Exemption Design and Human Capital Accumulation under Test-Based Promotion

Marcus A. Winters, Ph.D.* Abigail Francis†

Abstract

How institutional design shapes human capital accumulation depends on how policies reallocate individuals across investment environments. We study two regimes of Indiana's third-grade test-based promotion policy that differ in exemption structure. Leveraging regression discontinuity variation and quasi-experimental differences in exemption practices, we decompose effects across post-failure pathways. Retention followed by sustained grade reassignment generates large, persistent achievement gains, while promotion with remediation or accelerated cohort reentry yields smaller—and sometimes economically meaningful negative—effects. Dynamic skill formation depends on sustained instructional alignment and that institutional design determines both policy effectiveness and the LATE identified by promotion thresholds.

Acknowledgements: We thank the Indiana Department of Education for providing access to their data and offering valuable feedback throughout the research process. We 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 our own.

*Professor of Educational Leadership & Policy Studies, Boston University, Wheelock College of Education & Human Development. marcusw@bu.edu The author has no conflicting interest.

†Abigail Francis is a PhD student at the Boston University Wheelock College of Education & Human Development. *abbyf@bu.edu* The author has no conflicting interest.

1 Introduction

Human capital accumulation depends not only on how much instruction students receive, but on whether instruction is aligned with their existing skills. When institutions reassign individuals across discrete instructional ladders—such as grade levels in school—they alter both the timing and alignment of instruction. This paper studies how institutional design shapes dynamic skill formation in the context of test-based promotion policies. We show that policies governing grade reassignment affect achievement not simply through incentives or additional instructional time, but through sustained exposure to instruction aligned with students' ability. Moreover, we demonstrate that exemption design fundamentally alters both the of these policies on student learning and the causal parameter identified by standard regression discontinuity approaches.

Test-based promotion policies are widely used to prevent early-grade students from advancing without mastering foundational literacy skills. Under these policies, students who fail a gateway exam may be retained in grade, while those who pass are promoted on schedule. A substantial literature finds that policy-induced retention in early grades often raises short-run test scores, though effects frequently attenuate over time 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; Zhong, 2026).[1] However, in practice retention is rarely a binary treatment. Most states allow one or more "good cause" exemptions that permit failing students to be promoted, sometimes with remediation, and districts may apply these exemptions in heterogeneous ways. When such exemptions are prevalent, the discontinuity at the promotion cutoff identifies a weighted average of effects across distinct instructional trajectories. Understanding how institutional rules govern this reallocation is therefore essential for interpreting estimated local average treatment effects and for assessing the policy's impact on skill accumulation.

We examine two institutional regimes of Indiana's third-grade literacy policy that differ sharply in how exemptions are structured and applied. In the first regime, most students who

[1] We focus on test-based promotion policies for students in early elementary grades – typically third grade. Some states have similar promotion requirements for students in middle school grades.

failed the gateway exam were retained, but a substantial share of retained students subsequently skipped a grade to rejoin their original cohort. In the second regime, most failing students were promoted with mandated remedial instruction rather than being retained. These regimes generate variation not only in the extensive margin of retention, but also in the intensive margin governing post-retention instructional exposure. Leveraging regression discontinuity variation at the promotion cutoff and quasi-experimental variation in school-level exemption practices, we decompose treatment effects across these heterogeneous pathways.

Across both regimes, we find that sustained grade-level reassignment (i.e. retention followed by remaining behind one's original cohort) produces large and persistent achievement gains. In contrast, exemption pathways that either substitute remediation within an advanced-grade classroom or accelerate students back to their original cohort yield substantially smaller and sometimes negative effects relative to typical promotion. These findings are consistent with a model in which cumulative exposure to aligned instruction drives dynamic skill formation. More broadly, they illustrate how institutional design determines which complier margins are active and, consequently, what causal parameter is identified in threshold-based policy environments.

Our findings have clear and direct implications for the design of a large and rapidly growing public policy. Indiana's institutional variation is not unusual within the broader landscape of third-grade promotion policies. As summarized in Table A.1 in the Online Appendix, most states that implement test-based promotion allow one or more "good cause" exemptions for failing students. These exemptions commonly include promotion with mandated remediation, portfolio-based demonstrations of proficiency, or administrative discretion for specific student subgroups. States differ substantially in how strictly they apply promotion requirements. Florida's third-grade retention policy, for example, defines good-cause exemptions narrowly and historically retained a large share of students who failed the promotion exam (Greene and Winters, 2009; Schwerdt et al., 2017). By contrast, in Texas students who fail the reading assessment are frequently promoted following review by grade placement committees and provided mandated remediation (Jacob and Lefgren, 2004). In Colorado and North Carolina, portfolio pathways and discretionary exemptions similarly shift the policy toward conditional promotion with supplemental services rather

than mandatory retention (Colorado Department of Education, 2023; North Carolina Department of Public Instruction, 2023). Policymakers often debate whether to soften retention mandates by expanding exemption categories or substituting remediation in higher-grade classrooms. For example, Michigan recently repealed its retention requirement but kept literacy supports for failing students in place (Michigan Department of Education, 2024).

Allowing targeted exemptions may be desirable to avoid unnecessary retention. However, when exemptions accelerate students past valuable content or rely solely on in-grade remediation, they may substantially weaken the policy's impact for the marginal student. Designing effective promotion policies therefore requires careful attention to how exemptions reshape students' instructional environments.

We currently know little about the impact of exemptions under test-based promotion policies, despite the centrality to their design. Prior work acknowledges exemptions as part of the compliance environment for which researchers account when estimating the causal effect of policy-induced retention, but does not consider the longer-run implications of exemption use on later student outcomes.[2] A recent paper by Berne et al. (2026) further raises the concern that students promoted with an exemption are not conventional non-compliers in contexts where failing the test triggers remedial services even for students who are not retained – a common policy feature – potentially violating the exclusion restriction necessary for the commonly-applied fuzzy RD design to identify the LATE for policy-induced retention.

But from a policy perspective the availability and application of exemptions has broader implications than their threat to accurately identifying the LATE for policy-induced retention. Our analyses re-frame the focus of exemptions away from potential to bias RD estimates to the more pressing question facing schools and policymakers of whether the marginal student who has not demonstrated the required level of literacy skills is best served by receiving the policy's full treatment, a partial treatment with an exemption, or being typically promoted without further intervention. Across both iterations of the state's policy and employing different sources of identifying

[2]The lone exception in this literature comes from Greene and Winters (2009), who report that students who received a good-cause exemption under Florida's third grade retention policy appeared to make smaller learning gains than retained students in subsequent years, but their analysis did not exploit exogenous variation in exemption use.

variation, we find consistent evidence that students who fail the gateway reading test benefit from exposure to the full treatment under the policy. For the average failing student, receiving an exemption severely mitigates the treatment effect, or in some cases might even cause harms relative to typical promotion.

Our results for impacts under the the first iteration of Indiana's policy are relevant to research on tracking and grouping as a strategy to align instruction with student ability. Prior studies in this literature often focus on whether to advance gifted students in order to expose them to more challenging material (Tomlinson et al., 2003). There is also evidence that ability-aligned instruction has benefits for both low- and high-performing students within developing countries, where learning gaps are especially large (Duflo et al., 2011; Banerjee et al., 2016). The negative impacts from accelerating low-performing students that we identify are consistent with the logic underlying this framework and serve as unique evidence within the context of a developed country that the impacts of acceleration and skill-level mismatch are asymmetric across the ability distribution.

Our analysis also connects to the multi-margin IV literature showing that with heterogeneous responses and multiple treatment pathways standard estimands generally mix effects across distinct complier margins (Mountjoy, 2022; Kline and Walters, 2016; Mogstad et al., 2021; Bhuller and Sigstad, 2024). In our context, the relevant heterogeneity does not arise solely from differential compliance at the retention margin, but from variation in post-failure instructional trajectories. As a result, the same nominal policy can generate different estimated effects depending on how exemptions are structured and applied. More broadly, the results illustrate how policy design determines which complier margins are active and, consequently, what parameter is identified. In environments with multiple treatment margins, understanding how institutional features reallocate individuals across pathways is essential for interpreting IV estimates and for drawing policy conclusions.

2 Policy Setting

We consider two iterations of Indiana's third grade test-based promotion policy that are characterized by the structure and use of exemptions to the full treatment available to failing students during its respective time period. To aid our discussion, Figures A.1 and A.2 in the Online Appendix map the typical available pathways for students who entered third grade under each respective cohort type. Table 1 presents basic descriptive information by cohort and treatment status for all students who took the summer IREAD and those scoring near the passing threshold.

[TABLE 1 ABOUT HERE]

Both iterations of Indiana's policy required students to demonstrate a minimal level of literacy on the state's gateway reading test – the IREAD – before promotion to the fourth grade. Students with disabilities and English learners take the IREAD but are exempt from the policy and thus we exclude them from the analyses. IREAD is an untimed, computer-adaptive test. Both the content design and the passing score has remained constant.

Students in the entry cohorts we analyze first take the IREAD in early- to mid-March of third grade. Those who pass this spring IREAD administration are promoted to the fourth grade, are not required to attend summer school, and do not retake the test. Eligible students who fail the spring IREAD must attend summer school, after which they retake the test. Students who pass the summer re-take of the IREAD are promoted to the fourth grade, and those who fail are designated for retention. Our analyses include only eligible students who took the summer IREAD, which is administered after they have completed summer school.

We define the "full treatment" under the policy as third grade retention followed by cumulative promotion in subsequent years. The two iterations of the policy employed exemptions that reduced exposure to different aspects of this full treatment.

Students who entered the third grade between 2011-12 and 2015-16 experienced the first iteration of the policy. During these years, the state did not offer an explicit pathway to promote eligible students who had not passed the IREAD. Consequently, about 77% of students who failed the summer IREAD repeated third grade. The remaining 23% of failing students were promoted

despite receiving no formal exemption, and they did not take subsequent administrations of the IREAD.[3] Rather, for this cohort exemptions were applied to students after they had been retained. During these years many local LEAs used their historical discretion in determining student acceleration to push retained students forward. About 38% of all retained students in this cohort skipped the fourth grade. The state did not encourage schools to return retained students to their original cohort. We refer to students who entered the third grade during these years as the "skipping cohort."

Beginning with students who entered the third grade in 2016-17, the state altered its guidance to allow districts discretion to promote students who failed the summer IREAD. Schools were required to provide exempted students with remedial third grade reading instruction while they attended the fourth grade, and the students had to eventually pass the IREAD. The state did not specify how schools must supply these remedial services, and we do not observe the type of remediation individual schools provided to their students.[4] Though policymakers intended schools to use this exemption pathway as a targeted exception, in practice the retention rate among students who failed the summer IREAD declined from 77% to 9.6% following this change. Only 4.5% of retained students from this cohort skipped the fourth grade. We refer to students who entered third grade in fall of 2016 or later as the "remediation cohort."

3 Conceptual Framework

To fix ideas, this section develops a simple framework to consider how retention might affect student learning trajectories. Our framework builds on curriculum mismatch models (e.g. Pritchett and Beatty (2015)) by introducing institutional grade reassignment and discrete instructional

[3]Though we cannot verify why some failing students were promoted, conversations with state policymakers suggests that this non-compliance for this cohort could have been the result of some schools interpreting a loophole in the policy that students were required to be retained only in reading. We are aware of no statewide efforts to identify and further enforce these retentions.

[4]Given the high-profile nature of the policy and the need for exempted students to eventually pass the IREAD, we suspect differences across schools primarily in the type and intensity of remedial services provided to students, rather than their existence. In the context of Michigan's policy that had a similar remediation requirement for students who failed the test, Berne et al. (2026) report survey evidence that schools did provide additional supports for struggling students, and that the type of supports varied.

sequences. Whereas prior work emphasizes how system-wide curricular overreach can reduce learning for students who are academically behind, we examine how policy-induced reassignment across a fixed curriculum ladder changes alignment and, in turn, dynamic skill accumulation.

Student learning in year t depends on both entering skill and the level of instruction encountered. Let A_t denote the student's skill at the start of year t, and let L_t denote the instructional level of the grade in which they are placed, with $L_{t+1} > L_t$. Learning in year t is given by $G_t = h(A_t, L_t)$, where $h(\cdot)$ is concave and penalizes instructional mismatch. Learning is maximized when instruction is well aligned with entering skill, and declines as $|A_t - L_t|$ grows. Skill evolves according to $A_{t+1} = A_t + G_t$. This structure captures the basic intuition that students learn most when the curriculum is pitched at a level they can access, and that learning accumulates across instructional years.

Retention without exemption under the policy could generate two distinct advantages. First, repeating a grade increases alignment. Students who struggle in their initial attempt at the third grade but repeat the grade re-enter with higher skill and therefore encounter content that is better matched to their needs. Second, retention increases total instructional time by adding a year of schooling prior to exposure to higher-level content as they are cumulatively promoted in subsequent years.

The existence of exemptions creates multiple treatment pathways under the policy. All students who fail the test receive remedial instruction either when repeating the third grade or while attending the fourth grade, impacting their outcome by $\gamma^{remediation}$.[5] A portion $0 < \omega < 1$ of students who fail the summer IREAD repeat the third grade and remain behind their original cohort, exposing them to an additional year of instruction before they enter each subsequent grade, additionally impacting their outcome by $\gamma^{instruction}$. The net marginal treatment effect (MTE_{net}) is the weighted average of the effect for students who experienced the full intended treatment (MTE_{full}) and the effect on students who received an exemption available to their respective cohort (MTE_{exempt}).

[5] Our analyses assume that remediation is equivalent whether administered in the third or forth grade. In each case, students must demonstrate sufficient literacy ability by passing the IREAD test.

$$MTE_{net} = \omega MTE_{full} + (1 - \omega) MTE_{exempt} \tag{1}$$

$$MTE_{net} = \omega(\gamma^{remediation} + \gamma^{instruction}) + (1 - \omega)\gamma^{remediation} \tag{2}$$

The two exemption strategies reduce access to additional instruction in materially different ways. For failing students in the remediation cohort, exempted students are exposed to fourth grade instruction while they receive remediation to increase their literacy skills, but they do not receive an additional year of learning before entering subsequent grades. For failing students in the skipping cohort, however, receiving an exemption is not simply a weaker form of the full intended treatment under the policy because they are never exposed to fourth grade instruction. If students who skip the fourth grade enter fifth grade instruction with weaker cumulative preparation than marginally promoted students who have completed fourth grade material, then the effect of the exemption pathway for students in the skipping cohort could be negative relative to typical promotion.

4 Data and Sample Selection

To conduct our analyses, the Indiana Department of Education provided us with longitudinal administrative data containing demographic characteristics and relevant test scores for the universe of public school students in the state from 2007 through 2024. The estimation samples can differ substantially across analyses over time depending on whether outcomes for a cohort are observed for a given grade or year. To facilitate tracking of the estimation samples, Table A.2 in the Online Appendix describes the typical grade progression and observable scores for retained and promoted students from each cohort, and Table A.3 lists the entering third grade cohorts included in each analysis type in the paper.

As described above, the IREAD is the test that students are required to pass under the policy. Our analyses use scores from the summer administration of the IREAD, which is only taken

by students who did not pass in the spring. The outcome for each analysis is the student's score on the state's summative standardized test in ELA or math administered in the spring to students in grades three through eight. These tests are distinct from the IREAD. We observe scores on these tests for each year from 2007 through 2024, except for 2020 when students were not administered the tests.

We classify students by the year that they first entered the third grade. We analyze students who entered the third grade between 2011-12 and 2021-22, but not students who entered third grade in 2019-20 because they were not administered the IREAD and were not subjected to the policy.

To be included in any sample, a student must have taken the summer IREAD and must not be classified as having a disability or as an English learner. Analyses that exploit RD variation are further restricted to include only students whose summer IREAD score falls within the optimal bandwidth for the respective comparison.

We also must observe the student's outcome associated with the respective analysis. For example, a student who entered third grade in 2022 cannot be included when making same-grade comparisons of student outcomes in the sixth grade because we only observe two subsequent years of their test scores; students who entered third grade in 2017-18 are not included in same-year comparisons in the third year because they did not take standardized test in the spring of 2020, and so on.

Students took the ISTEP+ until the state changed to the ILEARN in 2019. Both tests use a vertically aligned scale such that a score is meant to represent the same level of ability regardless of the grade level of the test. But scaled scores are not comparable across the tests. When making same-grade comparisons, we standardize outcome test scores by grade and year, allowing the sample to include students who took either the ISTEP+ or the ILEARN, and we include a control for the test administered. We must use scale scores when making same-year comparisons. Because scale scores are not comparable across the tests, we only report same-year comparisons using the ISTEP+ test when analyzing the skipping cohort and scores on the ILEARN when analyzing the remediation cohort.

The sample is further restricted for regressions that use an instrument for the lagged percentage of students in the school who were granted an exemption. For the skipping cohort, this instrument is the percentage of retained students in the school who skipped the fourth grade the prior year. These analyses do not use students who entered third grade in 2011-12 because no students in the prior year were retained (or exempted) under the policy. Further, we exclude students in schools that retained no students the prior year because for them the instrument is undefined.

For the remediation cohort, when applying the instrumental variable approach we include only students who entered third grade in 2017-18 or 2018-19 because the combination of expected exclusions and exclusions due to the pandemic eliminate future entering third grade classes. The respective instrument is the lagged percentage of students who failed the summer IREAD in the school who were promoted and received remedial instruction in the fourth grade. As was the case for the prior cohort, we lose students from the first entering third grade class (2016-17) because we do not observe a lagged exemption for their school, as well as those from schools where no students failed the previous year. In addition, because students who entered third grade in 2019-20 did not take IREAD and were exempt from the policy, we lose both that group as well as students who entered third grade in 2020-21 because we do not observe a lagged exemption rate for them.

