing a respective administration of the IREAD on a student's summative test scores in the spring of their initial third grade year. The primary challenge with identification is that students who fail the IREAD may differ from those who passed in unobserved ways that are also correlated with later outcomes. To address this challenge I employ a sharp RD design that leverages plausibly exogenous variation in the possibility that a student is retained under the policy at the passing threshold on the respective IREAD administration.

Using either the second grade or spring third grade IREAD as the forcing variable, I estimate sharp RD regressions taking the form:

$$y_i = \alpha + \gamma X_i + f(a_i) + \beta \mathbb{1}(a_i < 0) + \epsilon_i \quad (2)$$

where y_i is the student's third grade outcome; X is a vector of observed baseline covariates; $f(a)$ is a function of the distance in scale scores from the passing threshold; $\mathbb{1}(a < 0)$ indicates that the student failed the test; and ϵ is a stochastic term.

Regressions only include non-exempt students with an IREAD score that falls within the MSE-optimal bandwidth on each side of the cutoff for the respective outcome (Calonico et al., 2017). I rebuild the bandwidth for each regression. Primary models use a uniform kernel and treat the relationship between the forcing variable and the outcome as a linear function allowed to vary on either side of the threshold. The estimates are not sensitive to these choices.

The key identifying assumption is that absent the policy the relationship between IREAD score and later outcomes would have been smooth at the cutoff. The most likely way that this assumption could be violated is if schools or students could manipulate scores in order to boost passing rates for those who scored just below the threshold. But the nature of computer-administered standardized testing makes it unlikely that scores could be systematically manipulated. Consistent with this expectation, the results reported in Table 1 find that the value of baseline characteristics generally balance at the threshold. In addition, density tests illustrated on Figure A.1 of the Online Appendix find no evidence that scores were manipulated at the passing threshold.

The primary outcome I consider is the student's scores on the state's summative test administered in April or May of their initial third grade year. Though I observe several student cohorts for subsequent years, spring of the student's initial third grade attempt is the only period in which I observe students prior to some of them being exposed to the bundle of treatments under the policy. Measuring the effect of retention under Indiana's policy is complicated by the use of various exemption strategies over time, and is beyond the scope of this paper.[3]

RESULTS

Figure 2 illustrates the relationship between scores on the respective administration of IREAD and third grade summative test score. For the analysis of failing the second grade administration, I also illustrate the relationship between the forcing variable and the likelihood a student is classified as having a disability in the third grade, which

would have made them exempt from the retention policy.[4] The estimate for the discontinuity at the threshold (β) appears above each respective figure. Failing IREAD in the second grade increased third grade ELA scores by about 0.063σ.

We also see some suggestive evidence of a small positive spill-over onto math scores, but the figure illustrates a dip in the relationship on the failing side of the threshold, and the precisely-estimated coefficient is insignificant. We see no significant relationship between failing the second grade IREAD and the likelihood a student obtained a good-cause exemption by being placed into special education, suggesting that schools did not systematically respond to the warning that a student could be retained by seeking to exclude students from the policy.

The bottom panel illustrates the relationship between scores on the March of third grade IREAD administration and scores on the state's summative standardized test. Failing the spring IREAD led to significant benefits in both ELA (0.039σ) and math (0.020σ). Though the magnitude of the effect from failing IREAD in the spring of the third grade seems relatively small, it is important to keep in mind that the outcome is measured only four- to six-weeks after the student failed the spring IREAD.[5]

Table 2 reports results from regressions restricted to students with particular baseline characteristics. Overall, the estimated impact of the threat of retention is similar across subgroups. There is some suggestive evidence that Black students disproportionally benefit and Hispanic students do not benefit from failing the second grade test, but the estimates for these subgroups are imprecise.

Finally, Table 3 reports the results from analyses that interact failing the IREAD with the pre-policy measure of district or school value-added in the respective subject. Including the measure of district/school quality into the model does not materially impact the coefficient on failing the IREAD, lending support to the primary identifying assumption for the primary analysis. Overall, the results suggest that the treatment effect does not depend on the average quality of the student's school. The magnitude of the lone marginally significant interaction with district quality for the analysis measuring the effect of failing the second grade IREAD on ELA scores is immaterial—taken at face value, it suggests that a 1σ increase in district quality decreases the treatment effect by about 0.01σ.

