EdWorkingPaper No. 22-688

Holding Back to Move Forward: The Effects of Retention in the Third Grade on Student Outcomes

NaYoung Hwang
University of New Hampshire

Cory Koedel
University of Missouri

We evaluate the effects of grade retention on students' academic, attendance, and disciplinary outcomes in Indiana. Using a regression discontinuity design, we show that third grade retention increases achievement in English Language Arts (ELA) and math immediately and substantially, and the effects persist into middle school. We find no evidence of grade retention effects on student attendance or disciplinary incidents, again into middle school. Our findings combine to show that Indiana's third grade retention policy improves achievement for retained students without adverse impacts along (measured) non-academic dimensions.

VERSION: December 2022

Suggested citation: Hwang, NaYoung, and Cory Koedel. (2022). Holding Back to Move Forward: The Effects of Retention in the Third Grade on Student Outcomes. (EdWorkingPaper: 22-688). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/mmxx-3e82

Holding Back to Move Forward:
The Effects of Retention in the Third Grade on Student Outcomes

NaYoung Hwang[1]
nayoung.hwang@unh.edu

Cory Koedel[2]
koedelc@missouri.edu

We evaluate the effects of grade retention on students' academic, attendance, and disciplinary outcomes in Indiana. Using a regression discontinuity design, we show that third grade retention increases achievement in English Language Arts (ELA) and math immediately and substantially, and the effects persist into middle school. We find no evidence of grade retention effects on student attendance or disciplinary incidents, again into middle school. Our findings combine to show that Indiana's third grade retention policy improves achievement for retained students without adverse impacts along (measured) non-academic dimensions.

Keywords: grade retention, student achievement, test-based retention policy, school discipline, absence

[1] NaYoung Hwang is an assistant professor in the Department of Education at the University of New Hampshire. Her research focuses on the ways educational policies and practices affect student learning and development with a particular emphasis on student-teacher relationships, teacher assignments, and school disciplinary policy.

[2] Cory Koedel is a professor in the Department of Economics and Truman School of Government and Public Affairs at the University of Missouri. His research is in the area of teacher quality and compensation, curriculum evaluation, and the efficacy of higher education institutions.

Acknowledgements: We appreciate Paul Hanselman, Thurston Domina, and Tutrang Nguyen, for their insightful suggestions on the earlier draft. This paper was supported by Notre Dame's Center for Research on Educational Opportunity (CREO) and the Institute of Educational Initiatives and their partnership with the Indiana Department of Education. We are grateful to the Indiana Department of Education for providing access to state administrative records and for supporting independent analyses. We are also grateful to Mark Berends and Roberto Peñaloza for supporting this project. All opinions expressed in this paper represent those of the authors and not necessarily the institutions with which they are affiliated. All errors in this paper are solely the responsibility of the authors.

Introduction

Grade retention is a potent but highly disruptive education intervention. Supporters of grade retention policies argue they provide students with the opportunity to master skills that are essential for learning in subsequent years, while opponents question the value of the extra time and emphasize the potential for negative social impacts. Despite mixed views among policymakers and educators, the implementation of grade retention policies is on the rise in the United States. As of 2019, 18 states had mandatory retention laws requiring students to repeat a grade if they do not attain grade-level proficiency on state assessments (Modan, 2019). Most policies focus on reading proficiency in early grades, with the rationale that early literacy skills are critical to students' long-term educational trajectories (Hernandez, 2011; Ferrer et al., 2015).

We contribute to the literature on grade retention by estimating the causal impacts of retention under Indiana's statewide, test-based retention policy. We estimate the impacts of retention on student achievement, attendance, and disciplinary outcomes. Like the slew of recent similar state policies, Indiana's policy is centered on students' third grade reading tests. Our analysis is based on state administrative data covering cohorts of third grade students from 2011-12 to 2016-17. For the earliest cohort, we track outcomes for five years after the retention event. We leverage Indiana's test-based retention rule in a regression-discontinuity (RD) framework for identification, which allows us to isolate the causal impacts of retention on marginally retained students.

We find that Indiana's retention policy has large positive short- and medium-term effects on same-grade student achievement in math and English Language Arts. The achievement effects are largest in the years more proximal to the retention event but remain substantial for the full five years over which we observe them. These findings corroborate several studies of similar early-grade retention policies in other U.S. states and school districts, which also show marked increases in student achievement after retention (in particular, see Green & Winters, 2007; Jacob & Lefgren, 2004; Schwerdt, West, & Winters, 2017). We find no evidence that retention affects students' disciplinary or attendance outcomes. These results contribute to a small and mixed literature on how similar policies affect school discipline and attendance (Martorell & Mariano, 2018; Özek, 2015). In summary, we find that Indiana's test-based retention policy has positive short- and medium-term effects on achievement for retained students and does not impact their disciplinary or attendance outcomes.