4.1 Within-Grade vs Within-Year Test Score Comparisons

Measuring the effect of grade retention on later student outcomes is complicated by the fact that in subsequent years retained students who remain behind their cohort are a grade lower than the counterfactual. There are two potential comparison strategies available: Same-Grade and Same-Year. Neither is an *a priori* superior strategy (Schwerdt et al., 2017).

The same-grade strategy compares retained students to promoted students after each completes a given grade. The benefit of the same-grade approach is that it incorporates the additional year of instruction a student receives as part of the retention treatment. Though we only observe test scores through eighth grade, retained students who are subsequently cumulatively promoted will have received an additional year of instruction when they graduate than if they had not

been retained. However, a concern with the same-grade approach is that treated students are a year older than comparison students when they take a grade-level test, so differences in their test scores could in part reflect maturity benefits.

Alternatively, the same-year strategy compares student outcomes after a given number of years, when fully-treated retained students are a grade below their promoted peers. The same-year comparison is appealing because it removes potential bias from maturity effects. But to the extent that the intended treatment includes exposure to an additional year of instruction the same-year comparison will understate the treatment effect. Further, these comparisons rely heavily on accurate vertical-alignment of standardized test scores across grades, which is difficult to accomplish in practice (Ballou, 2009). Validity studies of both ISTEP+ and ILEARN raised concerns about vertical scaling (Roeber and Briggs, 2016; Indiana Department of Education, 2021). We thus suggest caution when interpreting results from Same-Year analyses.

4.2 Controlling for Differences in District/School Quality

As we discuss in more detail below, our analyses differentiating the effect across the full-treatment and respective exemption pathways require that the district/school a student attends in the third grade conditionally impacts later outcomes only by influencing the likelihood they receive an exemption. A potential threat to identification is the possibility that a district/school's propensity for granting exemptions is associated with its independent impact on student learning.

A seemingly straightforward strategy to address this problem is to employ district/school fixed-effects. We report estimates from models that use school fixed-effects as a robustness test when analyzing the skipping cohort. However, this strategy is conceptually limited in our context where we suspect across-district/school differences should be more important than within-district/school differences in propensity to exempt students. More practically, the small annual within-district/school change in the instrument we use to identify effects for the remediation cohort produces very imprecise and unstable estimates when the regression incorporates fixed-effects.

We instead address this challenge by investigating regressions that control for a direct estimate of a school's persistent quality measured using data from pre-policy years. Using data

on students from grades three through five during the 2007-2011 school years as the outcome, we estimate simple value-added models using district fixed-effects (θ_d) and the student's lagged test score to predict test scores:

$$y_{idt} = \alpha + \phi y_{idt-1} + \theta_d + \mu \tag{3}$$

We capture $\hat{\theta_d}$, which gives the average test score gain made by students attending each district relative to an omitted comparison district in the state. We also use this strategy to generate an estimate for pre-policy school value-added $\hat{\theta_s}$.

We include this measure of persistent district/school quality as a control in alternative regression specifications, motivated by the idea that conditional on observed district/school value-added the remaining variation in its exemption-style is administrative rather than pedagogical. This approach assumes that district/school value-added from 2007-2011 is a reliable proxy for district/school quality from 2012-2024, and that changes in relative district/school quality after 2012 was conditionally unrelated to propensity to grant exemptions.

5 Measuring the Net Treatment Effect by Cohort

We begin our investigation by measuring the average net treatment effect from failing the summer IREAD under the two iterations of the policy. These analyses leverage plausibly exogenous variation in treatment status at the passing threshold on the summer IREAD within a sharp RD design. The key identifying assumption is that absent the policy the relationship between summer IREAD scores and later outcomes would have been smooth at the passing threshold. The nature of the summer IREAD as a computer-administered standardized test makes it unlikely that students or schools could systematically manipulate the scores. Consistent with this expectation, with few exceptions the results reported in Table 2 find balance in the value of baseline covariates at the threshold, and density tests illustrated on Figure A.3 find no evidence of a break in the distribution of summer IREAD scores at the passing threshold.

[TABLE 2 ABOUT HERE]

We estimate regressions measuring average outcome differences at the threshold between students who passed or failed the summer IREAD taking the form:

$$y_i = \alpha + \gamma X_i + f(a_i) + \beta_{net} \mathbb{1}(a_i < 0) + \epsilon_i \tag{4}$$

where y_i is the student's 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. Our primary regressions treat $f(a)$ as a linear function with a uniform kernel allowed to vary on either side of the threshold.[6] The sample for these RD regressions only include eligible students with a summer IREAD score that falls within the MSE-optimal bandwidth for the respective outcome (Calonico et al., 2017). We rebuild the bandwidth for each regression.

We are primarily interested in β_{net}, which under the classical RD assumptions identifies the causal effect of failing the summer IREAD. This estimate collapses the impact for both failing students who received the full intended treatment and those who received the exemption associated with their cohort into a single effect weighted by the share of students assigned to each respective treatment pathway.

The rows labeled "Net Effect" on Tables 3 and 4 report the effect of failing the summer IREAD by cohort in ELA and math, respectively. Overall, students from both cohorts who failed the summer IREAD experienced significant net benefits relative to students who passed. For the same-grade analyses, failing students from the skipping cohort make large initial gains that dissipate over time. As late as the seventh grade, impacts for the skipping cohort are significant and meaningful (0.118σ in ELA and 0.058σ in math). We also observe meaningful but imprecisely estimated benefits from the within-year comparisons, translating to about 0.053σ in ELA for the skipping cohort after five years. However, the same-year estimates do not find significant benefits for the skipping cohort in math over time, and the average net treatment effect after five years is negative and translates to about -0.09σ.

[6] Results reported in Tables A.4 through A.6 in the Online Appendix demonstrate that the results are not sensitive to these specification choices.

The pattern of results for the remediation cohort is different than for the skipping cohort, though it ultimately lands in a similar place over time. Students in the remediation cohort on average sustain a modest initial gain through middle school grades and across years. The net-impact of failing the summer IREAD for the remediation cohort as late as seventh grade and the fifth subsequent year are fairly similar to the respective estimate for the skipping cohort.

[TABLE 3 ABOUT HERE]

[TABLE 4 ABOUT HERE]

5.1 Describing Relative Influence of Each Treatment Pathway on the Net Treatment Effect

For each cohort, the net effect of failing the summer IREAD is the weighted average of the effect for students exposed to the full treatment and for students who received an exemption. The RD setting itself does not provide exogenous variation in pathway assignment among those who failed the summer IREAD necessary to distinguish the causal effect of receiving an exemption relative to receiving the full treatment. However, we can describe the relative influence of each treatment pathway on the observed outcome differences by estimating regressions that modify Equation (4) to replace the indicator for failing the summer IREAD with a vector of indicators classifying whether a failing student experienced the full treatment or received an exemption. Formally:

$$y_i = \alpha + \gamma X_i + f(a_i) + \sum_{k=1}^{K} \beta_k Type_k + \epsilon_i \tag{5}$$

For the skipping cohort, the vector $Type_k$ assigns each student who failed the summer IREAD into one of three mutually exclusive groups: 1) Failed but promoted to the fourth grade (i.e. non-compliers), 2) full treatment and 3) retained + skipped fourth grade. For students in the remediation cohort, $Type_k$ classifies failing students to either the full treatment or remediation in the fourth grade. Students who passed the summer IREAD, all but a trivial number of whom were promoted and received typical fourth grade instruction, serve as the omitted comparison group. The β_k's from estimating Equation (5) may in part reflect selection related to determining whether

a failing student receives and exemption, and thus are suggestive. Further, the standard errors do not account for first-stage uncertainty in pathway assignment among failing students, and thus are downward biased for the causal effect of one pathway relative to another.

The results from these analyses appear on Tables 3 and 4 below the dashed line separating the coefficients from the respective net-effect of failing the summer IREAD. This exercise reveals strong potential that the treatment effect under the policy varies substantially by exemption status. For both cohorts, students who received an exemption consistently post scores near or sometimes below the comparison group of students who passed the test, similar to what one would expect to observe from non-compliers. In contrast, for both cohorts students exposed to the full treatment sustained large initial test gains. Indeed, the coefficients measuring the difference at the threshold between those with or without an exemption relative to those who passed the summer IREAD are very similar across the cohorts.[7]

6 Relative Size of Treatment Pathway Groups

The extent to which the MTE of the full treatment differs from the net impact from failing the summer IREAD depends in part on the share of failing students who received an exemption.

For students in the skipping cohort, the descriptive statistics reported in Table 1 suggests that similar portions of failing students either experienced the full treatment or received an exemption. To further investigate, Table 5 reports the results from fuzzy RD regressions that use failing the summer IREAD as an instrument to estimate the LATE at the passing threshold for policy-induced third grade retention on a student's subsequent grade progression.

[TABLE 5 ABOUT HERE]

For students who ever attend the fourth grade, policy-induced retention delayed fourth grade entry by one year. But this effect drops sharply: Policy-induced third grade retention within

[7]Figures A.4 through A.7 in the Online Appendix illustrate the discontinuity by treatment type for each cohort, subject, and type of comparison. The figures reveal very similar patterns across each analysis, and throughout the distribution of the summer IREAD test.

this cohort on average delayed fifth grade entry by 0.416 years, and delayed entry into high school by 0.376 years.[8] This pattern is driven by a large group of retained students skipping the fourth grade. For students in the skipping cohort scoring at the threshold, third grade retention increased the probability of skipping the fourth grade by about 53.6 percentage points.

We next consider the relative size of the treated groups within the remediation cohort. Within a sample of eligible students in the remediation cohort, we estimate sharp RD regressions like Equation (4) but that use an indicator for whether the student was retained as the outcome. The results reported in Table 6 show that depending on the analysis sample, for students in the remediation cohort failing the summer IREAD increased the probability of repeating the third grade at the threshold by between $6.5 - 13\%$. Because within this cohort there are no conventional non-compliers who failed the summer IREAD and received typical fourth grade instruction, we can interpret that the remaining $87 - 93.5\%$ of failing students received an exemption.

[TABLE 6 ABOUT HERE]

7 Measuring the Effect of Full Treatment Relative to Exemption by Cohort Type

In this section, we seek to distinguish the causal effect of treatment for failing students who received the full treatment and those who obtained an exemption under each respective version of the policy. For both cohorts, our primary strategy is to exploit plausibly exogenous district- or school-level variation in propensity to grant exemptions to instrument for the conditional likelihood that a failing student receives an exemption. However, the different nature of the data and type of exemption across the iterations of the policy requires us to apply these instruments in slightly different manners to each respective cohort. We thus discuss the analyses for each respective cohort separately.

[8] This finding also implies that the taxpayer cost to retaining a student is substantially less than the state's annual per-pupil expenditure (Winters, 2023).

7.1 Skipping Cohort

7.1.1 Fuzzy RD: Effect of Retention and Full-Treatment Dosage

Our first strategy to differentiate the effect of full treatment relative to exemption under the policy for the skipping cohort is to apply two complementary fuzzy RD analyses. These analyses differ only in that they make opposing extreme assumptions about the influence that exempted students have on the reduced-form, and thus they produce bounds for the full treatment effect within the skipping cohort under the assumption that remediation and additional exposure to instruction are weakly beneficial on average.

We first use failing the summer IREAD to instrument for repeating the third grade, giving the LATE for policy-induced retention. This alone is a highly policy-relevant parameter. However, for our purposes it likely understates the full treatment effect because it treats retained students who received an exemption as equivalent compliers. In this way, the LATE for third grade retention is in essence the reduced-form for the full treatment effect. Returning to the framework we described in Section 3, the IV estimate for third grade retention under the policy will understate the intended treatment effect if $\gamma^{instruction} > 0$.

An alternative strategy uses failing the summer IREAD to instrument for whether the student is behind their original cohort during the respective grade or year analyzed. This strategy is akin to a dosage approach embedded within the fuzzy RD design. The first-stage treats retained students who skipped the fourth grade as non-compliers, which will lead the regression to overstate the full treatment effect if $\gamma^{remediation} > 0$.

Table 7 reports fuzzy RD estimates from both strategies. Regressions that instrument for repeating the third grade find that policy-induced retention on average leads to large immediate test score improvements that fade as the student progresses through middle school. The same-grade comparisons find marginally significant and economically meaningful average benefits from third grade retention as late as the eighth grade. Though estimated less precisely, the same-year analyses also find significant but smaller benefits after four years. The substantial negative coefficient for the within-year estimates in math after five years is somewhat concerning. However, in addition to

being estimated imprecisely, it exhibits an odd pattern relative to the earlier years and thus seems likely to be a spurious result, perhaps influenced by misalignment of the vertical scale scores on the ISTEP+ across the relevant grades.

[Table 7 ABOUT HERE]

The dosage-style regressions instrumenting for being behind the original cohort produce estimates that are more than twice the size of the LATE estimates for third grade retention, though they are measured even less precisely. Taken at face value, these estimates suggest that exposure to the full treatment improved ELA scores by about 0.186σ in the eighth grade, and about 0.14σ four and five years after initial retention.

7.1.2 Multi-Margin Deconstruction Approach

Our second strategy is to apply the methodology developed by Mountjoy (2022) to deconstruct the causal effect of failing the promotion cutoff into the marginal effects of the distinct post-failure treatment pathways. This approach exploits multiple sources of quasi-experimental variation to separately identify the effect of the full treatment and the effect of retain-then-skip relative to typical promotion. The discontinuity in exposure to treatment at the passing threshold within the fuzzy RD design (Z_{fail}) leverages exogenous variation in retention, and thus assignment to one of the two post-retention pathways. The second instrument exploits heterogeneity in how schools implement the retention policy, measured by the prior-year share of retained students who skipped the fourth grade (Z_{pskip}).

The deconstruction framework treats retention as a two-margin treatment. Identification relies on a 2×2 system in which Z_{fail} primarily affects the extensive margin of retention, while Z_{pskip} affects the intensive margin of post-retention assignment, but has little effect on overall retention rates. The interaction of these sources of variation allows the effects of the full-treatment and exemption pathways to be separately identified.

Table 8 reports results from regressions testing for balance in the value of observed char-

acteristics across Z_{pskip}.[9] Schools with higher propensity to skip retained students past the fourth grade have students within the estimation sample who are more likely to be male and Black, but are as likely to be eligible for Free or Reduced Priced Lunch and scored similarly on the summer IREAD assessment. Propensity to skip retained students is unrelated to pre-policy school value-added in ELA. The relationship between skipping propensity and the school's math value-added is significant but immaterial, translating to a decline of about 0.01 in math value-added for a 1σ increase in skip propensity. Results in Table A.7 show that controlling for the student's summer IREAD score does not meaningfully alter the estimates from these balance tests.

[TABLE 8 ABOUT HERE]

Identification of Z_{pskip} does not require that students are randomly assigned to schools. Rather, the key exclusion restriction is that, conditional on proximity to the promotion cutoff and baseline covariates, marginal students are quasi-randomly exposed to their school's prevailing implementation practices. The RD design identifies a population of marginal students near the cutoff who are comparable within schools, while variation in Z_{pskip} reflects cross-school differences in how retained students are assigned to post-retention pathways. The identifying assumption is that this school-level variation affects achievement only through shifting marginally retained students across pathways, and not through other bundled differences in instruction or supports.

A remaining concern is that cross-school differences in post-retention pathway assignment may be bundled with other aspects of how schools support marginally failing students. For example, schools that are more likely to return retained students to their original cohort may also differ in the structure or intensity of instructional supports provided during and after retention. Such bundled implementation would violate the exclusion restriction if it affected outcomes directly, rather than solely through pathway assignment. We address this concern in two ways. First, as a robustness check, Table A.8 in the Online Appendix shows that results from regressions that include school fixed-effects are similar in magnitude but much less precise than the main results. Second, Table A.9 in the Online Appendix shows that controlling for pre-policy school value-

[9] Figure A.8 plots the distribution of the lagged share of retained students who subsequently skipped a grade, illustrating substantial cross-school heterogeneity in post-retention pathway use.

added ($\hat{\theta}_s$) does not materially alter either the coefficient or associated standard error. This pattern of results further supports the interpretation that the reallocation instrument captures variation in pathway assignment rather than broader instructional quality.

Table 9 reports results from this deconstruction exercise. Columns 3–6 document the first-stage relationships between the instruments and treatment margins, and the final two columns report the estimated pathway-specific effects on achievement. As the instrument is measured at the school-level, for these analyses bootstrapped standard errors are clustered by school.

[TABLE 9 ABOUT HERE]

The first-stage results align closely with the institutional setting and provide support for the identifying variation. Failing the promotion cutoff substantially increases exposure to both pathways, raising the probability of both the full-treatment and retention-then-skip pathways. In contrast, the prior-year share of retained students who skipped ahead shifts students between the two pathways in opposite directions, with effects that sum to approximately zero. This pattern indicates that school practices reallocated retained students across pathways without materially affecting the overall likelihood of retention, which is precisely the variation required to separately identify pathway-specific effects.

Retention accompanied by remaining behind grade ($\hat{\beta}_{full}$) generates economically meaningful and significant estimated gains in both subjects that persist through middle school. In ELA, students exposed to the full treatment score about 0.44σ higher in the eighth grade and about 0.128σ higher after five years than if they had been typically promoted. We see similarly large impacts from the full treatment in math, with the continued exception of the odd pattern after five years. However, note that the results are imprecise, exhibiting wide confidence intervals.

In contrast, the estimated effect of retention followed by an exemption allowing cohort reentry ($\hat{\beta}_{return}$) is consistently negative but often insignificant relative to typical promotion. Although estimates for the exemption pathway are imprecise, the potential harms to student outcomes implied by the coefficient estimate are economically meaningful. In ELA, the coefficients suggest that retention followed by skipping the fourth grade leads students to score about -0.144σ in the eighth grade and -0.141σ after four years.