CONCLUSION

This paper provides causal evidence on how students respond to the possibility of grade retention under a test-based promotion policy. Exploiting sharp discontinuities at early passing thresholds on Indiana's third-grade literacy assessment, I show that failing early attempts at the test leads to immediate and measurable improvements in achievement. These gains emerge over very short time horizons and are observed on distinct summative assessments, indicating that they reflect real changes in learning rather than test-specific preparation.

The findings contribute to the literature on test-based promotion policies by documenting an incentive margin that has received comparatively little empirical attention. While a substantial body of research examines the effects of actually being retained, far less is known about how students, parents, and schools respond when retention is a credible

possibility. The results here show that exposure to salient performance thresholds can alter behavior and generate achievement gains even when the probability of eventual retention is low. This suggests that perceived stakes play an important role in shaping educational outcomes.

The short time horizon over which these effects emerge narrows the set of plausible mechanisms. Because outcomes are measured on statewide summative assessments distinct from the IREAD, the results are unlikely to reflect test-specific coaching or familiarity. Instead, the timing is most consistent with rapid changes in instructional intensity, student effort, or parental engagement following receipt of failing results. That measurable achievement gains appear within weeks of a failed exam highlights how quickly such responses can impact the outcomes of low-performing students.

From a policy perspective, the findings suggest that the design of accountability systems may matter not only through the severity of their ultimate consequences, but also through the timing and clarity of the signals they send. Early testing and intermediate thresholds may serve as mechanisms for identifying struggling students and mobilizing resources before more disruptive interventions occur. Understanding how these incentive effects interact with subsequent instructional pathways is an important direction for future work. More generally, these findings may be relevant for policies that rely on performance thresholds with delayed consequences, including probationary periods, early warnings, and multi-stage accountability systems.

Finally, the results highlight the value of distinguishing between realized treatments and anticipatory responses when evaluating policies that operate through performance thresholds. In settings where individuals face salient rules with delayed or uncertain consequences, behavioral responses to perceived risk may account for a meaningful share of observed policy effects. Carefully isolating these responses can improve both the interpretation of existing estimates and the design of future interventions.

REFERENCES

Aucejo, E., Romano, T., and Taylor, E. S. (2022). Does evaluation change teacher effort and performance? Quasi-experimental evidence from a policy of retesting students. Review of Economics and Statistics, 104(3):417–430.

Berne, J. S., Jacob, B. A., Weiland, C., and Strunk, K. O. (2026). Assessing the impact of grade retention: A cautionary tale of exclusion restriction violations. *Journal of Public Economics,* 253:105534.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2012). Salience theory of choice under risk. *Quarterly Journal of Economics,* 127(3):1243–1285.

Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2017). rdrobust: Software for Regression-Discontinuity designs. *Stata Journal,* 17(2):372–404.

Eren, O., Depew, B., and Barnes, S. (2017). Test-based promotion policies, dropping out, and juvenile crime. *Journal of Public Economics,* 153:9–31.

Hill, C. J., Bloom, H. S., Black, A. R., and Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives,* 2(3):172–177.

Hwang, N. and Koedel, C. (2025). Helping or hurting: The effects of retention in the third grade on student outcomes. *Educational Evaluation and Policy Analysis,* 47(1):65–88.

Indiana Department of Education (2021). ILEARN 2020–2021 technical report. Technical report, Indiana Department of Education. Accessed YYYY-MM-DD.

Indiana Department of Education (2024). Iread-3 assessment. Describes IREAD-3 as a summative assessment of foundational reading skills such as word recognition, vocabulary, and comprehension based on Indiana academic standards.

Jacob, B. A. and Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics,* 86(1):226–244.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica,* 47(2):263–291.

Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. *Quarterly Journal of Economics,* 121(4):1133–1165.

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.

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.

Slungaard Mumma, K. and Winters, M. (2023). The effect of retention under Mississippi's test-based promotion policy. Technical report, Wheelock Educational Policy Center.

Souders, S. and Bifulco, R. (2025). The effect of retention threat on student learning gains: Evidence from Ohio.