Background & Previous Literature

Theoretically, the impacts of grade retention on student outcomes are ambiguous. On the one hand, maturational theory suggests grade retention can benefit academically struggling students by giving them an extra year to master skills that are critical for educational progress (Crnic & Lamberty, 1994). Social comparison theory can also be used to support grade retention policies—if retained students compare themselves with their new classmates who are younger and less mature, repeating a grade likely improves academic self-efficacy, school engagement, and motivation (Wu et al., 2010). However, on the other hand, repeating a grade may demotivate students and create psychological stress, thereby hurting student development. Social comparison theory predicts students will be harmed by grade-retention policies if they compare themselves with their same-aged promoted peers, rather than their new same-grade (post-retention) peers (Kretschmann et al., 2019). Moreover, stigma theory posits that social disapproval stemming from grade retention can have long-lasting negative effects on students (Bos et al., 2013).

Given the stakes and controversy surrounding grade retention policies, it is unsurprising that they are widely studied. In their meta-analysis of the literature through 2007, Allen et al.

(2009) show the majority of studies estimate negative impacts of grade retention. However, they also identify research-design quality as a significant mediator of reported estimates—studies that use stronger research designs consistently find less negative or even positive impacts of grade retention. This is a sensible result given that grade retention is a treatment into which there is strong negative selection. Poorly-identified studies run the risk of confounding the causal impact of grade retention with negative selection into grade retention.

More recently, a small but growing literature uses more rigorous methods—and in particular, RD designs—to gain insight into the causal impacts of grade retention. RD is generally regarded as among the strongest non-experimental research designs available (e.g., see Chaplin et al., 2018). The RD approach is well-suited to study recently enacted grade retention policies because these policies use test-score cutoffs to assign students to the grade-retention treatment.

The newer, RD-based literature provides credible causal evidence on the impacts of grade retention, but the findings are mixed. Some studies find positive impacts (Diaz et al., 2021; Green & Winters, 2007; Schwerdt, West, & Winters, 2017), while others find mixed (Eren et al., 2017; Jacob & Lefgren, 2004, 2009), null (Martorell & Mariano, 2018), or negative impacts (Manacorda, 2012; Özek, 2015). While at first glance the mixed literature seems inconclusive, upon deeper review, some patterns emerge. Most notably, the impacts of grade retention on student outcomes are significantly more positive—although not exclusively positive—when retention occurs in an earlier grade. Most of the negative impacts of grade retention are found in instances where retention occurs in the sixth grade or higher. Indeed, Jacob and Lefgren (2009) and Diaz et al. (2021) show that a key channel through which early-grade retention positively impacts students in the long run is by reducing the likelihood of retention events in later grades.

A potential explanation for why grade retention effects are more negative in later grades is that the negative stigma and low levels of sense of belonging associated with retention may be more salient to older students (Anderson et al., 2005; Ou & Reynolds, 2010; Van Canegem et al., 2022).

Our Indiana evaluation contributes to the literature on early-grade retention policies, which show the most promise to date. Empirical evidence on the effects of third-grade retention on student achievement is available in two locales—Florida (Green & Winters, 2007; Schwerdt et al., 2017) and Chicago Public Schools (Jacob & Lefgren, 2004). Although the literature is small—certainly relative to the scope and significance of grade retention policies—the findings from these studies consistently show large and positive achievement effects. Our findings from Indiana, detailed below, further reinforce this result. While evidence from such a small number of studies should not be taken as conclusive, a consensus seems to be emerging that early-grade retention greatly improves student achievement, as intended.

By contrast, how early-grade retention impacts other student outcomes, such as student attendance and disciplinary events, is less clear. We are aware of just two studies that examine these outcomes and they reach different conclusions. Özek (2015) studies the Florida policy and finds retention greatly increases disciplinary incidents among retained students in the subsequent two years, although the effects dissipate thereafter. Conversely, Martorell and Mariano (2018) find no evidence that a similar retention policy in New York City affects students' discipline or attendance outcomes up to three years after the retention event.