The negative effect for students who obtained an exemption relative to if they had been typically promoted is consistent with the conceptual framework outlined in Section 3. Retained students who subsequently skip the fourth grade forgo exposure to fourth-grade instruction entirely. When these students rejoin their original cohort, they face higher-grade curricula with weaker cumulative preparation than either marginally promoted students or retained students who remained behind grade, both of whom experienced fourth grade instruction. In this way, cohort reentry may exacerbate instructional mismatch rather than alleviate it, potentially resulting in outcomes that are worse than typical promotion.

It is notable that the dosage and deconstruction approaches are not directly comparable. First, as described in Section 4, the sample used for the deconstruction analysis is more restrictive than the primary fuzzy RD sample. The two approaches also recover effects for different complier populations. The dosage estimator captures students whose grade progression is altered by policy-induced retention, while the deconstruction captures students whose post-retention pathway is shifted by school-level practices. Further, though the estimate for the full treatment effect from the deconstruction exercise at times yields larger impacts than the corresponding dosage-style estimates, the confidence intervals overlap in nearly all cases, reflecting substantial sampling uncertainty under both approaches. Thus, we do not interpret differences in point estimates as evidence that one estimand systematically exceeds the other, but rather as complementary evidence drawn from distinct sources of identifying variation.

7.2 Remediation Cohort

We next investigate the impact of retention relative to remediation in the fourth grade. As was the case for the skipping cohort, the central challenge is that students who fail the summer IREAD are non-randomly placed onto the full-treatment or exemption pathway. Under typical conditions, we would employ our estimates for the effect of failing the summer IREAD on the probability a student was retained that are reported in Table 6 as the first stage within a fuzzy RD design to measure the effect of policy-induced retention on later outcomes. However, under this iteration of the policy, failing students who were nonetheless promoted are not typical non-compliers because

they receive a different form of treatment (i.e. third grade remedial services in fourth grade), violating the exclusion restriction necessary for the fuzzy RD design (Berne et al., 2026). The large share of failing students who receive an exemption suggests that if it exists then such bias could be substantial in our context.

We address the identification challenge by leveraging plausibly exogenous variation across districts/schools in propensity to exempt failing students. Specifically, we use the percentage of failing students who were promoted to receive remedial reading instruction in the fourth grade within the student's district/school the previous year (Z_{prem}) to instrument for the probability they are retained.[10] These analyses include only students who failed the summer IREAD, all of whom were placed on one of the two available treatment pathways. The resulting IV estimand identifies the LATE of the full-treatment relative to receiving an exemption for students whose post-failure pathway is shifted by district/school exemption style.

We cannot apply this instrument within a multi-margin decomposition analysis like we did for the skipping cohort.[11] For students in the remediation cohort, the policy environment effectively collapses the failure margin on the summer IREAD such that the share of failing students who receive an exemption is so large that a multi-margin decomposition that separately instruments failure and exemption status is weakly identified in this setting. Our failers-only IV instead directly targets the empirically relevant margin of exemption assignment for which sufficient quasi-experimental variation exists.[12]

Causal identification requires that for students who failed the summer IREAD, Z_{prem} affects outcomes only by influencing whether they receive an exemption. The most likely threat to identification is the potential that a district/school's propensity to exempt failing students is

[10] Figure A.9 plots the distribution of the lagged share of failing students who received an exemption to be promoted to the fourth grade and receive remedial services, illustrating substantial cross-school heterogeneity in post-retention pathway use.

[11] A similar dosage strategy is also not available because students in the remediation cohort who received an exemption did not repeat the third grade.

[12] As reported on Table 6, crossing the promotion cutoff induces a large discontinuity in failure status but only an approximately 6-12 percentage point change in retention probability, rendering the second margin weak. Estimation of a two-margin system yields unstable first stages and imprecise second-stage estimates. Because all failing students receive either retention or remediation (with no exposure to typical fourth-grade instruction), the failers-only IV recovers the relevant causal contrast without relying on weak instruments.

correlated with its independent contribution to student learning.

Table 10 reports the results from regressions testing for covariate balance across Z_{prem} for each of the same-grade and same-year analyses.[13] Overall, observed student characteristics are balanced when using both the school and district-level sources of variation for Z_{prem}, with the exception that in the district-level analyses the instrument correlates with the likelihood a student is Hispanic. The value for the pre-policy district value-added is balanced in ELA, but there are significant differences in district-level math value-added and for both subjects at the school-level. However, the magnitude of these differences is again immaterial, implying that a 1σ increase in Z_{prem} is associated with 0.02 lower math value-added at both the district and school level. Nonetheless, this pattern suggests that accounting for differences in district/school quality may be important for identifying the main effect.

[TABLE 10 ABOUT HERE]

As an additional specification check on the instrument, Table 11 reports reduced-form regressions examining whether the relationship between Z_{prem} and later achievement differs for students just above and just below the promotion threshold. The estimation sample includes eligible students with summer IREAD scores within the bandwidth used for the RD analyses of the net effect on ELA outcomes. We interact Z_{prem} with an indicator for failing the summer IREAD.

This exercise provides a direct test of the instrument's exclusion restriction. If Z_{prem} captures broader differences in school quality or unobserved instructional inputs, it should predict outcomes for both failing and passing students near the threshold. In contrast, if the instrument operates solely by reallocating failing students across post-failure pathways, it should affect only those who failed the test, whose instructional trajectories are subject to exemption-based variation. Students who passed the test are not exposed to this margin and therefore provide a natural placebo group.

Consistent with the exclusion restriction, we find that higher school-level exemption propensity is associated with lower subsequent achievement among students who failed the summer IREAD, but—with only minor exceptions—is unrelated to outcomes for students who passed.

[13] Table A.10 in the Online Appendix reports first-stage results.

The absence of systematic effects among passing students provides evidence against the concern that Z_{prem} proxies for general school effectiveness rather than pathway assignment.

[TABLE 11 ABOUT HERE]

Finally, we might also be concerned with the interpretation of the results under the different policy regimes if the same districts/schools that skip retained students forward are also the schools that are most likely to remediate students who fail after the change in state guidance. Figures A.10 and A.11 in the Online Appendix illustrate the relationship between Z_{pskip} in the last observed year (2016) and Z_{prem} in its first observed year (2018) and the district and school levels, respectively. The figures illustrate meaningful variation in the value for both instruments that only modestly correlate (about 0.30).

Table 12 reports the IV estimates for the LATE of receiving the full intended treatment relative to receiving an exemption among students who failed the summer IREAD. Because the sample is more heavily restricted as described in Section 4, we can only make same-grade comparisons for impacts in sixth and seventh grade and same-year comparisons in years 3, 4, and 5. The pattern of results is similar whether the analysis employs variation in the instrument across districts or schools, though regressions that employ school-level variation in the exemption rates are considerably less precise.[14]

We find substantial benefits in both ELA and math for the full treatment relative to receiving an exemption to be remediated in the fourth grade. Failing students who are retained score about half a standard deviation higher in the seventh grade than if they had received an exemption. We also see substantial benefits from the same-year comparison after three years in ELA and after five years in math. However, all of the IV estimates exhibit wide confidence intervals, and thus we suggest caution when interpreting the magnitudes.

[TABLE 12 ABOUT HERE]

As a robustness check, for each analysis Table 12 also reports estimates from regressions that control for district/school value-added measured during the pre-policy years. Including this

[14] Analyses that use district-level variation cluster standard errors at the district level, and analyses that employ school-level variation cluster standard errors at the school level.

control does not materially alter the estimate for the full-treatment effect relative to receiving an exemption. This pattern of results supports the key identifying assumption for the IV analysis that school propensity to exempt failing students conditionally impacts later student learning only by altering the probability they receive an exemption.

7.3 Summary and Interpretation

The results across both policy regimes suggest that what matters for the marginal failing student is not simply whether the student is formally retained, but whether retention results in sustained realignment of instructional exposure.

The reduced-form RD estimates show that failing the promotion cutoff increases achievement on average under both regimes. But the pathway decomposition reveals that these net effects mask substantial heterogeneity in post-failure trajectories. In both cohorts, students exposed to the full intended treatment experience large and persistent gains. In contrast, students who receive exemptions exhibit much smaller gains, and in some specifications perform no better or worse than typically promoted students.

The skipping cohort provides especially sharp evidence on the mechanism. Students who were retained and remained behind their cohort show substantial and sustained improvements. Students who were retained but subsequently skipped fourth grade, thereby rejoining their original cohort, do not share those gains and in several specifications appear to suffer economically meaningful losses relative to if they had been typically promoted. Though imprecisely estimated, the difference between these pathways is large relative to conventional policy effect sizes in education. This pattern is difficult to reconcile with a purely motivational interpretation of retention and instead aligns closely with a model in which cumulative exposure to grade-aligned instruction drives skill formation.

The remediation cohort isolates a related but distinct margin. Using cross-district and cross-school variation in exemption propensity, we estimate the causal effect of retention relative to promotion with remediation. The IV results indicate that full retention substantially outperforms remediation delivered in the subsequent grade. While these estimates are imprecise, their

magnitudes are large and consistent across subjects and comparison strategies. Remediation in an advanced-grade classroom appears to be, at best, an imperfect substitute for an additional year of grade-level instruction.

8 Conclusion

Test-based promotion policies are often evaluated as binary interventions. This paper shows that such a characterization is incomplete. In practice, the effects of these policies depend critically on how exemptions reshape students' instructional trajectories.

Using two institutional regimes of Indiana's third-grade literacy policy, we document three central findings. First, crossing the promotion threshold generates meaningful achievement gains on average, confirming prior evidence that retention policies can raise short-run performance. Second, these average effects conceal substantial heterogeneity across post-failure pathways. Students who repeat third grade and remain behind their original cohort experience large and persistent gains. Third, exemptions that accelerate students back to their original cohort or substitute remediation within an advanced-grade classroom substantially weaken—and in some cases may reverse—those gains.

Taken together, the results indicate that the effectiveness of retention policies operates primarily through sustained instructional alignment. An additional year of grade-level instruction appears to generate durable improvements in achievement for marginally failing students. By contrast, partial treatments that attempt to preserve on-time grade progression reduce exposure to aligned content and produce much smaller benefits.

These findings have direct implications for policy design. Allowing exemptions may serve important equity or political objectives, particularly given the disruptive nature of grade retention. But widespread or poorly targeted exemptions can fundamentally alter the parameter a retention policy identifies and substantially dilute its impact. Policymakers designing early-literacy interventions should therefore attend not only to the retention cutoff itself, but also to the structure and implementation of post-failure pathways.

More broadly, the mechanism we identify extends beyond early-grade promotion policies to a wide class of institutional settings in which individuals are reassigned across discrete investment environments. Many education and labor market interventions—such as remedial college placement, special education assignment, vocational tracking, workforce retraining, or even sentencing and rehabilitation programs—shift individuals between instructional or developmental ladders that differ in both level and sequencing of content. In such environments, the impact of reassignment depends not only on additional time spent in training, but on whether that time generates sustained alignment between task difficulty and underlying skill. Policies that partially reassign individuals while preserving rapid progression may weaken alignment and attenuate long-run returns, whereas policies that durably realign exposure can alter dynamic skill accumulation paths. More generally, our results underscore that in multi-pathway institutional settings, treatment effects are shaped by how policy reallocates individuals across investment trajectories, and empirical estimates must be interpreted through the lens of these underlying margin shifts.

References

Ballou, D. (2009). Test scaling and value-added measurement. *Education Finance and Policy*, 4(4):351–383.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., Shotland, M., and Walton, M. (2016). Mainstreaming an effective intervention: Evidence from randomized evaluations of "Teaching at the Right Level" in India. Technical report, National Bureau of Economic Research.

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.

Bhuller, M. and Sigstad, H. (2024). 2sls with multiple treatments. *Journal of Econometrics*, 242(1):105785.

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

Colorado Department of Education (2023). Read Act annual report.

Duflo, E., Dupas, P., and Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review*, 101(5):1739–1774.

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. (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.

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.

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.

Kline, P. and Walters, C. R. (2016). Evaluating public programs with close substitutes: The case of Head Start. *Quarterly Journal of Economics*, 131(4):1795–1848.

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.

Michigan Department of Education (2024). Read by Grade Three guide.

Mogstad, M., Torgovitsky, A., and Walters, C. R. (2021). The causal interpretation of two-stage least squares with multiple instrumental variables. *American Economic Review*, 111(11):3663–3698.

Mountjoy, J. (2022). Community colleges and upward mobility. *American Economic Review*, 112(8):2580–2611.

North Carolina Department of Public Instruction (2023). Read to Achieve annual report.

Pritchett, L. and Beatty, A. (2015). Slow down, you're going too fast: Matching curricula to student skill levels. *International Journal of Educational Development*, 40:276–288.

Roeber, E. and Briggs, D. C. (2016). An investigation into the validity of the Indiana 2015 ISTEP+ assessment program. Technical report, Indiana State Board of Education.

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.

Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R., Brimijoin, K., Conover, L. A., and Reynolds, T. (2003). Differentiating instruction in response to student readiness, interest, and learning profile in academically diverse classrooms: A review of literature. *Journal for the Education of the Gifted*, 27(2-3):119–145.

Winters, M. A. (2023). The cost of retention under a test-based promotion policy for taxpayers and students. *Educational Evaluation and Policy Analysis*, 45(4):688–695.

Zhong, J. (2026). Early grade retention harms adult earnings. *American Economic Journal: Applied Economics*.

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.

Tables

Table 1: Observed Characteristics by Cohort

	All Administered IREAD				Within 25 Points			
	All	Passed	Failed	Retained	All	Passed	Failed	Retained
	Skipping Cohort: Entered Grade 3 2011-12 through 2015-16							
Male	0.532	0.534	0.528	0.524	0.522	0.520	0.525	0.522
	(0.499)	(0.499)	(0.499)	(0.499)	(0.500)	(0.500)	(0.499)	(0.500)
Black	0.347	0.283	0.460	0.433	0.385	0.352	0.427	0.430
	(0.476)	(0.450)	(0.498)	(0.496)	(0.487)	(0.478)	(0.495)	(0.495)
Hispanic	0.073	0.075	0.070	0.072	0.072	0.071	0.074	0.076
	(0.260)	(0.263)	(0.256)	(0.259)	(0.259)	(0.257)	(0.261)	(0.265)
White	0.504	0.566	0.397	0.426	0.467	0.499	0.425	0.424
	(0.500)	(0.496)	(0.489)	(0.494)	(0.499)	(0.500)	(0.494)	(0.494)
FRL	0.734	0.687	0.817	0.807	0.768	0.741	0.801	0.807
	(0.442)	(0.464)	(0.387)	(0.395)	(0.422)	(0.438)	(0.399)	(0.395)
Retained 3rd	0.301	0.022	0.771	1.000	0.364	0.032	0.791	1.000
	(0.459)	(0.147)	(0.420)	(0.000)	(0.481)	(0.176)	(0.407)	(0.000)
Skipped 4th	0.133	0.003	0.352	0.377	0.166	0.004	0.373	0.453
	(0.340)	(0.054)	(0.478)	(0.485)	(0.372)	(0.065)	(0.484)	(0.498)
N	23,212	14,561	8,651	9,680	10,495	5,905	4,590	3,819
	Remediation Cohort: Entered Grade 3 2016-17 through 2020-21 (Excluding 2019-20)							
Male	0.506	0.509	0.503	0.513	0.498	0.503	0.492	0.505
	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)
Black	0.326	0.244	0.414	0.329	0.318	0.299	0.340	0.309
	(0.469)	(0.430)	(0.493)	(0.470)	(0.466)	(0.458)	(0.474)	(0.463)
Hispanic	0.091	0.085	0.097	0.077	0.092	0.087	0.097	0.070
	(0.287)	(0.279)	(0.296)	(0.266)	(0.289)	(0.282)	(0.296)	(0.256)
White	0.505	0.593	0.410	0.512	0.513	0.538	0.484	0.540
	(0.500)	(0.491)	(0.492)	(0.500)	(0.500)	(0.499)	(0.500)	(0.499)
FRL	0.697	0.639	0.760	0.778	0.695	0.678	0.715	0.782
	(0.459)	(0.480)	(0.427)	(0.416)	(0.460)	(0.467)	(0.452)	(0.413)
Failed summer test	0.495	0.000	1.000	0.920	0.470	0.000	1.000	0.894
	(0.500)	(0.000)	(0.000)	(0.271)	(0.499)	(0.000)	(0.000)	(0.308)
Retained in grade 3	0.052	0.008	0.096	1.000	0.051	0.010	0.097	1.000
	(0.221)	(0.090)	(0.295)	(0.000)	(0.220)	(0.100)	(0.296)	(0.000)
Skipped grade 4	0.003	0.000	0.006	0.045	0.004	0.000	0.008	0.067
	(0.054)	(0.020)	(0.074)	(0.207)	(0.061)	(0.016)	(0.087)	(0.251)
N	43,166	21,845	21,356	3,479	14,853	7,878	6,975	756

Table 2: Tests for Covariate Balance: RD Variation at Threshold

	(1) Male	(2) Black	(3) Hispanic	(4) White	(5) FRL
Skipping Cohort	-0.013	0.021	0.000	-0.024	-0.026
	(0.021)	(0.021)	(0.010)	(0.021)	(0.018)
Remediation Cohort	0.014	0.029*	-0.010	-0.007	0.009
	(0.015)	(0.015)	(0.010)	(0.016)	(0.014)

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. Column labels report the dependent variable. All regressions control for the forcing variable, which is the distance between the student's summer IREAD score and the passing threshold. Robust standard errors reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Net Effect of Failing Summer IREAD: ELA