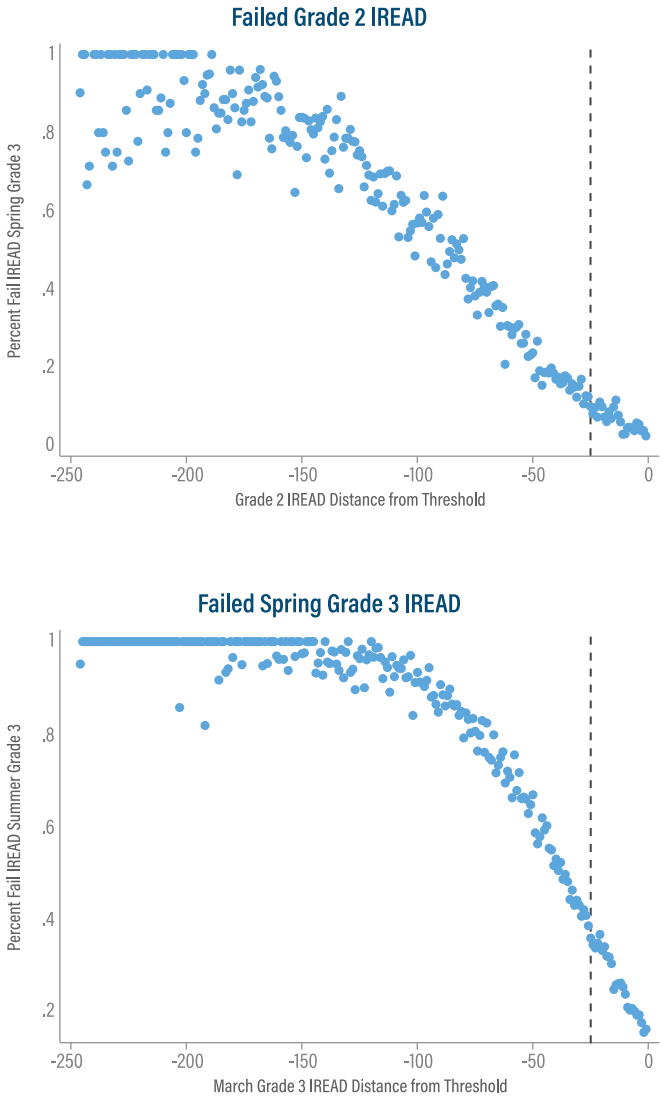
Unpublished manuscript.

DATA AND REPLICATION

The analyses use restricted administrative data from the Indiana Department of Education (IDOE). Access to these data requires a data-use agreement with each agency. Upon publication, replication code and detailed documentation will be made available upon request.