Understanding these non-academic impacts is critical if policymakers are to make scientifically informed decisions. If the impacts are negative, as indicated by Özek's (2015) investigation in Florida, it implies a direct tradeoff to the achievement benefits of grade retention. Alternatively, if the impacts are null (or even positive) as indicated by Martorell and Mariano (2018), it suggests the achievement gains of retention are not offset by adverse consequences along these other dimensions.

Indiana Policy Details & Data

Indiana's retention policy requires students to be proficient in reading at the completion of the third grade before promotion to the fourth grade. Starting with the 2011-12 academic year, students who did not score at the level of proficient or above on the state-mandated Indiana Reading Evaluation and Determination test (IREAD-3) were required to repeat the grade unless they qualified for an exemption. Exempted students—English language learners, students with disabilities, and students who had previously been retained twice—could be promoted even if their IREAD-3 test scores did not reach the threshold. All third graders get two chances to pass the exam. The first exam is in March and the second is over the summer, in either June or July. Students who do not pass the first exam receive intensive remediation prior to the second exam.

We use administrative data from the Indiana Department of Education (IDOE) from academic years 2011-12 through 2016-17 for our analysis. We study four year-cohorts of students in the third grade, from 2011-12 through 2014-15. We follow all students through 2016-17, which is up to five years after the initial retention event (or potential event). We begin tracking outcomes in our preferred specifications, which compare marginally retained and promoted students when they are in the same grade, in the 4th grade. The last grade in which we compare same-grade outcomes is the 7th grade (for the initial 2011-12 cohort), five years after the retention event.

Our preferred estimates make same-grade comparisons between marginally retained and promoted students, but same-age comparisons are also informative and as such, we show them as well. The same-age comparisons occur in different grades. For our achievement evaluation we use scores on the Indiana State test—the Indiana Statewide Testing for Educational Progress-Plus, or ISTEP+—which purported to be a vertically scaled test in English Language Arts (ELA) and mathematics. In principle the vertical scaling should allow us to compare student achievement in different grades during the same year (i.e., at the same age), although with the caveat that vertical scaling is difficult to achieve in practice despite the claims of test publishers (Ballou, 2009). We provide a deeper discussion of the tradeoff in using same-grade versus same-age comparisons in the extensions section below.

Table 1 provides descriptive statistics for our data. The first vertical panel shows descriptive statistics for all third graders in Indiana, then further divides students based on retention/promotion status. Overall, approximately 1.8% of third graders were retained during our study period. Not surprisingly, retained students have lower average achievement, are disciplined at a higher rate, and are absent more frequently than their promoted peers in the year prior to the retention event. In addition, students enrolled for Free or Reduced-Price Lunch (FRL), Individualized Education Program (IEP) students, and Black and Hispanic students are overrepresented among retained students.

Next, the second vertical panel of Table 1 documents the analytic sample we use for our analysis. We make two notes about the analytic sample. First, we exclude English Language Learner (ELL) and IEP students because they qualify for exemptions to retention policy, making our research design ill-suited to evaluate them. Second, recall from above that when students fail the spring IREAD-3 test, they receive intensive instruction and re-take the test in the summer. Then, the promotion decision is based on the summer test. An RD design built around either the spring or summer test can be justified. Using the summer test will be more efficient because it is the decisive test. Put another way, an RD design built around the spring test will be "fuzzier"—and thus less efficient—because performance on the spring test is one step removed from the retention decision rule. However, a concern with using the summer test is that if the test re-take rate is low, and/or selection into retaking the test is endogenous, using the summer-test sample could cause bias in our evaluation. In Table 1, we show that the summer test re-take rate is very high—over 96 percent.[1] This largely negates the latter concern and accordingly, we use as our analytic sample the population of students who take the summer test each year.

In terms of outcomes, we estimate the achievement effects of the retention policy using ELA and math test scores from the ISTEP+. Our disciplinary outcome is the sum of unique suspension and expulsion events. We measure student absences as the number of school days missed by the student for an unapproved reason (i.e., unexcused absences). Summary statistics for all of these measures are provided in Table 1.

Methods

The RD design permits causal inference by comparing students whose likelihood of assignment to treatment changes discontinuously at a specific point in the distribution of an underlying continuous (or roughly continuous) variable, referred to as the "running variable (Lee & Lemieux, 2010)." In our application, small differences in students' third grade test scores around the policy cutoff generate samples of comparable students—along observed and unobserved dimensions under the RD identifying assumptions—who differ only by whether they are retained.