	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Year 2	Year 3	Year 4	Year 5
					Skipping Cohort				
Net Effect	0.276***	0.133***	0.058**	0.118***	0.056*	0.365	2.617	4.926*	4.583
	(0.029)	(0.027)	(0.026)	(0.032)	(0.034)	(1.699)	(2.351)	(2.813)	(4.103)
Full Treatment	0.455***	0.352***	0.187***	0.299***	0.217***	-0.322	7.251***	6.884***	6.621*
	(0.027)	(0.030)	(0.029)	(0.034)	(0.037)	(1.869)	(2.465)	(2.526)	(3.669)
Retain & Skip	0.000	0.026	-0.007	0.035	-0.024	-0.200	1.333	-3.065	2.251
	(.)	(0.029)	(0.029)	(0.036)	(0.034)	(1.900)	(2.711)	(3.271)	(4.068)
Promoted	-0.055*	-0.090***	-0.020	-0.060	-0.049	-2.197	-4.306	-3.257	-7.356
	(0.032)	(0.031)	(0.034)	(0.038)	(0.045)	(1.943)	(3.070)	(3.570)	(5.160)
					Remediation Cohort				
Net Effect	0.055***	0.053*	0.031	0.102***	0.063*	2.654	1.580	8.148***	10.712
	(0.021)	(0.030)	(0.034)	(0.031)	(0.034)	(2.992)	(4.894)	(2.519)	(7.226)
Full Treatment	0.320***	0.305***	0.077**	0.316***	0.203***	11.212***	6.908	10.549***	13.651**
	(0.030)	(0.036)	(0.038)	(0.037)	(0.039)	(2.815)	(4.337)	(3.032)	(5.827)
Remediation 4th	0.045**	0.035	0.019	0.062**	0.042	1.399	0.565	5.553**	3.384
	(0.020)	(0.030)	(0.034)	(0.030)	(0.033)	(3.261)	(4.925)	(2.456)	(5.446)

Notes: Table reports results from regressions measuring the effect of failing the summer IREAD on ELA scores. For each cohort, the results labeled "Net Effect" are from sharp RD estimates and report the coefficient for the variable indicating if the student failed the summer IREAD, as specified in Equation 4. Results from separate regressions deconstructing patterns by the treatment pathway among students who failed the test, as specified in Equation 5 are reported below the dashed line for the respective cohort. $FullTreatment$ indicates failing student repeated third grade and remained behind their original cohort; $Retain\&Skip$ indicates failing student repeated third grade then rejoined their cohort; $Promoted$ indicates failing student was typically promoted to the fourth grade; and $Remediation4th$ indicates failing student promoted to fourth grade and received third grade remedial reading instruction. Table reports results from both same-grade and same-year comparisons, described at the top of the column. Forcing variable is the distance between the student's summer IREAD score and the passing threshold. All regressions control for student demographic characteristics observed at baseline. Dependent variable for same-grade analyses reported in standard deviation units. Dependent variable for same-year analyses reported in scale scores for the respective test. Same-year samples include only students administered the ISTEP+ for the skipping cohort, and only students administered the ILEARN for the remediation cohort. Regressions 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 4: Net Effect of Failing Summer IREAD: Math

	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Year 2	Year 3	Year 4	Year 5
					Skipping Cohort				
Net Effect	0.277***	0.157***	0.066**	0.058*	0.008	-2.014	-0.723	0.100	-7.572*
	(0.032)	(0.029)	(0.032)	(0.033)	(0.034)	(1.970)	(2.414)	(2.672)	(3.965)
Full Treatment	0.437***	0.383***	0.145***	0.204***	0.157***	-4.299*	2.064	-1.518	-13.323***
	(0.030)	(0.033)	(0.033)	(0.035)	(0.037)	(2.232)	(2.537)	(2.611)	(3.795)
Retain & Skip	0.000	0.007	-0.044	-0.030	-0.069**	-2.395	-2.700	-1.608	-3.681
	(.)	(0.032)	(0.034)	(0.038)	(0.035)	(2.167)	(2.688)	(3.148)	(4.539)
Promoted	0.045	0.049	0.054	-0.053	-0.104**	2.700	-3.090	1.122	-8.923
	(0.035)	(0.036)	(0.041)	(0.040)	(0.049)	(2.457)	(3.141)	(3.682)	(5.694)
					Remediation Cohort				
Net Effect	0.056**	0.133***	0.038	0.049	0.026	10.123**	4.236	4.794*	2.520
	(0.023)	(0.042)	(0.040)	(0.030)	(0.044)	(4.536)	(6.565)	(2.886)	(8.845)
Full Treatment	0.322***	0.322***	0.052	0.269***	0.127***	20.316***	19.198***	19.902***	7.353
	(0.034)	(0.045)	(0.041)	(0.037)	(0.040)	(4.538)	(5.553)	(3.598)	(5.810)
Remediation 4th	0.037	0.109***	0.014	0.003	0.020	8.539*	1.884	0.020	3.036
	(0.023)	(0.040)	(0.037)	(0.029)	(0.038)	(5.109)	(6.114)	(2.845)	(7.402)

Notes: Table reports results from regressions measuring the effect of failing the summer IREAD on math scores. For each cohort, the results labeled "Net Effect" are from sharp RD estimates and report the coefficient for the variable indicating if the student failed the summer IREAD, as specified in Equation 4. Results from separate regressions deconstructing patterns by the treatment pathway among students who failed the test, as specified in Equation 5 are reported below the dashed line for the respective cohort. $FullTreatment$ indicates failing student repeated third grade and remained behind their original cohort; $Retain\&Skip$ indicates failing student repeated third grade then rejoined their cohort; $Promoted$ indicates failing student was typically promoted to the fourth grade; and $Remediation4th$ indicates failing student promoted to fourth grade and received third grade remedial reading instruction. Table reports results from both same-grade and same-year comparisons, described at the top of the column. Forcing variable is the distance between the student's summer IREAD score and the passing threshold. All regressions control for student demographic characteristics observed at baseline. Dependent variable for same-grade analyses reported in standard deviation units. Dependent variable for same-year analyses reported in scale scores for the respective test. Same-year samples include only students administered the ISTEP+ for the skipping cohort, and only students administered the ILEARN for the remediation cohort. Regressions 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 5: Fuzzy RD Estimates for Effect of Grade 3 Retention on Grade Trajectory

	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9
			Years Until Enter Grade			
Retained (IV)	1.007***	0.416***	0.404***	0.397***	0.379***	0.376***
	(0.004)	(0.019)	(0.020)	(0.020)	(0.021)	(0.022)
			Probability Retained in Grade			
Retained (IV)	-0.019*	-0.010***	-0.004	-0.005	-0.000	-0.001
	(0.010)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
			Probability Skipped Grade			
Retained (IV)	0.536***	0.013***	0.001	-0.003	0.011**	
	(0.018)	(0.004)	(0.003)	(0.004)	(0.005)	

Notes: Table reports results from fuzzy RD regressions that use failing the summer IREAD to instrument for probability the student repeated the third grade. Top of the column reports the grade level for which the comparison is made. Outcome described in the sub-header above each row: *Years Until Enter Grade* outcome variable is the number of years between entering the respective grade and the initial third grade year for the student; *Probability Retained in Grade* outcome variable indicates whether the student repeated the respective grade; and *Probability Skipped Grade* outcome variable indicates that in subsequent years the student was observed in the grade prior and then the grade following the respective grade. 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 *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 6: RD Estimates for Effect of Failing Summer IREAD on Probability Retained in Third Grade: Remediation Cohort

	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Fail	0.065***	0.078***	0.089***	0.124***	0.113***
	(0.005)	(0.008)	(0.008)	(0.012)	(0.009)
	Year 2	Year 3	Year 4	Year 5	
Fail	0.088***	0.112***	0.124***	0.130***	
	(0.009)	(0.008)	(0.013)	(0.010)	

Notes: Regressions include eligible students from the remediation cohort observed in given grade or year with summer IREAD score within bandwidth used for analysis measuring net treatment effect for the respective same-grade or same-year comparison in ELA. Dependent variable is an indicator for whether the student repeated the third grade. Running variable is test score difference from the promotion cutoff. Regressions also control for observed baseline student demographic characteristics. Robust standard errors reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Fuzzy RD Estimates for Effect of Third Grade Retention and Being Behind Original Cohort: Skipping Cohort

	Grade 4	Grade 5		Grade 6		Grade 7		Grade 8	
Instrumented:	Retained	Retained	Behind	Retained	Behind	Retained	Behind	Retained	Behind
ELA(Z)	0.437***	0.173***	0.407***	0.086**	0.236***	0.151***	0.355***	0.079*	0.186**
	(0.038)	(0.037)	(0.086)	(0.035)	(0.081)	(0.041)	(0.079)	(0.041)	(0.090)
Math(Z)	0.444***	0.201***	0.452***	0.071*	0.203**	0.100**	0.165*	-0.006	0.000
	(0.055)	(0.039)	(0.089)	(0.038)	(0.093)	(0.044)	(0.085)	(0.042)	(0.101)
	Year 1	Year 2		Year 3		Year 4		Year 5	
ELA	4.015**	0.534	0.859	5.693*	9.480	6.384**	12.529*	5.536	12.338
	(1.897)	(2.316)	(5.310)	(3.262)	(6.668)	(3.199)	(6.434)	(4.706)	(8.859)
Math	1.714	-1.210	-3.462	-1.044	-4.298	0.070	2.842	-9.551**	-11.269
	(2.354)	(2.604)	(5.767)	(3.222)	(6.433)	(3.218)	(6.600)	(4.537)	(8.456)

Notes: Table reports results from fuzzy RD regressions. For each respective grade/year, the left column ("Retained") reports results from regressions that use failing the summer IREAD to instrument for third grade retention, and the right column ("Behind") reports results from regressions that use failing the summer IREAD to instrument for being behind the original third grade cohort. For same-grade analyses, the second approach instruments for the student entering the respective grade a year later than their original cohort, and for same-year estimates it instruments for being a grade level behind their original cohort after the respective number of years. Top panel reports results from same-grade test score comparisons; bottom panel reports results from same-year test score comparisons. Same-grade test scores are standardized by grade and year. Same-year test scores include scores on the ISTEP+ measured on the test's vertically aligned scale. For context, in 2018 a standard deviation on the ISTEP+ scale score in grade three through eight was about 87 in ELA and about 84 in math. 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 8: Test for Covariate Balance: Lag Percentage Retained Students Skip Fourth Grade

	Male	Black	Hispanic	White	FRL	$\hat{\theta}_{ela}$	$\hat{\theta}_{math}$
Grade 5	0.042***	0.131***	0.000	-0.138***	0.017	0.004	-0.032**
	(0.016)	(0.038)	(0.011)	(0.036)	(0.019)	(0.012)	(0.014)
Grade 6	0.048***	0.135***	-0.001	-0.149***	0.031*	0.007	-0.034**
	(0.016)	(0.038)	(0.010)	(0.037)	(0.019)	(0.012)	(0.015)
Grade 7	0.037*	0.101**	-0.002	-0.112***	-0.027	0.020	-0.013
	(0.020)	(0.042)	(0.012)	(0.041)	(0.022)	(0.013)	(0.015)
Grade 8	0.049**	0.133***	-0.009	-0.126***	0.005	-0.004	-0.034**
	(0.022)	(0.042)	(0.012)	(0.040)	(0.023)	(0.013)	(0.015)
Year 2	0.041***	0.124***	-0.000	-0.134***	0.014	0.005	-0.032**
	(0.016)	(0.038)	(0.011)	(0.036)	(0.019)	(0.012)	(0.014)
Year 3	0.046***	0.135***	0.002	-0.147***	0.029	0.006	-0.033**
	(0.018)	(0.038)	(0.011)	(0.037)	(0.021)	(0.012)	(0.014)
Year 4	0.047**	0.112***	-0.005	-0.109***	-0.028	0.019	-0.017
	(0.021)	(0.041)	(0.012)	(0.041)	(0.024)	(0.013)	(0.015)
Year 5	0.031	0.111***	-0.012	-0.112***	0.011	-0.002	-0.031**
	(0.020)	(0.042)	(0.012)	(0.040)	(0.022)	(0.014)	(0.015)

Notes: Table reports results from sharp RD regressions testing balance for the value of covariates relative to the lagged percentage of retained students within the student's initial third grade school who skipped the fourth grade for students who entered third grade during the skipping cohort years. Each cell is the result of a different regression. Each row describes the same-grade or same-year comparison for the estimation sample. Column labels report the dependent variable. Samples include only eligible students with an observed outcome for the respective same-grade or same-year analysis with a summer IREAD score that falls within the bandwidth for the respective outcome analysis. $\hat{\theta}_{ela}$ and $\hat{\theta}_{math}$ are the estimated school value-added in the respective subject measured using data on students in grades 3-5 from 2007-2011 for the school the student attended in their initial third grade year. Forcing variable is the distance between the student's summer IREAD score and the passing threshold. No other control variables are included in the regressions. Robust standard errors clustered by initial-third-grade school reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Multi-Margin Deconstruction: Effect of Full Treatment and Skipping Exemption

			First-Stages				***Impact by Margin***	
	N	Clust.	$Z_{fail} \rightarrow R_{full}$	$Z_{fail} \rightarrow R_{return}$	$Z_{pskip} \rightarrow R_{full}$	$Z_{pskip} \rightarrow R_{return}$	$\hat{\beta}_{full}$	$\hat{\beta}_{return}$
					English Language Arts			
Grade 5	6154	871	0.334***	0.452***	-0.169***	0.132***	0.658***	-0.380***
			(0.024)	(0.024)	(0.014)	(0.019)	(0.143)	(0.118)
Grade 6	6052	687	0.325***	0.438***	-0.155***	0.112***	0.251**	0.098
			(0.019)	(0.020)	(0.012)	(0.015)	(0.115)	(0.094)
Grade 7	4445	522	0.360***	0.385***	-0.208***	0.156***	0.499***	-0.134
			(0.025)	(0.024)	(0.017)	(0.018)	(0.118)	(0.114)
Grade 8	3741	486	0.311***	0.572***	-0.136***	0.193***	0.440***	-0.144
			(0.031)	(0.036)	(0.022)	(0.027)	(0.187)	(0.118)
					Mathematics			
Grade 5	5962	864	0.333***	0.454***	-0.171***	0.128***	0.945***	-0.487***
			(0.023)	(0.025)	(0.015)	(0.019)	(0.160)	(0.132)
Grade 6	5512	676	0.326***	0.441***	-0.170***	0.130***	0.401***	-0.071
			(0.025)	(0.027)	(0.016)	(0.019)	(0.125)	(0.111)
Grade 7	4571	526	0.364***	0.386***	-0.220***	0.162***	0.493***	-0.335***
			(0.026)	(0.027)	(0.019)	(0.021)	(0.137)	(0.124)
Grade 8	4398	510	0.338***	0.543***	-0.128***	0.172***	0.399*	-0.359***
			(0.025)	(0.029)	(0.021)	(0.024)	(0.208)	(0.123)
					English Language Arts			
Year 2	6532	935	0.370***	0.422***	-0.185***	0.133***	13.835**	-13.410**
			(0.025)	(0.025)	(0.015)	(0.018)	(7.093)	(6.593)
Year 3	4661	837	0.396***	0.357***	-0.210***	0.132***	17.685***	-13.848
			(0.027)	(0.026)	(0.019)	(0.019)	(7.553)	(8.837)
Year 4	2860	580	0.448***	0.449***	-0.176***	0.207***	18.266*	-12.267
			(0.040)	(0.042)	(0.032)	(0.033)	(10.405)	(12.365)
Year 5	1858	401	0.584***	0.316***	-0.168***	0.177***	11.139	12.840
			(0.040)	(0.039)	(0.039)	(0.035)	(19.795)	(23.871)
					Mathematics			
Year 2	6577	931	0.377***	0.414***	-0.174***	0.118***	22.674***	-22.367***
			(0.025)	(0.024)	(0.014)	(0.018)	(8.296)	(8.198)
Year 3	4516	828	0.398***	0.355***	-0.208***	0.132***	26.611***	-31.105***
			(0.028)	(0.026)	(0.019)	(0.020)	(8.138)	(9.356)
Year 4	2637	560	0.459***	0.442***	-0.177***	0.215***	19.804**	-20.072
			(0.041)	(0.043)	(0.032)	(0.032)	(10.880)	(12.211)
Year 5	1512	372	0.587***	0.313***	-0.154***	0.176***	-13.900	-28.529
			(0.041)	(0.046)	(0.039)	(0.035)	(22.314)	(35.898)

Notes: Table reports multi-margin RD estimates deconstructing the effect of third-grade retention on same-grade and same-year test scores into two pathways: retention with students remaining behind grade (R_{full}) and retention followed by cohort reentry (R_{return}). The running variable is distance from the promotion cutoff on the summer IREAD test. Instruments include an indicator for failing the cutoff (Z_{fail}) and school-level percent of retained students who skipped ahead the prior year (Z_{skip}). $Z_{fail} \rightarrow R_{full}$ labels first-stage estimates for the effect of Z_{fail} on the likelihood retained and remain behind, and so on. $\hat{\beta}_{full}$ and $\hat{\beta}_{return}$ label the estimated effect of the respective pathway relative to fourth grade promotion. Bootstrapped standard errors clustered at the initial-third-grade school level reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Test for Covariate Balance: Lag Remediation