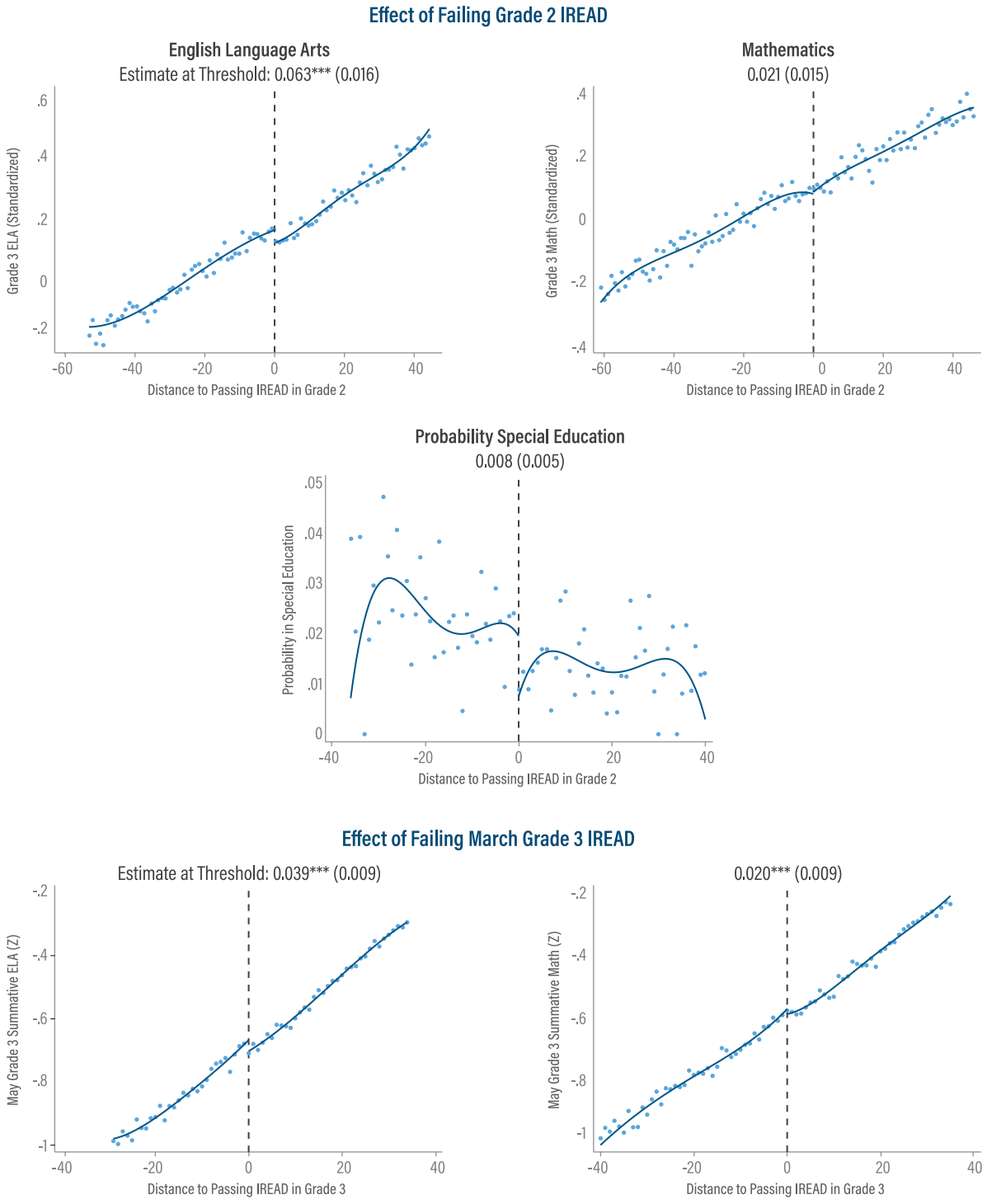
Figures

Figure 1: Percentage of Students who Fail Summer Grade 3 IREAD by Prior IREAD Score



Notes: Figure illustrates percentage of students who failed the final attempt at the IREAD in the summer of initial third grade year (y-axis) by each possible failing score on the respective prior administration of the IREAD (x-axis). Students who pass the IREAD do not take later administrations of the test. Vertical line indicates IREAD score 25 points below the passing threshold, which approximates the lower-end of the bandwidth used to create the estimation samples for RD analyses described in the Identification Strategy section.

Figure 2: RD Estimates: Effect of Failing IREAD



Notes: Figure illustrates sharp RD estimates for the effect of failing respective administration of the IREAD on student test scores and probability of placement into special education in the third grade. Test scores are from the state's summative standardized test, either ISTEP+ or ILEARN depending on year, administered in April or May of the student's initial third grade year. X-axis is the distance between the student's score and the passing score on the test, and the vertical line marks the passing threshold. Figures produced using the rdplot command in STATA, with the regression lines fit using a uniform kernel with slopes allowed to vary across the threshold. Above each figure is reported the coefficient estimate and (robust standard error) for the effect at the threshold. Regressions control for student gender, race/ethnicity, and eligibility for free or reduced price lunch in their initial third grade year. Forcing variable is the distance between the student's IREAD score and passing score. Regressions estimated using rdrobust command in STATA, apply optimal bandwidths, and a uniform kernel. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Tables