As noted above, some students in Indiana qualify for policy exemptions—namely, ELL and IEP students—and we drop these students from our evaluation. There is also non-compliance

[1] And even this number understates the coverage of the summer test, as the dominator includes students who exit Indiana Public Schools after the spring test.

with the policy for other reasons that cannot be addressed by sample adjustments (e.g., a parent refusing, or insisting, to have their child held back), with the end result being that the discontinuity in our application is fuzzy. A fuzzy RD is one in which there is a discontinuous jump in the likelihood of treatment at the cutoff, but assignment to treatment around the cutoff is not absolute. The fuzziness in the discontinuity in our application is not a threat to causal identification, although it results in less efficient estimation compared to a sharp discontinuity.

The standard approach to estimation in the presence of a fuzzy regression discontinuity is to use the discontinuity as an instrument for treatment. We follow this approach by estimating the following two-stage model. Equation (1) describes the first stage:

(1) $Retention_{i3} = \beta_0 + \beta_1 Below_{i3} + \beta_2 f(Runvar_{i3}) + \beta_3 Below_{i3} * g(Runvar_{i3}) + \beta_4 S_{i3} + e_{i3}$

In the equation, $Retention_{i3}$ is an indicator for whether student *i* is retained in the third grade, $Below_{i3}$ is an indicator for whether the student scored below the cutoff on the reading test, and $Runvar_{i3}$ is the IREAD-3 test score itself, or the running variable. The functions $f(Runvar_{i3})$ and $g(Runvar_{i3})$ allow for flexibility in the functional forms mapping the running variable to the retention event on both sides of the cutoff. The vector S_{i3} includes indicators for the following student characteristics: FRL enrollment, gender, and race/ethnicity. e_{i3} is the error term, which we cluster at the school level throughout.

We then estimate the following second-stage model:

(2) $Y_{ig} = \gamma_0 + \gamma_1 \widehat{Retention}_{i3} + \gamma_2 f(Runvar_{i3}) + \beta_3 Below_{i3} * g(Runvar_{i3}) + \beta_4 S_{ig} + \pi_{ig}$

where Y_{ig} indicates a student outcome of interest in grade *g*, where *g* takes values from 4-7. The outcomes are ELA and math ISTEP+ scores, disciplinary outcomes, and attendance. $\widehat{Retention}$ is the fitted value from equation (1) and isolates variation from differences in retention generated by the test-score cutoff. γ_1 is the parameter of interest and can be interpreted as the causal effect of retention under the RD identifying assumptions.

In our preferred models, we specify f(.) and g(.) in equations (1) and (2) as simple linear functions of student test scores and use a bandwidth of 25 test score points on each side of the cutoff. We also consider models that specify f(.) and g(.) and quadratic functions of test scores, and models that use alternative bandwidths ranging from 15-45 test score points. Importantly, none of our findings are substantively sensitive to how we specify f(.) and g(.), or to the choice of bandwidth (see below).

Results

RD Validation

We begin by providing evidence in support of the fundamental identifying assumption of the RD design—that conditional on the model described in the preceding section, students just below and above the cutoff are similar along observed and unobserved dimensions. It is not possible to prove this assumption with certainty, but its plausibility is typically probed in two ways. First is to test for evidence of running-variable manipulation, which could lead to observed or unobserved imbalance around the cutoff. Manipulation of test scores in our context could be caused by students or administrators. Student-driven manipulation would be difficult because retention among the summer test-taking sample is based on the single summer test, without the possibility of retakes. Administrator manipulation also seems unlikely outside of professional malpractice (such as what has been documented instances of teacher cheating—e.g., see Apperson et al., 2016; Jacob & Levitt, 2003).

Still, it is useful to provide empirical support. To this end, we conduct a density test to look for abnormalities in the distribution of the running variable around the test cutoff. The presence of an abnormality would raise concerns about manipulation through an unanticipated channel. Figure 1 shows the distribution of students' third-grade summer scores. Visually, the distribution is fairly smooth through the cutoff (at a score on the IREAD-3 test of 446). A formal density test following Cattaneo, Jansson, & Ma (2018) corroborates the visual evidence, revealing no indication of running-variable manipulation.[2]

The other common test for violations of the primary identifying assumption of RD is a test for observed differences in the sample just above and below the cutoff. There should be no such differences if the RD is dividing otherwise similar students into the treatment and control conditions. We test for observed differences coinciding with the discontinuity by estimating the model described by equations (1) and (2), except we remove the S-vector and instead use these variables as dependent variables. If the RD is operating as intended, the discontinuity should not systematically predict any of the exogenous variables in the S-vector. Table 2 shows that this is the case in our application.