	Male	Black	Hispanic	White	FRL	$\hat{\theta}_{ela}$	$\hat{\theta}_{math}$
				School-Level Variation			
Grade 6	0.019	0.141**	0.021	-0.180***	0.026	-0.066***	-0.086***
	(0.037)	(0.065)	(0.021)	(0.060)	(0.029)	(0.018)	(0.024)
Grade 7	0.009	0.125*	0.037*	-0.173***	0.061*	-0.069***	-0.092***
	(0.036)	(0.066)	(0.022)	(0.059)	(0.033)	(0.019)	(0.024)
Year 3	0.023	0.133**	0.028	-0.176***	0.031	-0.068***	-0.091***
	(0.036)	(0.066)	(0.021)	(0.060)	(0.028)	(0.018)	(0.024)
Year 4	0.006	0.131**	0.030	-0.175***	0.037	-0.069***	-0.091***
	(0.037)	(0.066)	(0.022)	(0.059)	(0.033)	(0.018)	(0.024)
Year 5	0.016	0.107	0.040*	-0.165***	0.049	-0.069***	-0.092***
	(0.038)	(0.068)	(0.023)	(0.060)	(0.033)	(0.018)	(0.024)
				District-Level Variation			
Grade 6	-0.004	0.132	0.081***	-0.234**	0.013	-0.019	-0.073**
	(0.042)	(0.108)	(0.023)	(0.106)	(0.037)	(0.030)	(0.036)
Grade 7	-0.002	0.134	0.087***	-0.242**	0.051	-0.017	-0.070*
	(0.041)	(0.114)	(0.024)	(0.108)	(0.039)	(0.030)	(0.036)
Year 3	0.007	0.131	0.084***	-0.231**	0.018	-0.019	-0.075**
	(0.042)	(0.109)	(0.023)	(0.107)	(0.037)	(0.030)	(0.036)
Year 4	-0.013	0.137	0.080***	-0.239**	0.042	-0.017	-0.070*
	(0.042)	(0.112)	(0.024)	(0.108)	(0.043)	(0.030)	(0.036)
Year 5	0.010	0.116	0.091***	-0.228**	0.015	-0.017	-0.069*
	(0.039)	(0.116)	(0.025)	(0.107)	(0.040)	(0.031)	(0.036)

Notes: Table reports results from regressions testing balance for the value of covariates relative to the lagged percentage of failing students within the student's initial third grade school who were promoted and received remedial instruction in the fourth grade with an exemption for students who entered third grade during the remediation cohort years. Each cell is the result of a different regression. Table reports results from regressions where the instrument (Z_{prem}) measures percent of failing students who received an exemption within the student's school or district, respectively. Column labels report the dependent variable. Samples include only eligible students with an observed outcome for the respective same-grade or same-year analysis with a summer IREAD score that falls within the bandwidth for the respective outcome analysis. $\hat{\theta}_{ela}$ and $\hat{\theta}_{math}$ are the estimated school or district value-added in the respective subject measured using data on students in grades 3-5 from 2007-2011 for the school the student attended in their initial third grade year. No other control variables are included in the regressions. Robust standard errors reported in parentheses. Regressions that employ school-level variation cluster standard errors at the school level, and regressions that employ district-level variation cluster standard errors at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Reduced-Form Relationship Between Z_{prem} and Outcomes on Either Side of Passing Threshold

	Grade 6	Grade 7	Year 3	Year 4	Year 5
School-Level Variation					
English Language Arts					
Fail IREAD	0.208***	0.168***	10.226***	7.937**	26.263*
	(0.050)	(0.053)	(3.906)	(3.936)	(15.739)
Z_{prem}	0.038	0.055	3.739	6.733*	11.815
	(0.046)	(0.044)	(3.653)	(3.758)	(14.912)
$Fail \times Z_{prem}$	-0.245***	-0.169***	-13.336***	-9.708**	-12.544
	(0.062)	(0.064)	(4.792)	(4.754)	(15.622)
Mathematics					
Fail IREAD	0.175***	0.134**	16.347***	10.366*	13.133
	(0.053)	(0.054)	(4.862)	(5.341)	(8.651)
Z_{prem}	0.001	0.018	0.634	3.449	3.331
	(0.051)	(0.044)	(5.038)	(4.548)	(5.485)
$Fail \times Z_{prem}$	-0.249***	-0.144**	-24.211***	-13.065**	-20.322
	(0.066)	(0.064)	(6.168)	(6.471)	(13.087)
District-Level Variation					
English Language Arts					
Fail IREAD	0.243***	0.229***	11.862**	10.957**	14.805
	(0.070)	(0.068)	(5.236)	(4.676)	(12.485)
Z_{prem}	0.045	0.076	5.188	9.711**	-8.809
	(0.064)	(0.055)	(4.969)	(4.655)	(14.192)
$Fail \times Z_{prem}$	-0.267***	-0.243***	-14.176**	-12.697**	5.869
	(0.097)	(0.091)	(6.997)	(6.215)	(16.702)
Mathematics					
Fail IREAD	0.233***	0.172**	21.952***	13.281*	15.611
	(0.069)	(0.073)	(6.666)	(6.775)	(10.669)
Z_{prem}	0.015	0.045	1.746	6.938	7.284
	(0.065)	(0.053)	(6.326)	(5.437)	(6.286)
$Fail \times Z_{prem}$	-0.308***	-0.185**	-29.543***	-15.911*	-22.016
	(0.085)	(0.087)	(8.251)	(8.295)	(14.722)

Notes: Table reports results of regressions investigating relationship between Z_{prem} and student outcomes for students who passed or failed summer IREAD. Sample includes eligible students from remediation cohort with summer IREAD score within optimal bandwidth for respective analysis measuring the net effect of failing the test on later outcomes. All regressions control for third grade entry year, observed baseline demographic characteristics, and difference between student's summer IREAD score and the passing threshold. Outcomes for same-grade comparisons reported in standard deviation units, and outcomes for same-year comparisons reported in scale scores on the ILEARN test. Standard errors reported in parentheses clustered at school-level for analyses that exploit school-level variation and clustered at district level for analyses that exploit district-level variation. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: IV Estimates for the LATE of Retention Relative to Promotion for Retained Students

	Grade 6		Grade 7		Year 3		Year 4		Year 5	
District-Level Variation										
English Language Arts										
Retained (IV)	0.695***	0.702***	0.583***	0.596***	30.724***	28.993**	14.670	15.108	12.605	7.165
	(0.122)	(0.170)	(0.114)	(0.155)	(10.253)	(12.391)	(9.325)	(13.524)	(11.979)	(14.372)
$\hat{\theta}_{ela}$		0.792***		0.505**		67.304***		49.115***		45.183*
		(0.219)		(0.239)		(16.062)		(18.682)		(26.006)
Mathematics										
Retained (IV)	0.794***	0.769***	0.611***	0.564***	74.032***	69.169***	38.069***	34.759**	50.049**	48.503
	(0.163)	(0.210)	(0.122)	(0.164)	(14.893)	(17.787)	(13.211)	(17.308)	(21.676)	(32.069)
$\hat{\theta}_{math}$		0.958***		0.568***		96.692***		57.606***		73.142
		(0.241)		(0.194)		(22.445)		(17.246)		(83.948)
School-Level Variation										
English Language Arts										
Retained (IV)	0.800***	0.836***	0.459***	0.500***	38.904**	39.136**	11.736	14.798	6.425	3.638
	(0.185)	(0.239)	(0.173)	(0.185)	(15.343)	(18.847)	(14.018)	(15.923)	(16.597)	(20.703)
$\hat{\theta}_{ela}$		0.022		-0.061		1.727		-0.530		-3.560
		(0.154)		(0.145)		(11.791)		(12.611)		(17.742)
Mathematics										
Retained (IV)	0.906***	0.889***	0.493***	0.492***	87.184***	84.033***	34.064*	33.742	58.106	45.364
	(0.233)	(0.300)	(0.170)	(0.180)	(21.956)	(27.059)	(18.157)	(20.910)	(39.265)	(32.542)
$\hat{\theta}_{math}$		0.456***		0.083		47.909***		13.723		74.204
		(0.175)		(0.121)		(15.822)		(13.234)		(47.630)

Standard errors in parentheses

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

Notes: Table reports results from IV regressions measuring the effect of retention relative to fourth-grade promotion. Column labels report the same-grade or same-year comparison. For each comparison, the table reports results from regressions without and with a control for estimated school or district value-added in the respective subject measured using data on students in grades 3-5 from 2007-2011 for the district/school the student attended in their initial third grade year ($\hat{\theta}_{ela}$ and $\hat{\theta}_{math}$). All regressions also control for observed student baseline characteristics, an indicator for year entered third grade, and the difference between the student's score on the summer IREAD and the passing threshold. Outcomes for same-grade comparisons reported in standard deviation units, and outcomes for same-year comparisons reported in scale scores on the ILEARN test. Robust standard errors reported in parentheses. Regressions that employ school-level variation cluster standard errors at the school level, and regressions that employ district-level variation cluster standard errors at the district level. * $p < 0.10$, **

Online Appendix

A Exhibits

Table A.1: Third-Grade Reading Retention Policy Characteristics by State

	Timeline			Exemptions									Stringency	
State	First	Expand	Limit	Any	Promo	IEP	ELL	Prev Ret	Alt Assess	Port	Sum Sch	Prof Judg	Remed	Mand Ret
Alabama	2021	0	0	1	1	1	1	1	0	1	0	0	1	1
Arizona	2010	0	1	1	1	1	1	1	1	1	0	0	1	1
Arkansas	2009	1	0	1	1	1	0	0	0	0	0	0	1	1
Delaware	2012	0	0	1	1	1	1	1	1	1	0	0	1	1
Florida	2002	1	0	1	1	1	1	1	1	1	1	0	1	1
Indiana	2012	1	1	1	1	1	1	1	0	0	1	0	1	1
Iowa	2012	0	0	1	1	0	0	0	0	0	1	0	1	1
Mississippi	2013	1	1	1	1	1	1	1	1	1	0	0	1	1
Missouri	2012	0	0	1	1	1	0	0	0	0	0	0	1	1
Nevada	2015	1	0	1	1	0	0	0	0	0	0	1	1	0
N. Carolina	2012	1	0	1	1	1	1	0	1	1	1	0	1	1
Ohio	2012	0	1	1	1	1	1	1	1	0	0	0	1	0
Oklahoma	2012	1	0	1	1	0	0	1	1	1	0	1	1	0
S. Carolina	2014	1	0	1	1	1	1	1	1	1	0	0	1	1
Tennessee	2012	1	1	1	1	1	1	0	1	0	0	1	1	0
W. Virginia	2013	0	0	1	1	1	0	0	0	0	1	0	1	1
Wisconsin	2012	0	0	1	1	0	0	0	0	0	0	0	1	1
DC	2014	0	0	0	0	0	0	0	0	0	0	0	1	1

Notes: Binary indicators (0/1) denote absence/presence of features. Timeline: First = year of first policy; Expand = policy expanded after adoption; Limit = policy limited after adoption. Exemptions: Promo = promotion; Prev Ret = previously retained; Alt Assess = alternative assessment; Port = portfolio; Sum Sch = summer school; Prof Judg = professional judgment. Stringency: Remed = requires remediation; Mand Ret = mandates retention. Data reflect most expansive policy version.

Table A.2: Map of Cohort Timelines

Year	2012		2013		2014		2015		2016		2017		2018		2019		2020		2021		2022		2023	
	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P
2012	3	3																						
2013	3	4	3	3																				
2014	4	5	3	4	3	3																		
2015	5	6	4	5	3	4	3	3																
2016	6	7	5	6	4	5	3	4	3	3														
2017	7	8	6	7	5	6	4	5	3	4	3	3												
2018	8		7	8	6	7	5	6	4	5	3	4	3	3										
2019			8		7	8	6	7	5	6	4	5	3	4	3	3								
2020					8		7	8	6	7	5	6	4	5	3	4	3	3						
2021							8		7	8	6	7	5	6	4	5	3	4	3	3				
2022									8		7	8	6	7	5	6	4	5	3	4	3	3		
2023											8		7	8	6	7	5	6	4	5	3	4	3	3
2024													8		7	8	6	7	5	6	4	5	3	4

Notes: Numbers indicate grade level for students cumulatively promoted in subsequent years after potential treatment by group and entering third grade year. R = Retained cohort; P = Promoted cohort. Year (rows) indicates the academic year of observation; year (columns) indicates the cohort's third-grade year.

Table A.3: Entering Third Grade Cohorts Included in Estimation Samples

	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Year 2	Year 3	Year 4	Year 5
Net Effect of Failing Summer IREAD									
2011-12	✓	✓	✓	✓	✓	✓	✓	✓	✓
2012-13	✓	✓	✓	✓	✓	✓	✓	✓	✓
2013-14	✓	✓	✓	✓	✓	✓	✓	✓	
2014-15	✓	✓	✓	✓	✓	✓	✓		
2015-16	✓	✓	✓	✓	✓	✓			
2016-17	✓	✓	✓	✓	✓	✓		✓	✓
2017-18	✓	✓	✓	✓	✓		✓	✓	✓
2018-19	✓	✓	✓	✓	✓	✓	✓	✓	✓
2019-20									
2020-21	✓	✓	✓			✓	✓		
2021-22	✓	✓				✓			
2022-23	✓								
Fuzzy RD: Skipping Cohort									
2011-12	✓	✓	✓	✓	✓	✓	✓	✓	✓
2012-13	✓	✓	✓	✓	✓	✓	✓	✓	✓
2013-14	✓	✓	✓	✓	✓	✓	✓	✓	
2014-15	✓	✓	✓	✓	✓	✓	✓		
2015-16	✓	✓	✓	✓	✓	✓			
Multi-Margin Deconstruction: Skipping Cohort									
2011-12									
2012-13	✓	✓	✓	✓	✓				
2013-14	✓	✓	✓	✓	✓				
2014-15	✓	✓	✓	✓	✓	✓	✓	✓	
2015-16	✓	✓	✓	✓	✓	✓	✓		✓
IV Effect of Retention vs Remediation: Remediation Cohort									
2016-17									
2017-18			✓	✓	✓		✓	✓	✓
2018-19			✓	✓	✓		✓	✓	✓
2019-20									
2020-21									
2021-22									
2022-23									

Table A.4: Fuzzy RD Estimates for Effect of Third Grade Retention and Being Behind Original Cohort for Skipping Cohort: MSETWO Bandwidths + Triangular Kernel

	Grade 4	Grade 5		Grade 6		Grade 7		Grade 8	
		Retained	Behind	Retained	Behind	Retained	Behind	Retained	Behind
ELA(Z)	0.451***	0.171***	0.412***	0.095***	0.230***	0.165***	0.341***	0.083**	0.197**
	(0.049)	(0.036)	(0.085)	(0.037)	(0.086)	(0.039)	(0.080)	(0.040)	(0.097)
Math(Z)	0.433***	0.205***	0.482***	0.070*	0.179*	0.084**	0.182**	0.001	0.010
	(0.055)	(0.038)	(0.085)	(0.042)	(0.100)	(0.042)	(0.091)	(0.045)	(0.102)
	Year 1	Year 2		Year 3		Year 4		Year 5	
ELA	4.452**	0.118	0.268	3.552	7.070	6.155*	12.989**	5.648	9.899
	(1.830)	(2.275)	(5.144)	(3.335)	(6.571)	(3.147)	(6.437)	(5.052)	(8.765)
Math	1.829	-1.248	-2.470	-1.767	-3.403	-0.651	-0.516	-9.194*	-15.881*
	(2.265)	(2.550)	(5.822)	(3.344)	(6.592)	(3.013)	(6.139)	(4.695)	(8.385)

Notes: Table reports results fuzzy RD regressions. For each respective grade/year, the left column ("Retained") reports results from models that use failing the summer IREAD to instrument for third grade retention, and the right column ("Behind") reports results from models that use failing the summer IREAD to instrument for being behind the original third grade cohort. For same-grade analyses, the second approach instruments for the student entering the respective grade a year later than their original cohort, and for same-year estimates it instruments for being a grade level behind their original cohort after the respective number of years. Top panel reports results from same-grade test score comparisons; bottom panel reports results from same-year test score comparisons. Same-grade test scores are standardized by grade and year. Same-year test scores include scores on the ISTEP+ measured on the test's vertically aligned scale. For context, in 2018 a standard deviation on the ISTEP+ scale score in grade three through eight was about 87 in ELA and about 84 in math. 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 msetwo optimal bandwidths, and a triangular kernel. Robust standard errors reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Fuzzy RD Estimates for Effect of Third Grade Retention and Being Behind Original Cohort for Skipping Cohort: MSERD Bandwidths + Uniform Kernel

	Grade 4	Grade 5		Grade 6		Grade 7		Grade 8	
		Retained	Behind	Retained	Behind	Retained	Behind	Retained	Behind
ELA(Z)	0.448***	0.176***	0.403***	0.085***	0.208***	0.151***	0.321***	0.089**	0.218***
	(0.035)	(0.032)	(0.073)	(0.033)	(0.072)	(0.039)	(0.074)	(0.036)	(0.085)
Math(Z)	0.430***	0.213***	0.496***	0.082**	0.146*	0.071**	0.155*	0.010	0.013
	(0.042)	(0.036)	(0.083)	(0.035)	(0.079)	(0.036)	(0.080)	(0.040)	(0.095)
	Year 1	Year 2		Year 3		Year 4		Year 5	
ELA	6.714***	0.357	0.761	2.661	5.059	5.154	7.238	8.553*	14.698*
	(1.429)	(2.015)	(4.298)	(2.542)	(4.981)	(3.193)	(5.922)	(4.685)	(8.678)
Math	2.787	-1.767	-3.774	0.101	0.196	-2.762	-1.215	-10.129**	-17.256**
	(1.927)	(2.359)	(5.038)	(3.032)	(5.874)	(2.864)	(6.217)	(4.554)	(7.765)