Table 1: Tests for Covariate Balance

	Male	Black	Hispanic	White	FRL	$\hat{\theta}_d$ ELA	$\hat{\theta}_d$ Math	$\hat{\theta}_s$ ELA	$\hat{\theta}_s$ Math
Grade 3 Spring	0.000 (0.006)	-0.003 (0.006)	-0.004 (0.004)	0.009 (0.006)	0.017** (0.007)	-0.000 (0.001)	-0.001 (0.001)	0.001 (0.002)	-0.000 (0.002)
Grade 2	0.037** (0.018)	0.012 (0.012)	0.003 (0.009)	-0.001 (0.015)	-0.007 (0.017)	-0.008** (0.003)	-0.004 (0.004)	-0.006 (0.004)	-0.006 (0.005)

Notes: Table reports results from sharp RD regressions testing balance for the value of covariates at the passing threshold on respective administrations of the IREAD. Each cell is the result of a different regression. Each row reports the administration of the IREAD considered in the analysis. Column labels report the dependent variable. Regressions control for distance between the student's IREAD score and passing score (i.e. the forcing variable). Robust standard errors reported in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

Table 2: RD Estimates by Student Characteristics

	Male	Black	Hispanic	White	FRL
Second Grade IREAD					
ELA	0.053** (0.021)	0.129*** (0.044)	0.003 (0.044)	0.055*** (0.018)	0.061*** (0.023)
N	22521	6606	4094	32749	23360
Math	-0.002 (0.021)	0.066* (0.039)	-0.021 (0.053)	0.012 (0.018)	0.028 (0.020)
N	22512	6600	4090	32747	23352
Spring Third Grade IREAD					
ELA	0.037*** (0.012)	0.032* (0.016)	0.011 (0.029)	0.048*** (0.011)	0.029*** (0.011)
N	324869	84526	50467	490141	302326
Math	0.024* (0.013)	0.026 (0.018)	0.012 (0.032)	0.025** (0.013)	0.013 (0.011)
N	325260	84665	50499	490411	302682

Notes: Table reports results from estimating Equation (2) to measure the effect of failing the second grade or March of third grade administration of IREAD on third grade summative test scores in samples restricted to students with particular baseline characteristics. Regressions control for student gender, race/ethnicity, and eligibility for free or reduced price lunch in their initial third grade year. Forcing variable is the distance between the student's summer IREAD score and passing score. Models estimated using rdrobust command in STATA, apply optimal bandwidths, and a uniform kernel. Robust standard errors reported in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

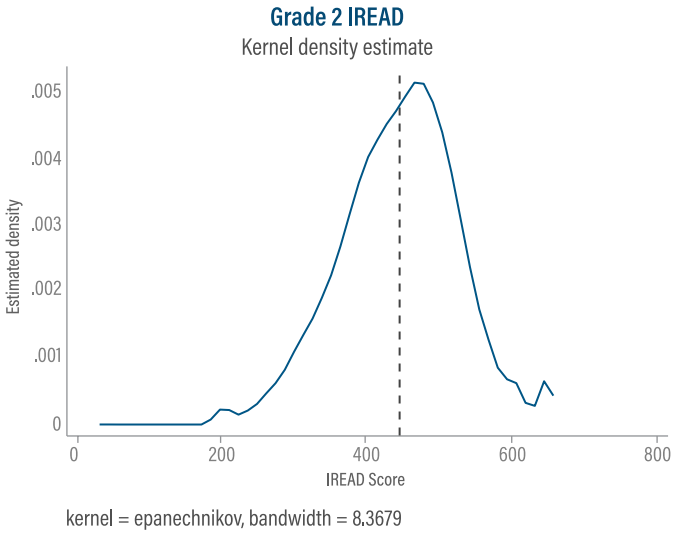
Table 3: RD Estimates, Interaction with District/School Value-Added