Instrument Strength

Table 3 shows results from estimating the first-stage regression in equation (1) for each outcome-by-grade-level regression that we consider. Each cell in the table is from a different first-stage regression. The rows show the type of outcome, and the columns show the grade level at which the outcome is assessed.

The table reports the first-stage coefficient on the instrument with its standard error below in parenthesis, followed by first-stage F-statistic and the outcome-by-grade-level sample size. The latter is a function of (1) the number of cohorts used in estimation—recall that we use more cohorts in more proximal grades—and (2) the number of students with scores inside the 25-point

[2] Specifically, we test statistically for evidence of bunching around the cutoff and cannot reject the null hypothesis that the distribution is smooth (p-value: 0.91).

bandwidth. The primary takeaway from Table 3 is that the first stage is consistently strong throughout our analysis. In turn, this implies our instrumental variables (IV) estimates in the second stage will be well-powered with minimal bias due to instrumentation strength (Hahn & Hausman, 2005).

We also briefly draw attention to the fact that the first-stage coefficients are increasing with the grade level in Table 3. In results omitted for brevity, we show this is due to a cohort-composition effect combined with the fact that there was greater compliance with the test-based policy when it was first implemented in Indiana. That is, the coefficients are larger in the later grades because those estimates are based on data from the initial cohorts only, which had higher policy compliance rates. This does not affect causal inference from the RD estimates (because it does not affect the internal validity of the discontinuity for each cohort), although it does shift the weight in our estimates toward the earlier cohorts because they contribute more identifying variation to the RD parameter.[3]

Primary Findings

Table 4 shows our primary RD estimates of the achievement, disciplinary, and attendance impacts of grade retention. The structure of the table mirrors Table 3; each cell shows results from a different regression.

Focusing on the achievement outcomes in rows (1) and (2) first, we find large short-term effects of retention on same-grade achievement. In both ELA and math, retained students score over 0.50 student standard deviations higher on the 4th-grade ISTEP+ than their marginally promoted peers. These short-term effects are directionally aligned with the most-closely related evidence from Florida and Chicago (Green & Winters, 2007; Jacob & Lefgren, 2004; Schwerdt

[3] We confirm this explanation by re-estimating the first stage model for each grade-outcome combination using only data from the first cohort of students, in which case the first-stage coefficients become much more uniform.

et al., 2017). In terms of magnitude, they are a close match to the estimates from the two Florida studies, and larger than the estimates from Chicago.

As students progress through school, the achievement effects of retention attenuate. The pattern of attenuation is also a close match to the pattern in Schwerdt, West, and Winters (2017), which is the only similar study with a long enough data panel to facilitate a comparison. Still, despite the attenuation, by the 7th grade marginally retained students in Indiana significantly outperform marginally promoted students in ELA. In math, our 7th-grade point estimate is educationally meaningful but statistically imprecise.

Next, we turn to students' disciplinary and absence outcomes in rows (3) and (4) of Table 4. Here we find no evidence of retention impacts. For both outcomes our point estimates are inconsistently signed across grades, and none are statistically significant. Our findings in this regard are consistent with Martorell and Mariano (2018), who also find no evidence of disciplinary or attendance impacts of grade retention. They are at odds with Özek (2015), who finds that retention greatly increases disciplinary incidents. The addition of evidence from our single study does not resolve the ambiguity in the literature (and in fact, there may not be ambiguity if the prior conflicting findings are due to location-specific factors), but it is a step toward the development of a more robust literature on this question. Our findings in Indiana provide support for the claim that early-grade retention does not adversely affect these non-academic outcomes.

Figures 2 and 3 provide visual complements to the regression estimates in Table 4. The figures show reduced-form relationships between student scores near the cutoff and each outcome in each grade—i.e., they show what are referred to as intent-to-treat (ITT) effects in the parlance of the program evaluation literature. Our fuzzy RD estimates in Table 4 are connected to the graphs in Figures 2 and 3 through the first-stage coefficients in Table 3. Specifically, the estimates in Table 4 are treatment effects of retention, recovered by (effectively) dividing the reduced-form gaps in the figures by the first-stage regression coefficients from Table 3 to offset the attenuating effect (in the figures) of policy non-compliance.