Notes: Table reports results fuzzy RD regressions. For each respective grade/year, the left column ("Retained") reports results from models that use failing the summer IREAD to instrument for third grade retention, and the right column ("Behind") reports results from models that use failing the summer IREAD to instrument for being behind the original third grade cohort. For same-grade analyses, the second approach instruments for the student entering the respective grade a year later than their original cohort, and for same-year estimates it instruments for being a grade level behind their original cohort after the respective number of years. Top panel reports results from same-grade test score comparisons; bottom panel reports results from same-year test score comparisons. Same-grade test scores are standardized by grade and year. Same-year test scores include scores on the ISTEP+ measured on the test's vertically aligned scale. For context, in 2018 a standard deviation on the ISTEP+ scale score in grade three through eight was about 87 in ELA and about 84 in math. 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 msetwo optimal bandwidths, and a triangular kernel. Robust standard errors reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: Fuzzy RD Estimates for Effect of Third Grade Retention and Being Behind Original Cohort for Skipping Cohort: MSERD Bandwidths + Triangular Kernel

	Grade 4	Grade 5		Grade 6		Grade 7		Grade 8	
		Retained	Behind	Retained	Behind	Retained	Behind	Retained	Behind
ELA(Z)	0.438***	0.177***	0.391***	0.086***	0.169**	0.157***	0.324***	0.084**	0.203**
	(0.040)	(0.032)	(0.063)	(0.032)	(0.066)	(0.037)	(0.075)	(0.038)	(0.092)
Math(Z)	0.432***	0.205***	0.467***	0.068*	0.167*	0.075**	0.153*	-0.003	0.012
	(0.041)	(0.037)	(0.078)	(0.039)	(0.089)	(0.036)	(0.081)	(0.045)	(0.099)
	Year 1	Year 2		Year 3		Year 4		Year 5	
ELA	5.683***	0.546	1.254	3.257	5.598	6.058*	12.547**	5.733	10.200
	(1.655)	(1.843)	(4.317)	(2.894)	(5.164)	(3.124)	(6.306)	(4.983)	(8.611)
Math	2.507	-1.816	-3.275	-0.556	-0.908	-0.680	-1.205	-9.181**	-15.894*
	(2.075)	(2.166)	(5.065)	(2.950)	(5.698)	(3.010)	(6.051)	(4.681)	(8.383)

Notes: Table reports results fuzzy RD regressions. For each respective grade/year, the left column ("Retained") reports results from models that use failing the summer IREAD to instrument for third grade retention, and the right column ("Behind") reports results from models that use failing the summer IREAD to instrument for being behind the original third grade cohort. For same-grade analyses, the second approach instruments for the student entering the respective grade a year later than their original cohort, and for same-year estimates it instruments for being a grade level behind their original cohort after the respective number of years. Top panel reports results from same-grade test score comparisons; bottom panel reports results from same-year test score comparisons. Same-grade test scores are standardized by grade and year. Same-year test scores include scores on the ISTEP+ measured on the test's vertically aligned scale. For context, in 2018 a standard deviation on the ISTEP+ scale score in grade three through eight was about 87 in ELA and about 84 in math. 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 msetwo optimal bandwidths, and a triangular kernel. Robust standard errors reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: Test for Covariate Balance: Lag Percentage Retained Students Skip Fourth Grade, Control for Summer IREAD Score

	Male	Black	Hispanic	White	FRL	$\hat{\theta}_{ela}$	$\hat{\theta}_{math}$
Grade 5	0.042***	0.133***	0.000	-0.140***	0.018	0.004	-0.032**
	(0.016)	(0.037)	(0.011)	(0.035)	(0.018)	(0.012)	(0.014)
Grade 6	0.048***	0.135***	-0.001	-0.149***	0.031*	0.007	-0.034**
	(0.016)	(0.038)	(0.010)	(0.036)	(0.019)	(0.012)	(0.014)
Grade 7	0.037*	0.104**	-0.001	-0.115***	-0.025	0.019	-0.014
	(0.020)	(0.041)	(0.012)	(0.040)	(0.022)	(0.013)	(0.015)
Grade 8	0.049**	0.132***	-0.009	-0.125***	0.004	-0.004	-0.034**
	(0.022)	(0.041)	(0.012)	(0.039)	(0.023)	(0.013)	(0.015)
Year 2	0.041***	0.125***	-0.000	-0.135***	0.015	0.005	-0.032**
	(0.016)	(0.037)	(0.011)	(0.035)	(0.018)	(0.012)	(0.014)
Year 3	0.046***	0.139***	0.002	-0.151***	0.031	0.005	-0.033**
	(0.017)	(0.038)	(0.011)	(0.036)	(0.020)	(0.012)	(0.014)
Year 4	0.047**	0.116***	-0.006	-0.113***	-0.026	0.018	-0.018
	(0.021)	(0.041)	(0.012)	(0.041)	(0.024)	(0.013)	(0.015)
Year 5	0.031	0.110***	-0.012	-0.111***	0.011	-0.002	-0.031**
	(0.020)	(0.041)	(0.012)	(0.039)	(0.022)	(0.014)	(0.015)

Notes: Table reports results from sharp RD regressions testing balance for the value of covariates relative to the lagged percentage of retained students within the student's initial third grade school who skipped the fourth grade for students who entered third grade during the skipping cohort years. Regressions differ from those reported in Table 8 by adding a control for the difference between the student's summer IREAD score and the passing threshold. Each cell is the result of a different regression. Each row describes the same-grade or same-year comparison for the estimation sample. Column labels report the dependent variable. Samples include only eligible students with an observed outcome for the respective same-grade or same-year analysis with a summer IREAD score that falls within the bandwidth for the respective outcome analysis. $\hat{\theta}_{ela}$ and $\hat{\theta}_{math}$ are the estimated school value-added in the respective subject measured using data on students in grades 3-5 from 2007-2011 for the school the student attended in their initial third grade year. Forcing variable is the distance between the student's summer IREAD score and the passing threshold. No other control variables are included in the regressions. Robust standard errors clustered by initial-third-grade school reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.8: Mountjoy Deconstruction with School Fixed-Effects

			First-Stages				***Impact by Margin***	
	N	Clust.	$Z_{fail} \rightarrow R_{full}$	$Z_{fail} \rightarrow R_{return}$	$Z_{pskip} \rightarrow R_{full}$	$Z_{pskip} \rightarrow R_{return}$	$\hat{\beta}_{full}$	$\hat{\beta}_{return}$
					English Language Arts			
Grade 5	6154	871	0.322***	0.475***	-0.093***	0.038***	0.796	-0.394
			(0.024)	(0.025)	(0.015)	(0.021)	(0.488)	(0.384)
Grade 6	6052	687	0.319***	0.453***	-0.077***	0.011***	0.127	0.191
			(0.019)	(0.021)	(0.013)	(0.018)	(0.771)	(0.554)
Grade 7	4445	522	0.344***	0.411***	-0.131***	0.105***	0.574***	-0.065
			(0.024)	(0.024)	(0.017)	(0.023)	(0.204)	(0.191)
Grade 8	3741	486	0.301***	0.590***	-0.100***	0.146***	0.733**	-0.200
			(0.032)	(0.037)	(0.022)	(0.028)	(0.289)	(0.172)
					Mathematics			
Grade 5	5962	864	0.321***	0.480***	-0.099***	0.039***	1.061***	-0.469*
			(0.023)	(0.025)	(0.017)	(0.019)	(0.401)	(0.276)
Grade 6	5512	676	0.314***	0.464***	-0.104***	0.051***	0.368	-0.042
			(0.025)	(0.028)	(0.018)	(0.022)	(0.241)	(0.187)
Grade 7	4571	526	0.352***	0.409***	-0.149***	0.119***	0.528***	-0.287*
			(0.026)	(0.027)	(0.021)	(0.025)	(0.190)	(0.171)
Grade 8	4398	510	0.329***	0.556***	-0.084***	0.119***	0.654*	-0.418**
			(0.025)	(0.030)	(0.021)	(0.025)	(0.336)	(0.197)
					English Language Arts			
Year 2	6532	935	0.341***	0.456***	-0.075***	0.020***	20.741	-12.992
			(0.025)	(0.026)	(0.014)	(0.019)	(31.527)	(24.458)
Year 3	4661	837	0.249***	0.459***	-0.078***	0.031***	94.197**	-48.730*
			(0.025)	(0.028)	(0.016)	(0.022)	(47.717)	(28.547)
Year 4	2860	580	0.330***	0.566***	-0.069***	0.148***	35.095	-12.819
			(0.038)	(0.043)	(0.029)	(0.036)	(36.050)	(23.373)
Year 5	1858	401	0.552***	0.362***	-0.191***	0.184***	29.894	-21.909
			(0.041)	(0.042)	(0.043)	(0.039)	(31.821)	(36.456)
					Mathematics			
Year 2	6577	931	0.344***	0.449***	-0.052***	-0.009***	43.527	-33.005
			(0.024)	(0.024)	(0.013)	(0.018)	(509.718)	(440.759)
Year 3	4516	828	0.251***	0.459***	-0.074***	0.029***	113.789*	-64.453*
			(0.026)	(0.029)	(0.017)	(0.023)	(71.957)	(39.990)
Year 4	2637	560	0.331***	0.567***	-0.064***	0.156***	26.681	-10.146
			(0.039)	(0.044)	(0.031)	(0.036)	(64.277)	(37.753)
Year 5	1512	372	0.544***	0.372***	-0.189***	0.173***	-9.297	-46.913
			(0.046)	(0.051)	(0.046)	(0.043)	(30.741)	(46.289)

Notes: Table reports multi-margin RD estimates deconstructing the effect of third-grade retention on same-grade and same-year test scores into two pathways: retention with students remaining behind grade (R_{remain}) and retention followed by cohort reentry (R_{return}). Regressions include a fixed effect for school attended during initial third grade year. The running variable is distance from the promotion cutoff on the summer IREAD test. Instruments include an indicator for failing the cutoff (Z_{fail}) and school-level percent of retained students who skipped ahead the prior year (Z_{skip}). $Z_{fail} \rightarrow R_{full}$ labels first-stage estimates for the effect of Z_{fail} on the likelihood retained and remain behind, and so on. $\hat{\beta}_{full}$ and $\hat{\beta}_{return}$ label the estimated effect of the respective pathway relative to fourth grade promotion. Bootstrapped standard errors clustered at the school level reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.9: Mountjoy Deconstruction Controlling for Pre-Policy School Value-Added

			First-Stages				***Impact by Margin***	
	N	Clust.	$Z_{fail} \rightarrow R_{full}$	$Z_{fail} \rightarrow R_{return}$	$Z_{pskip} \rightarrow R_{full}$	$Z_{pskip} \rightarrow R_{return}$	$\hat{\beta}_{full}$	$\hat{\beta}_{return}$
					English Language Arts			
Grade 5	6154	871	0.337	0.447	-0.166	0.130	0.689***	-0.392***
			(0.022)	(0.024)	(0.014)	(0.019)	(0.142)	(0.113)
Grade 6	6052	687	0.329	0.431	-0.152	0.110	0.265**	0.068
			(0.019)	(0.020)	(0.013)	(0.017)	(0.120)	(0.096)
Grade 7	4445	522	0.360	0.382	-0.205	0.157	0.513***	-0.168
			(0.027)	(0.026)	(0.019)	(0.019)	(0.105)	(0.124)
Grade 8	3741	486	0.313	0.570	-0.135	0.191	0.420**	-0.160
			(0.033)	(0.038)	(0.023)	(0.027)	(0.188)	(0.116)
					Mathematics			
Grade 5	5962	864	0.335	0.450	-0.170	0.127	0.863***	-0.435***
			(0.025)	(0.024)	(0.016)	(0.020)	(0.173)	(0.143)
Grade 6	5512	676	0.328	0.435	-0.170	0.130	0.358***	-0.056
			(0.024)	(0.025)	(0.017)	(0.019)	(0.133)	(0.114)
Grade 7	4571	526	0.367	0.379	-0.221	0.164	0.474***	-0.328**
			(0.028)	(0.024)	(0.019)	(0.021)	(0.129)	(0.128)
Grade 8	4398	510	0.341	0.537	-0.127	0.165	0.329*	-0.348***
			(0.027)	(0.032)	(0.019)	(0.022)	(0.186)	(0.135)
					English Language Arts			
Year 2	6532	935	0.372	0.418	-0.181	0.130	14.964**	-14.426**
			(0.025)	(0.024)	(0.015)	(0.017)	(6.922)	(6.824)
Year 3	4661	837	0.399	0.352	-0.208	0.133	20.510***	-16.861*
			(0.030)	(0.026)	(0.020)	(0.021)	(7.206)	(9.049)
Year 4	2860	580	0.455	0.443	-0.174	0.206	18.217*	-14.671
			(0.043)	(0.042)	(0.032)	(0.031)	(10.310)	(12.836)
Year 5	1858	401	0.594	0.305	-0.173	0.173	12.454	6.804
			(0.045)	(0.046)	(0.042)	(0.035)	(20.054)	(27.700)
					Mathematics			
Year 2	6577	931	0.379	0.410	-0.172	0.115	18.021**	-18.776**
			(0.024)	(0.023)	(0.015)	(0.018)	(8.251)	(8.313)
Year 3	4516	828	0.402	0.348	-0.209	0.134	25.132***	-29.528***
			(0.029)	(0.027)	(0.018)	(0.020)	(7.921)	(10.296)
Year 4	2637	560	0.467	0.433	-0.180	0.215	15.548	-18.692
			(0.041)	(0.042)	(0.033)	(0.034)	(9.761)	(11.806)
Year 5	1512	372	0.599	0.300	-0.161	0.174	-15.984	-30.479
			(0.043)	(0.044)	(0.043)	(0.033)	(20.564)	(34.251)

Notes: Table reports multi-margin RD estimates deconstructing the effect of third-grade retention on same-grade and same-year test scores into two pathways: retention with students remaining behind grade (R_{remain}) and retention followed by cohort reentry (R_{return}). Regressions include control for pre-policy value-added in respective subjects for school attended during initial third grade year. The running variable is distance from the promotion cutoff on the summer IREAD test. Instruments include an indicator for failing the cutoff (Z_{fail}) and school-level percent of retained students who skipped ahead the prior year (Z_{skip}). $Z_{fail} \rightarrow R_{full}$ labels first-stage estimates for the effect of Z_{fail} on the likelihood retained and remain behind, and so on. $\hat{\beta}_{full}$ and $\hat{\beta}_{return}$ label the estimated effect of the respective pathway relative to fourth grade promotion. Bootstrapped standard errors clustered at the school level reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.10: First Stage Estimates: Using Z_{pret} to Instrument for Third Grade Retention

	Grade 6		Grade 7		Year 3		Year 4		Year 5	
English Language Arts										
School	-0.270***	-0.226**	-0.273***	-0.226**	-0.284***	-0.244***	-0.270***	-0.231**	-0.278***	-0.230**
	(0.066)	(0.089)	(0.065)	(0.095)	(0.066)	(0.093)	(0.066)	(0.093)	(0.067)	(0.098)
F-stat	17.02	6.41	17.52	5.68	18.65	6.84	16.80	6.23	17.18	5.48
District	-0.422***	-0.400***	-0.405***	-0.379***	-0.428***	-0.407***	-0.421***	-0.402***	-0.409***	-0.380***
	(0.049)	(0.094)	(0.050)	(0.092)	(0.050)	(0.093)	(0.050)	(0.097)	(0.051)	(0.093)
F-stat	73.52	18.10	65.49	16.81	73.93	19.28	72.27	17.11	65.34	16.81
Mathematics										
School	-0.276***	-0.234**	-0.276***	-0.230**	-0.287***	-0.249***	-0.273***	-0.234**	-0.282***	-0.238**
	(0.066)	(0.092)	(0.066)	(0.096)	(0.066)	(0.095)	(0.067)	(0.095)	(0.067)	(0.101)
F-stat	17.35	6.49	17.63	5.71	18.80	6.94	16.66	6.13	17.53	5.52
District	-0.424***	-0.410***	-0.410***	-0.390***	-0.429***	-0.417***	-0.423***	-0.409***	-0.414***	-0.396***
	(0.050)	(0.092)	(0.051)	(0.091)	(0.050)	(0.090)	(0.050)	(0.095)	(0.051)	(0.091)
F-stat	73.24	19.76	65.57	18.44	73.58	21.23	70.52	18.56	66.72	18.73