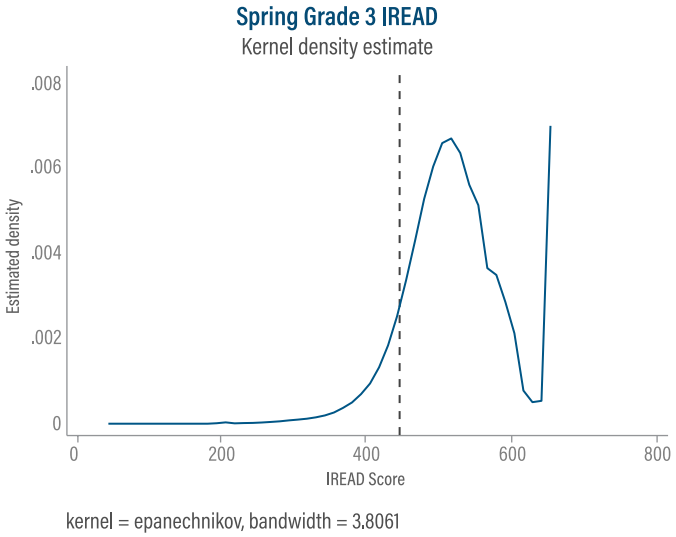
	ELA		Math	
Second Grade IREAD				
Fail	0.055*** (0.013)	0.055*** (0.014)	0.011 (0.013)	0.010 (0.013)
$\hat{\theta}_d$ x Fail	-0.165* (0.087)		-0.069 (0.069)	
$\hat{\theta}_s$ x Fail		-0.002 (0.061)		-0.014 (0.048)
Spring Third Grade IREAD				
Fail	0.032*** (0.008)	0.032*** (0.008)	0.017* (0.009)	0.017* (0.009)
$\hat{\theta}_d$ x Fail	0.067 (0.047)		0.028 (0.045)	
$\hat{\theta}_s$ x Fail		0.052 (0.032)		-0.011 (0.032)

Notes: Table reports results from estimating Equation (2) and adding interactions between failing and estimated district or school fixed-effect to measure the effect of failing the second grade or March of third grade administration of IREAD on third grade summative test scores. Regressions control for student gender, race/ethnicity, and eligibility for free or reduced price lunch in their initial third grade year. Forcing variable is the distance between the student's summer IREAD score and passing score. Regressions estimated using OLS within samples restricted to students with an IREAD scores within the bandwidth for the respective analysis reported on Table A.3. Robust standard errors reported in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

ONLINE APPENDIX

Figure A.1: Distribution of IREAD Scores





Notes: Figure illustrates kernel densities for various administrations of the IREAD, used as forcing variables in the analysis. P-value for test for discontinuity at the passing score (446) for each respective test are: 2nd Grade: 0.296; Spring of 3rd Grade: 0.781; Summer of 3rd Grade: 0.991.

Table A.1: Observed Characteristics: Grade 2 IREAD

	All Administered IREAD			Within 25 Points		
	All	Passed	Failed	All	Passed	Failed
Male	0.486 (0.500)	0.477 (0.499)	0.495 (0.500)	0.486 (0.500)	0.492 (0.500)	0.478 (0.500)
Black	0.137 (0.344)	0.072 (0.258)	0.204 (0.403)	0.113 (0.317)	0.106 (0.308)	0.121 (0.326)
Hispanic	0.099 (0.299)	0.079 (0.269)	0.120 (0.325)	0.097 (0.296)	0.093 (0.291)	0.101 (0.302)
White	0.688 (0.463)	0.776 (0.417)	0.597 (0.490)	0.717 (0.450)	0.731 (0.444)	0.702 (0.457)
FRL	0.504 (0.500)	0.381 (0.486)	0.631 (0.483)	0.499 (0.500)	0.474 (0.499)	0.527 (0.499)
$\hat{\theta}_d$ ELA	-0.000 (0.085)	0.016 (0.081)	-0.019 (0.086)	0.004 (0.082)	0.004 (0.082)	0.003 (0.081)
$\hat{\theta}_d$ Math	0.000 (0.101)	0.017 (0.096)	-0.019 (0.104)	0.006 (0.097)	0.007 (0.097)	0.005 (0.097)
$\hat{\theta}_s$ ELA	0.023 (0.118)	0.045 (0.114)	-0.001 (0.118)	0.028 (0.113)	0.029 (0.114)	0.027 (0.112)
$\hat{\theta}_s$ Math	0.012 (0.145)	0.031 (0.143)	-0.009 (0.145)	0.016 (0.138)	0.017 (0.139)	0.015 (0.138)
N	49,724	25,271	24,453	11,925	6,342	5,583

Notes: Table reports mean and (standard deviation).