Robustness

We examine the robustness of our findings to two features of estimation: the functional form we use to control for the running variable and the bandwidth we specify around the test-score cutoff. We discuss the results briefly here and provide full details in Appendix Tables 1 and 2.

First, recall from above that we specify f(.) and g(.) in equations (1) and (2) as simple linear functions of the running variable. This decision is visually supported by Figures 2 and 3 but to confirm robustness, we also estimate versions of our main models that specify f(.) and g(.) as quadratic functions of the running variable. Appendix Table 1 shows that this specification adjustment has no qualitative bearing on our findings, although our estimates from the models that use the quadratic specifications of f(.) and g(.) are less precise.

Next, we consider the sensitivity of our findings to alternative bandwidths around the test score cutoff. Recall that we use a bandwidth of 25 test points on each side of the cutoff in all of our models thus far. The 25-point bandwidth corresponds to just over half of a student standard deviation of the IREAD-3 test. To arrive at this bandwidth, we used the procedure of Calonico, Catteneo, and Farrell (2020) to calculate the optimal bandwidth for each of the 16 RD models in Table 4. The 25-point bandwidth is the minimum of the 16 optimal-bandwidth values, which range from 25-44 points across outcomes and grades. In Appendix Table 2, we re-estimate all of our models using bandwidths of 15, 35, and 45 points. This range includes an even more conservative bandwidth (15 points) and encompasses the full range of optimal bandwidth values based on Calonico, Catteneo, and Farrell (2020). Appendix Table 2 confirms that none of our findings are substantively sensitive to bandwidth choice over the wide range of bandwidths we consider.

Extensions

In this section we report on two extensions of our primary analysis. First, in Table 5, we test for heterogeneity in the effects of grade retention by student gender, race-ethnicity, and FRL status. We are not well-powered statistically to detect modest effect heterogeneity, but we can observe effect heterogeneity if it is substantial. While there are some fluctuations in the estimates throughout Table 5, our summary interpretation is that there is no evidence of substantial effect heterogeneity along the observable dimensions we can test.

In our second extension, we provide *same-age* estimates of grade retention to complement our preferred *same-grade* estimates presented above. To elaborate on the difference, in our same-grade comparisons thus far we compare marginally retained and promoted students when they are in the same grade, but in different years. The complementary same-age comparisons are between the same marginally retained and promoted students, but in the same year and different grades. The modeling structure is the same between these two types of comparisons, as shown in equations (1) and (2). The difference is in which outcomes are compared.

Schwedt, West, and Winters (2017) provide a thoughtful discussion of the merits of these two different types of estimates. Both are defensible, and both have limitations. Conceptually, we prefer the same-grade comparisons because they fit most logically with the intent of a retention policy within a standards-based education framework. That is, in a standards-based framework, promotion should be based on meeting the standard. In principle, age should not be a factor. As a counterpoint, however, this view ignores the obvious practical point that age *is* a factor, at least to some degree.[4] The age-based comparison also makes sense within a larger policy framework outside of the bounds of the education system. Such a framework would include the opportunity cost of retention in the form of extending the K-12 education period; the age-based comparison incorporates this cost, whereas the grade-based comparison does not.

There are also identification tradeoffs associated with switching between same-grade and same-age comparisons. In same-grade comparisons, treated students are older and have attended school longer at the time of the comparison. Schwedt, West, and Winters (2017) note these factors can be viewed as confounders if the goal is to narrowly identify the effect of the retention event. However, in a standards-based framework, we believe it is more appropriate to view these "confounders" as part of the bundle of treatments associated with retention. Said another way, the purpose of retention is partly to increase maturity, and years of schooling, conditional on the grade. The identification challenge for same-age comparisons is with respect to the achievement outcomes, which require stronger measurement assumptions to facilitate comparisons across grades. In particular, and as noted above, it must be assumed that test scores are vertically scaled, which means that a test "point" in any grade represents the same amount of knowledge, regardless of which grade it is from. Many test publishers claim their tests are vertically scaled, but research (and even simple empirical properties of the exams) suggests this is often not the case (Ballou, 2009; Schwerdt et al., 2017).[5]

Ultimately, given that both same-grade and same-age comparisons are reasonable, and also limited, we do not take a firm stand on the "right" way to estimate the effects of grade

[4] In fact, this is made explicit in the Indiana policy in extreme cases—e.g., students who are retained for two consecutive years are exempt from the test-based rule for promotion in the following year.
[5] The same-grade comparisons circumvent the scaling problem by standardizing test scores within grades and comparing the positions of marginally retained and promoted students in the same-grade achievement distribution.

retention. We lead with the same-grade comparisons because we prefer them conceptually and technically. But for readers who prefer same-age comparisons, we provide these in Table 6.