Figure A.1: Map of Student Grade Progression for Skipping Cohort

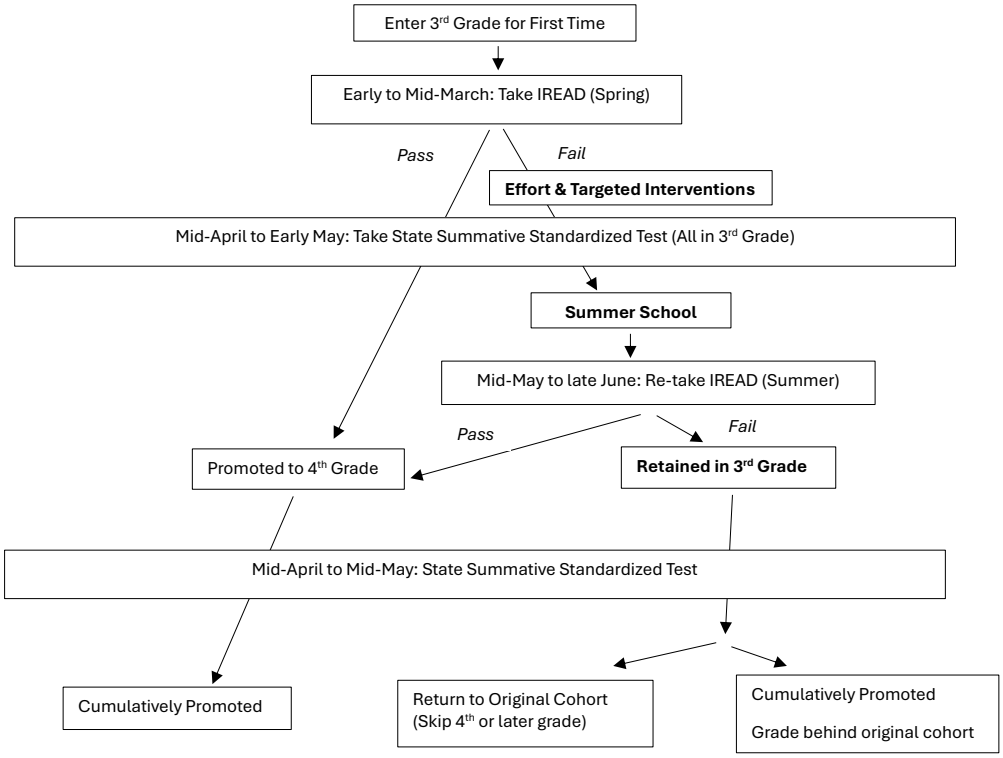


Figure A.2: Map of Student Grade Progression for Remediation Cohort

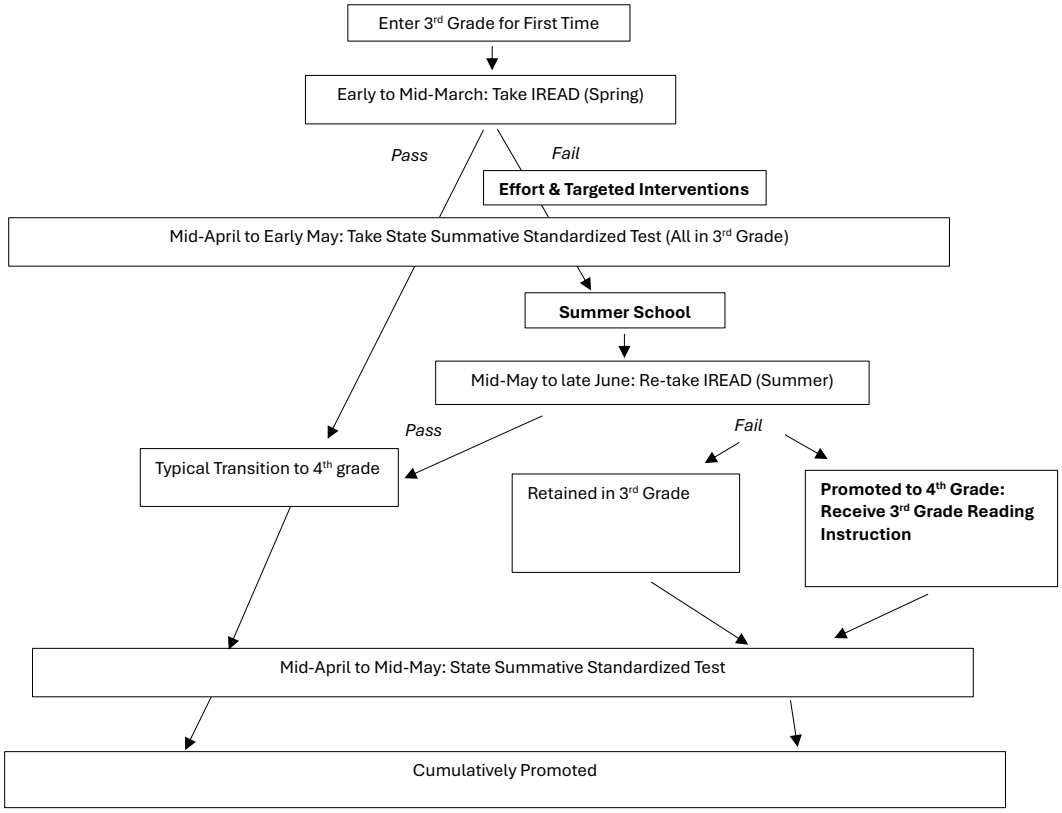
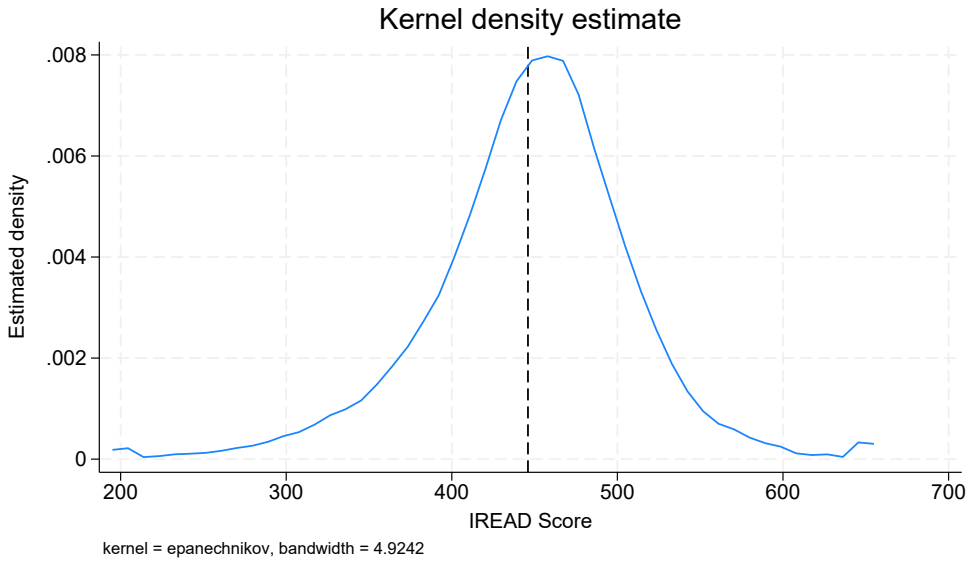
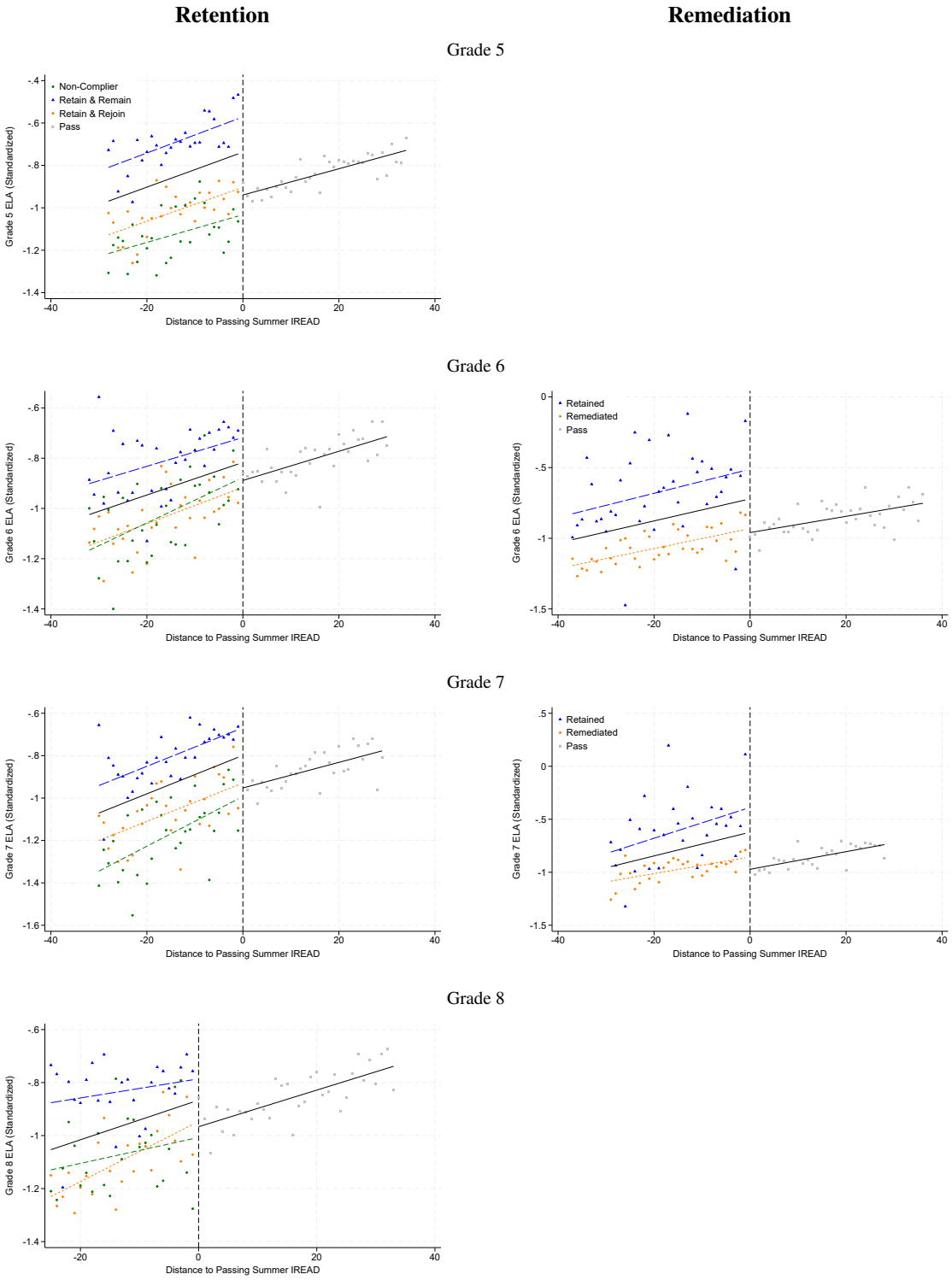


Figure A.3: Distribution of Summer IREAD Scores



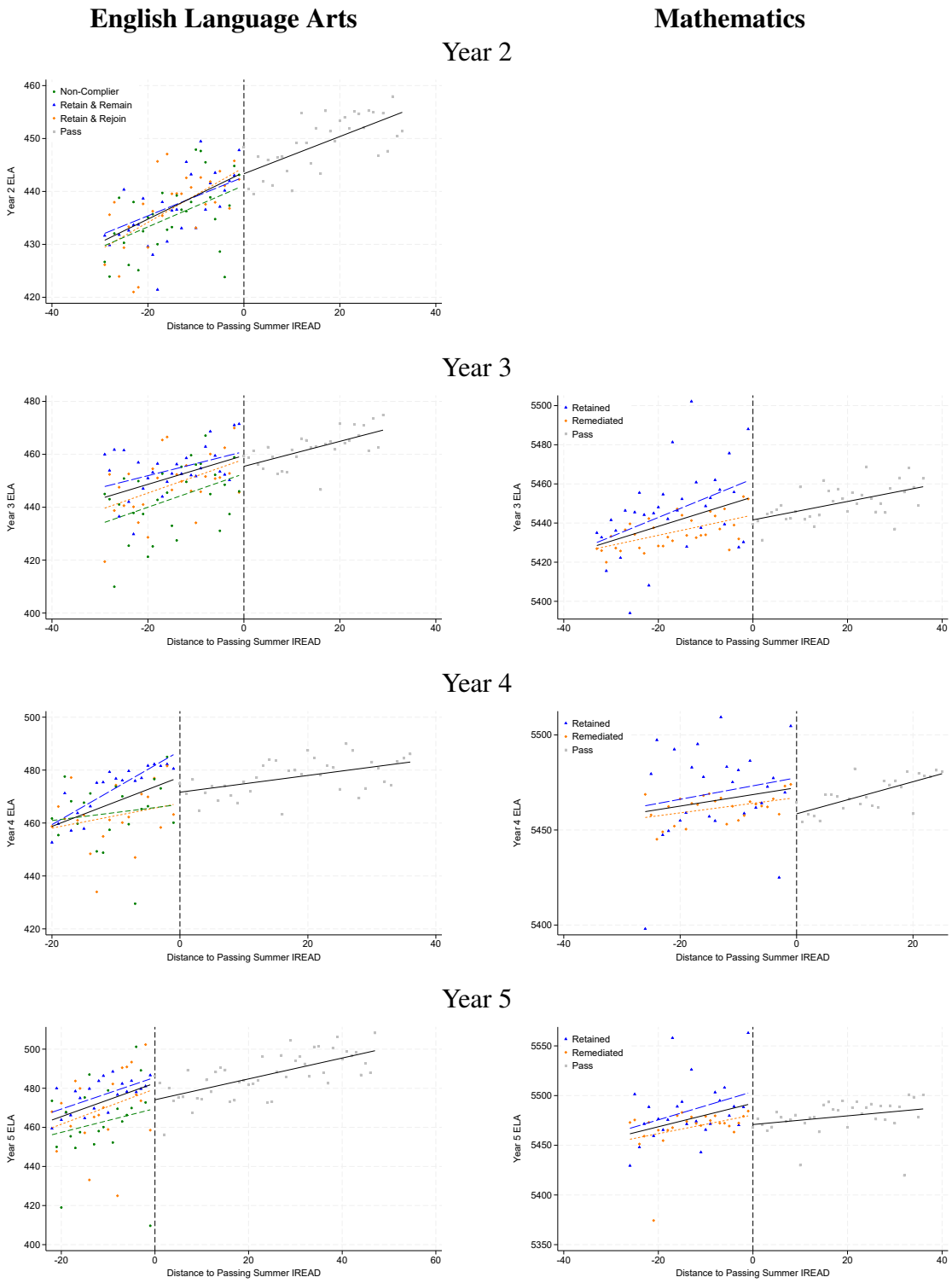
Notes: Figure illustrates kernel densities for summer administration of the IREAD. P-value for test for discontinuity at the passing score (446) is 0.991.

Figure A.4: Deconstructing the Effect of Failing Summer IREAD by Treatment Pathway: Same-Grade, ELA



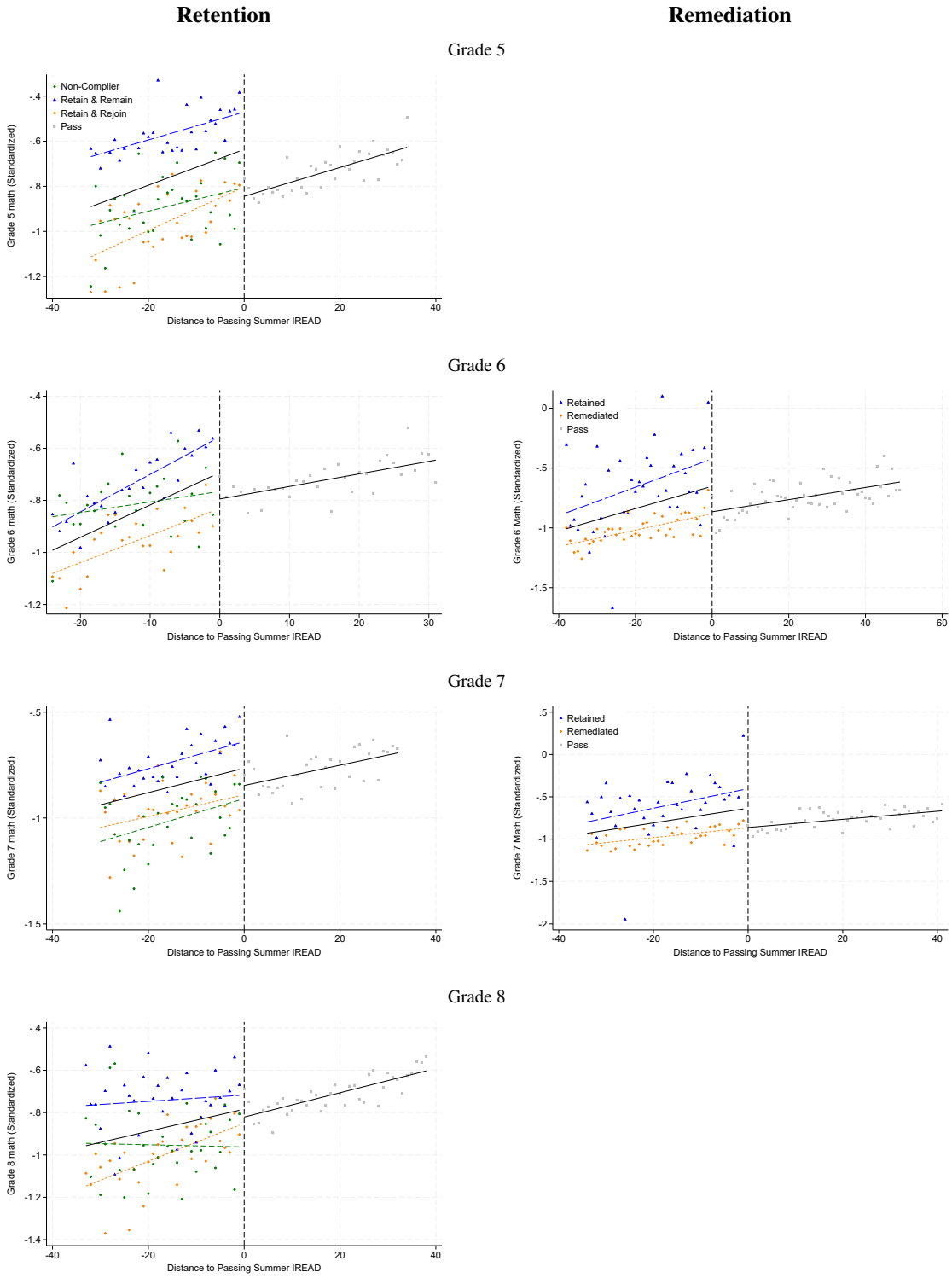
Notes: Figure illustrates results from estimating Equation (5) making within-grade comparisons. X-axis is the distance between student's score on the summer IREAD and the passing score. Y-axis is the student's score on the respective test, standardized by grade and year. Points on the figure illustrate average score by group for students with a particular summer IREAD score. Grey dots represent students with passing scores; Blue squares are students who failed, repeated third grade, and entered the respective grade a year behind their original cohort; orange triangles are students who failed and received the exemption for their respective cohort type; and green dots are students who failed but were typically promoted to the fourth grade. Lines represent local regressions fit for the respective observations. Black lines illustrate separately estimated reduced form regressions for the effect of failing summer IREAD that do not distinguish between student types.