Table A.2: Observed Characteristics: Spring Grade 3 IREAD

	All Administered IREAD			Within 25 Points		
	All	Passed	Failed	All	Passed	Failed
Male	0.482 (0.500)	0.478 (0.499)	0.517 (0.500)	0.503 (0.500)	0.501 (0.500)	0.506 (0.500)
Black	0.125 (0.331)	0.101 (0.301)	0.346 (0.476)	0.237 (0.425)	0.211 (0.408)	0.286 (0.452)
Hispanic	0.075 (0.263)	0.074 (0.261)	0.085 (0.278)	0.083 (0.276)	0.083 (0.275)	0.083 (0.276)
White	0.726 (0.446)	0.752 (0.432)	0.492 (0.500)	0.606 (0.489)	0.633 (0.482)	0.556 (0.497)
FRL	0.447 (0.497)	0.416 (0.493)	0.722 (0.448)	0.634 (0.482)	0.612 (0.487)	0.677 (0.468)
$\hat{\theta}_d$ ELA	0.004 (0.088)	0.009 (0.086)	-0.040 (0.089)	-0.020 (0.088)	-0.016 (0.088)	-0.029 (0.088)
$\hat{\theta}_d$ Math	0.004 (0.102)	0.008 (0.100)	-0.037 (0.106)	-0.018 (0.104)	-0.013 (0.103)	-0.026 (0.105)
$\hat{\theta}_s$ ELA	0.029 (0.128)	0.036 (0.127)	-0.032 (0.128)	-0.006 (0.127)	0.000 (0.126)	-0.018 (0.128)
$\hat{\theta}_s$ Math	0.014 (0.145)	0.020 (0.144)	-0.039 (0.148)	-0.016 (0.145)	-0.010 (0.144)	-0.026 (0.146)
N	677,662	608,150	69,512	98,120	64,338	33,782

Notes: Table reports mean and (standard deviation).

Table A.3: RD Estimates for Effect of Failing IREAD by Test Administration

	Grade 2	Grade 3		
		All	Retention	Remediation
ELA	0.063*** (0.016)	0.039*** (0.009)	0.026** (0.013)	0.046*** (0.011)
Math	0.021 (0.015)	0.020** (0.009)	0.025* (0.015)	0.015 (0.012)
Special Ed	0.008 (0.005)			

Notes: Table reports results from estimating Equation (2) to measure the effect of failing the second grade or March of third grade administration of IREAD on third grade summative test scores and probability placed into special education. Regressions control for student gender, race/ethnicity, and eligibility for free or reduced price lunch in their initial third grade year. Forcing variable is the distance between the student's summer IREAD score and passing score. Models estimated using rdrobust command in STATA, apply optimal bandwidths, and a uniform kernel. For the Grade 3 analyses, the final two columns report results for models restricted to include only students who entered the third grade within either the retention or remediation treatment cohort, respectively. Robust standard errors reported in parentheses. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$

ENDNOTES

1 The improvements from failing the spring of second grade administration of the IREAD that I identify are similar to Souders and Bifulco (2025)'s recent estimates for the effect of failing a fall administration of Ohio's gateway test on spring test scores.

2 Beginning with students who entered the third grade in 2016–17, the state gave schools greater discretion to promote students who failed the summer IREAD and provide them with remedial third grade reading instruction while they attend the fourth grade. Policymakers intended this "promotion + remediation" pathway as an exception, not the norm. In practice, however, the percentage of students who failed the summer IREAD who were retained declined from 77% to 9.6% following this change.

3 In a related paper, I employ both an instrumental variable and a multi-margin decomposition strategy to investigate the impact of retention under the various iterations of the policy on student outcomes over time.

4 English learners are also exempt from the retention policy. However, because of the nature of the classification process, it is rare for a student to be newly identified as an English learner after they have been in the school for a year or more.

5 Results reported on Table A.3 in the Online Appendix find that the effect of failing the spring IREAD was similar for students in the retention and remediation cohorts.

SUGGESTED CITATION

Winters, M. A. (2026). *Student Responses to the Credible Possibility of Grade Retention* (Working Paper 2026–4). 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 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

WHEELOCK EDUCATIONAL POLICY CENTER