We make two presentational notes about Table 6 before discussing the substantive findings. First, Table 6 has one more column than our preceding tables. This is because we can begin to make same-age comparisons between marginally retained and promoted students one year earlier—specifically, in the year after the summer test, when retained students are in the third grade and promoted students are in the fourth grade (whereas for same-grade comparisons, the first post-retention common grade is grade 4). Second, for our achievement models, we report our estimates in raw test-score points rather than standard deviation units. Under the assumption that the ISTEP+ is vertically scaled, we can compare test-point gaps between retained and promoted students in different grades similarly to how we compare standardized test scores within grades in our preceding models.

Turning to the substance of Table 6, the most straightforward takeaway is that our null findings for disciplinary and absence outcomes in rows (3) and (4) are substantively unaffected by switching to the same-age comparisons. The achievement estimates indicate positive and statistically significant effects on ELA achievement up to five years after the retention event, and null effects on math achievement throughout. The magnitudes of the ELA estimates in Table 6 are somewhat smaller and statistically less certain than the analogous same-grade estimates in Table 4. This is unsurprising because retained students are now being compared to promoted students without the benefit of the additional year of maturity or extra year of schooling bundled into the same-grade retention effects. To get a rough sense of the magnitude difference, note that the average student standard deviation of ELA scores in grades 3-8 in our sample is about 53 points. So, for instance, the 9.74-point effect on the ELA score 3 years after retention in Table 6 corresponds to an effect size in standard deviation units of about 0.18.

Discussion and Conclusion

We estimate the short- and medium-run impacts of retention in the third grade on students' academic and non-academic outcomes. Using a regression discontinuity design built around Indiana's test-based grade retention policy, we show that retention increases student achievement substantially, echoing findings from earlier studies (Green & Winters, 2007; Jacob & Lefgren, 2004; Schwerdt et al., 2017). The positive achievement effects of retention persist up to five years after the retention event and are stronger—at least suggestively—in ELA than in mathematics.

We do not find any evidence that grade retention impacts students' attendance or disciplinary outcomes. These null findings are important in light of evidence that disciplinary incidents in particular (such as suspensions) are associated with a host of negative student outcomes (Hwang, 2018; Monahan et al., 2014; Raffaele Mendez, 2003). Our null results for these outcomes contribute to a small and mixed literature. They corroborate findings from Martorell and Mariano (2018), who show that early-grade retention does not affect these outcomes for up to three years post-retention in New York City. In Indiana, we find no evidence of retention effects up to five years later. The five-year timespan afforded by our data panel includes two years of middle-school enrollment for retained students, which is notable given the well-documented increase in disciplinary incidents as students move from elementary to middle schools (Eccles et al., 1997; Figlio, 2007).

Taken on the whole, our findings of positive achievement effects of the Indiana policy, coupled with the lack of negative effects on attendance and disciplinary outcomes, suggest grade retention is a promising intervention for students who are struggling academically early in their schooling careers. In the context of the larger literature, our results fit with an emerging theme that timing greatly affects the impacts of grade retention. Retention policies in later grades—e.g., in middle school—mostly have negative impacts on a variety of student outcomes (Manacorda, 2012; Jacob & Lefgren, 2009), whereas retention policies in earlier grades tend to be beneficial (albeit with some exceptions; e.g., see Özek, 2015).

We conclude with two general caveats to our findings. First, while we contribute to a thin literature on the effects of retention on students' non-academic outcomes, we are constrained by what we can measure. Future studies that can expand the range of non-academic outcomes to include concepts such as a sense of belonging, self-efficacy, self-esteem, and confidence would be valuable additions to the literature. Second, we note the standard RD qualification that our findings are identified locally by comparing marginally retained and promoted students. Our estimates may not apply to students with scores far from the cutoff. That said, in our context, marginal students are of the greatest policy interest. For example, it is not useful to understand the effect of retention for students who score well above the cutoff and are not under consideration for retention; and while in principle understanding the effect of retention on students well below the cutoff may be of interest, such a parameter would apply to very few students because few students' have scores this low.