Figure A.5: Deconstructing the Effect of Failing Summer IREAD by Treatment Pathway: Same-Year, ELA



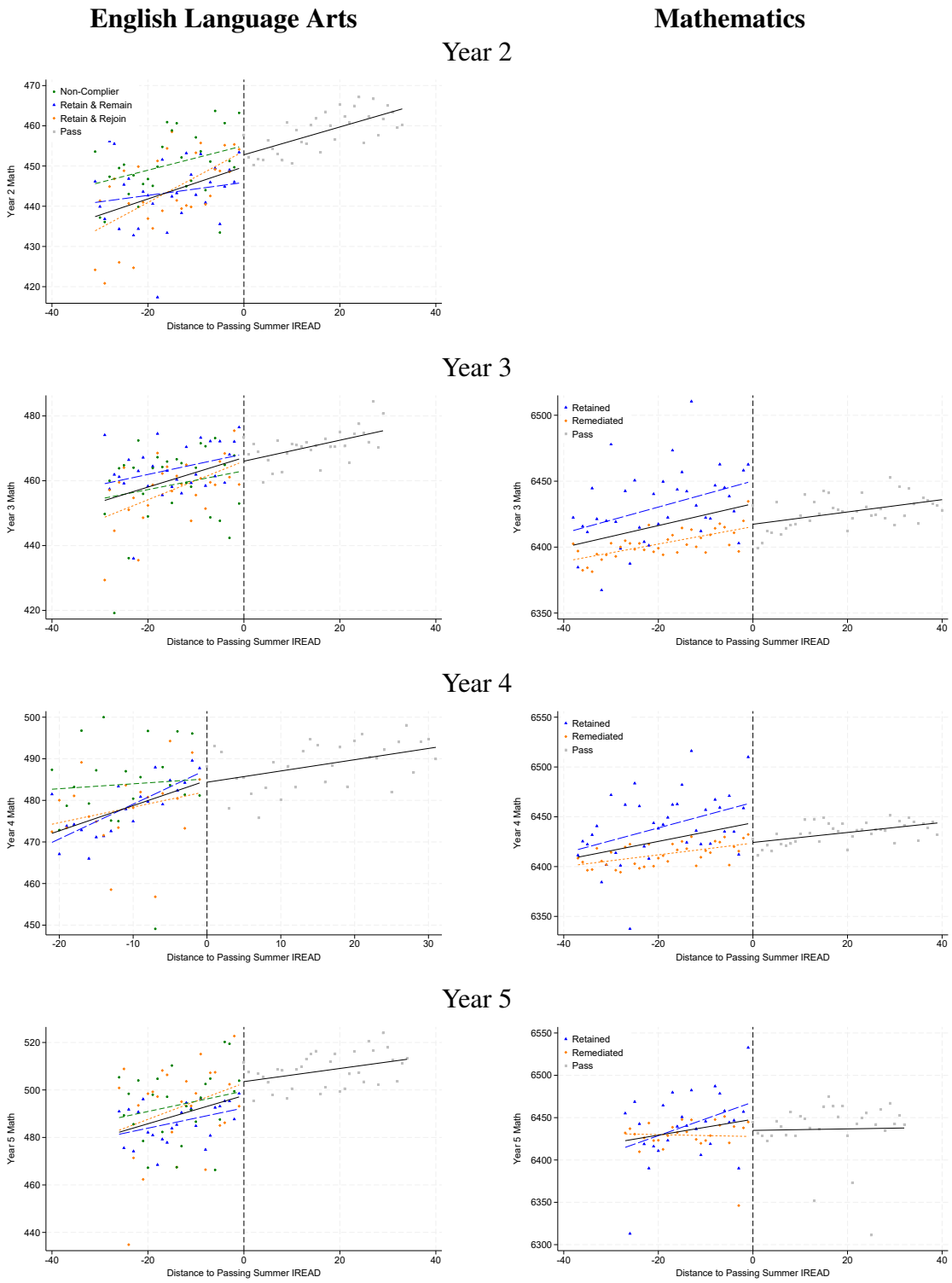
Notes: Figure illustrates results from estimating Equation (5) to make within-year comparisons. X-axis is the distance between student's score on the summer IREAD and the passing score, indicated by the vertical line. Y-axis is the student's score on the respective test, as measured by the vertical scale on the ISTEP+. Points on the figure illustrate average score by group for students with a particular summer IREAD score. Grey dots represent students with passing scores; Blue squares are students who failed, repeated third grade, and were a grade level behind their original cohort the respective year; orange triangles are students who failed and received the exemption for their respective cohort type; and green dots are students who failed but were typically promoted to the fourth grade.. Lines represent local regressions fit for the respective observations. Black lines illustrate separately estimated reduced form regressions for the effect of failing summer IREAD that do not distinguish between student types.

Figure A.6: Deconstructing the Effect of Failing Summer IREAD by Treatment Pathway: Same-Grade, Math



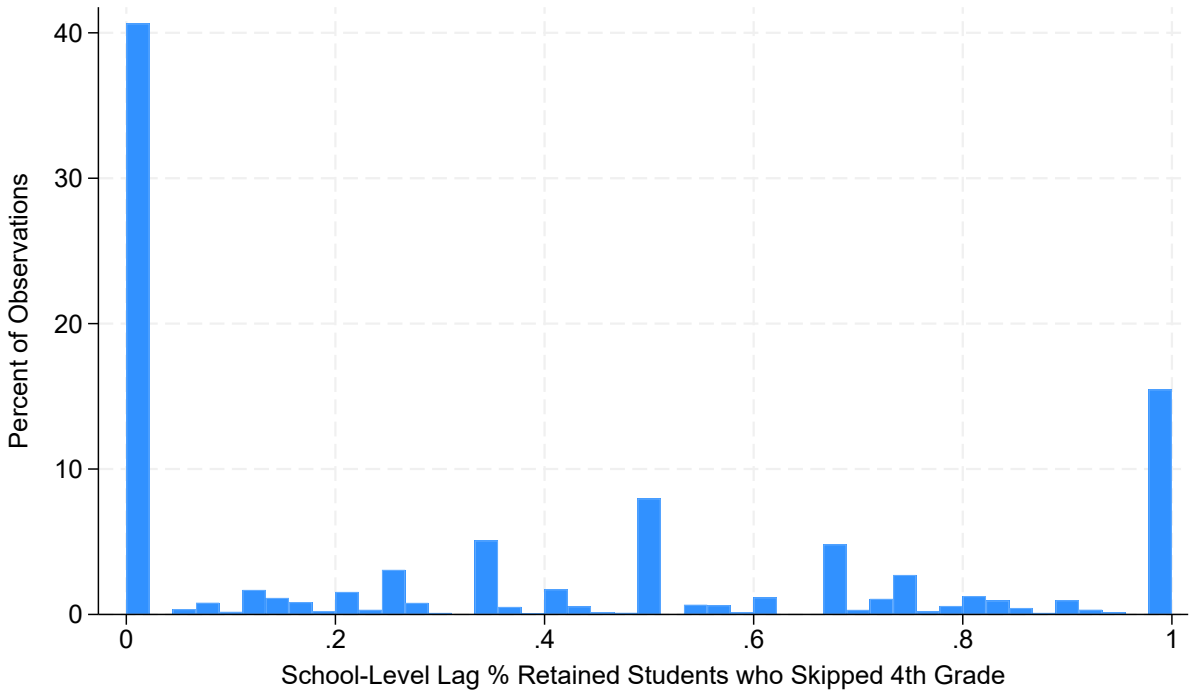
Notes: Figure illustrates results from estimating Equation (5) making within-grade comparisons. X-axis is the distance between student's score on the summer IREAD and the passing score. Y-axis is the student's score on the respective test, standardized by grade and year. Points on the figure illustrate average score by group for students with a particular summer IREAD score. Grey dots represent students with passing scores; Blue squares are students who failed, repeated third grade, and entered the respective grade a year behind their original cohort; orange triangles are students who failed and received the exemption for their respective cohort type; and green dots are students who failed but were typically promoted to the fourth grade. Lines represent local regressions fit for the respective observations. Black lines illustrate separately estimated reduced form regressions for the effect of failing summer IREAD that do not distinguish between student types.

Figure A.7: Deconstructing the Effect of Failing Summer IREAD by Treatment Pathway: Same-Year, Math



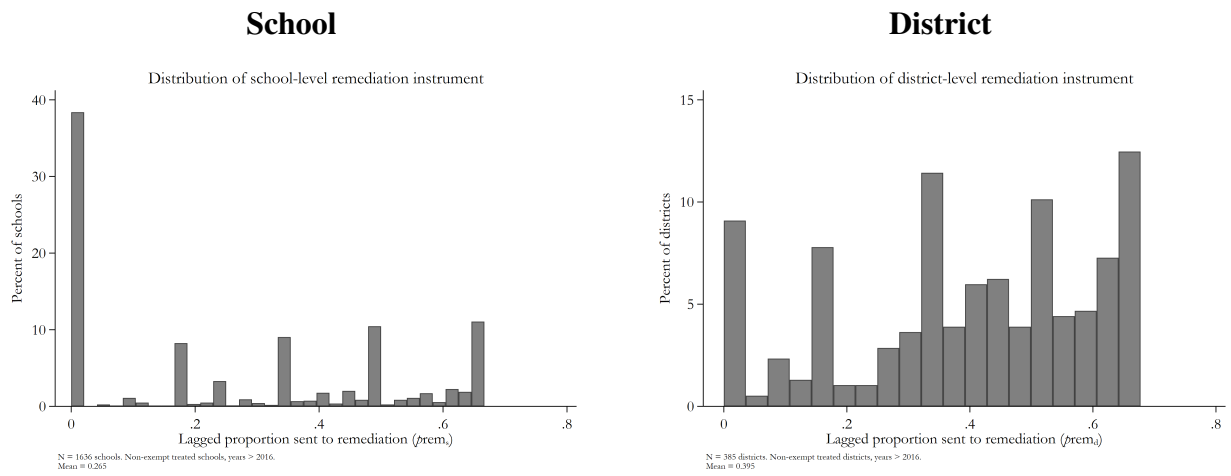
Notes: Figure illustrates results from estimating Equation (5) to make within-year comparisons. X-axis is the distance between student's score on the summer IREAD and the passing score, indicated by the vertical line. Y-axis is the student's score on the respective test, as measured by the vertical scale on the ISTEP+. Points on the figure illustrate average score by group for students with a particular summer IREAD score. Grey dots represent students with passing scores; Blue squares are students who failed, repeated third grade, and were a grade level behind their original cohort the respective year; and orange triangles are students who failed, repeated third grade, but were in the same grade as their original cohort in the respective year. Lines represent local regressions fit for the respective observations. Black lines illustrate separately estimated reduced form regressions for the effect of failing summer IREAD that do not distinguish between student types.

Figure A.8: Histogram for Lagged % Skipped



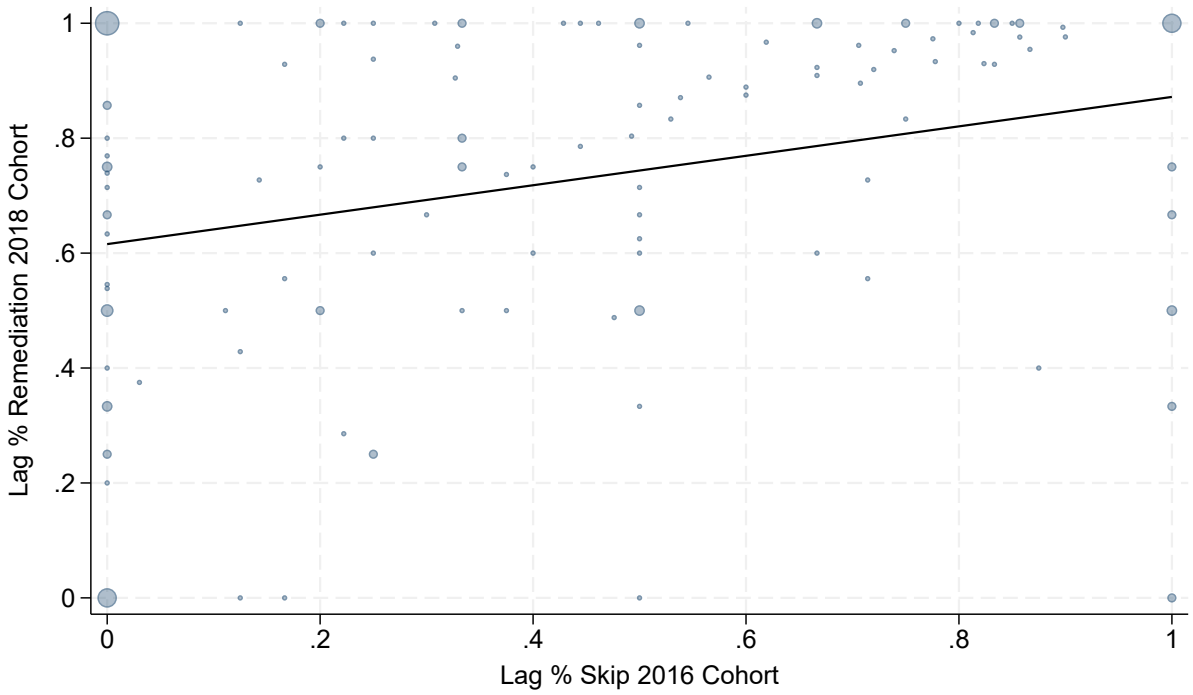
Notes: The figure shows the distribution, across schools, of the prior-year share of retained third-grade students who subsequently skipped the fourth grade. The distribution illustrates heterogeneity in post-retention pathway use across schools.

Figure A.9: Histogram for Lagged % Remediated



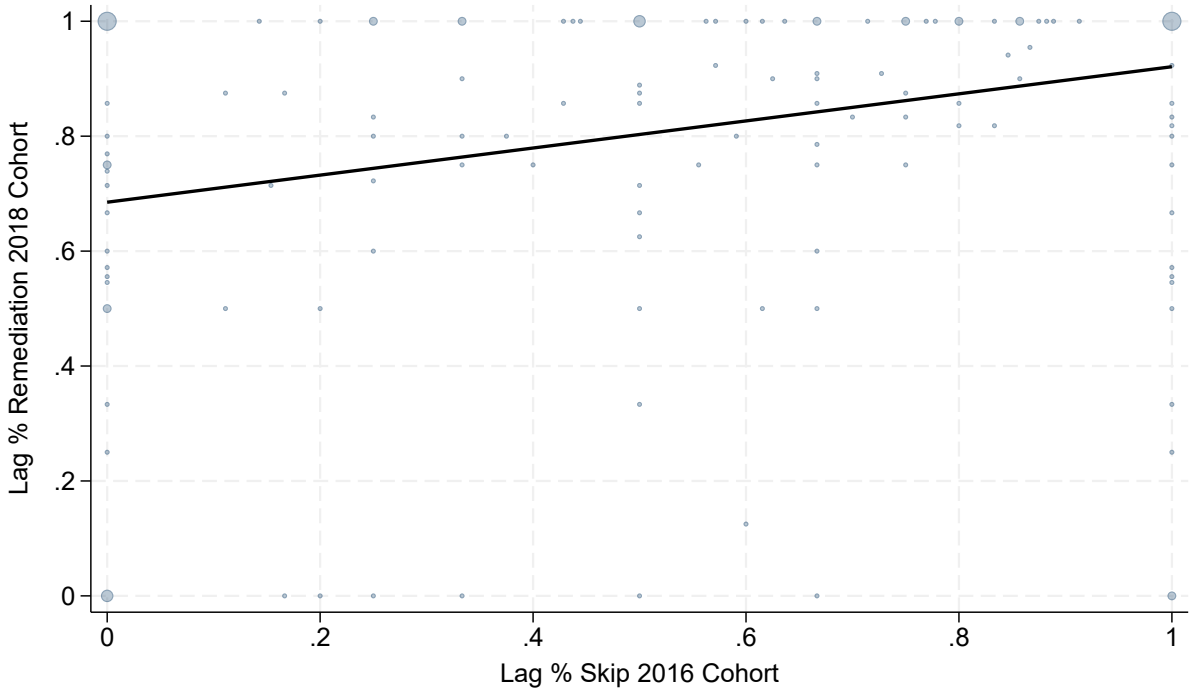
Notes: Each panel displays the distribution of remediation assignment rates across schools (left) and districts (right), where each bar represents the share of institutions with a given percentage of students assigned to remediation. The sample is restricted to schools and districts with at least one student who remained in remediation after 2016.

Figure A.10: Relationship Between District Lag % Skipped (2016) and Lag % Remediation (2018)



Notes: Figure illustrates relationship between district-level Z_{pskip} in the last observed year of its value (2016) and Z_{prem} in the first observed year of its value (2018). To be included, district must have retained at least 1 student in 2015 and have had at least 1 student fail the summer IREAD in 2017. Size of dot represents number of observations with the same value. Black line is a simple regression illustrating the relationship and has slope 0.2896.

Figure A.11: Relationship Between School Lag % Skipped (2016) and Lag % Remediation (2018)



Notes: Figure illustrates relationship between school-level Z_{pskip} in the last observed year of its value (2016) and Z_{prem} in the first observed year of its value (2018). To be included, schools must have retained at least 1 student in 2015 and have had at least 1 student fail the summer IREAD in 2017. Size of dot represents number of observations with the same value. Black line is a simple regression illustrating the relationship and has slope 0.2920.