References

Allen, C. S., Chen, Q., Willson, V. L., & Hughes, J. N. (2009). Quality of research design moderates effects of grade retention on achievement: A meta-analytic, multilevel analysis. *Educational Evaluation and Policy Analysis*, *31*(4), 480-499.

Anderson, G. E., Jimerson, S. R., & Whipple, A. D. (2005). Student ratings of stressful experiences at home and school: Loss of a parent and grade retention as superlative stressors. *Journal of Applied School Psychology*, *21*(1), 1-20.

Apperson, J., Bueno, C., and Sass, T. (2016). Do the cheated ever prosper? The long-run effects of test-score manipulation by teachers on student outcomes. CALDER Working Paper No. 155. Washington, DC: CALDER.

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

Bos, A. E., Pryor, J. B., Reeder, G. D., & Stutterheim, S. E. (2013). Stigma: Advances in theory and research. *Basic and applied social psychology*, *35*(1), 1-9.

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, *23*(2), 192-210.

Cattaneo, M. D., Jansson, M., & Ma, X. (2018). Manipulation testing based on density discontinuity. *The Stata Journal*, *18*(1), 234-261.

Chaplin, D.D., Cook, T.D., Zurovac, J., Coopersmith, J.S., Finucane, M.M., Vollmer, L.N., and Morris, R.E. (2018). The internal and external validity of the regression discontinuity design: A meta-analysis of 15 within-study comparisons. *Journal of Policy Analysis and Management* 37(2), 403-429.

Crnic, K., & Lamberty, G. (1994). Reconsidering school readiness: Conceptual and applied perspectives. *Early Education and Development*, *5*(2), 91-105.

Diaz, J., Grau, N., Reyes, T., and Rivera, J. (2021). The impact of grade retention on juvenile crime. *Economics of Education Review* 84, 1-16.

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Mac Iver, D. (1997). Development during adolescence: The impact of stage–environment fit on young adolescents' experiences in schools and in families (1993). In J. M. Notterman (Ed.), *The evolution of psychology: Fifty years of the American Psychologist* (pp. 475–501). American Psychological Association.

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

Ferrer, E., Shaywitz, B. A., Holahan, J. M., Marchione, K. E., Michaels, R., & Shaywitz, S. E. (2015). Achievement gap in reading is present as early as first grade and persists through adolescence. *The Journal of pediatrics*, *167*(5), 1121-1125.

Figlio, D. N. (2007). Boys named Sue: Disruptive children and their peers. *Education finance and policy*, *2*(4), 376-394.

Greene, J. P., & Winters, M. A. (2007). Revisiting grade retention: An evaluation of Florida's test-based promotion policy. *Education Finance and Policy*, *2*(4), 319-340.

Hahn, J., and Hausman, J. (2005). Estimation with valid and invalid instruments. *Annales D'Economie et de Statistique* 79-80, 25-57.

Hernandez, D. J. (2011). Double jeopardy: How third-grade reading skills and poverty influence high school graduation. Annie E. Casey Foundation.

Hwang, N. (2018). Suspensions and achievement: Varying links by type, frequency, and subgroup. *Educational Researcher*, *47*(6), 363-374.

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

Jacob, B. A., & Lefgren, L. (2009). The effect of grade retention on high school completion. *American Economic Journal: Applied Economics 1*(3), 33-58.

Jacob, B. A., & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *The Quarterly Journal of Economics*, *118*(3), 843-877.

Kretschmann, J., Vock, M., Lüdtke, O., Jansen, M., & Gronostaj, A. (2019). Effects of grade retention on students' motivation: A longitudinal study over 3 years of secondary school. *Journal of educational psychology*, *111*(8), 1432-1446.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of economic literature*, *48*(2), 281-355.

Manacorda, M. (2012). The cost of grade retention. *Review of Economics and Statistics*, *94*(2), 596-606.

Martorell, P., & Mariano, L. T. (2018). The causal effects of grade retention on behavioral outcomes. *Journal of Research on Educational Effectiveness*, *11*(2), 192-216.

Modan, N. (2019). 50 States of Ed Policy: Do third grade retention policies work? https://www.educationdive.com/news/the-50-states-of-education-policy-do-3rd-grade-retention-policies-work/559741/

Monahan, K. C., VanDerhei, S., Bechtold, J., & Cauffman, E. (2014). From the school yard to the squad car: School discipline, truancy, and arrest. *Journal of youth and adolescence*, *43*(7), 1110-1122.

Ou, S. R., & Reynolds, A. J. (2010). Grade retention, postsecondary education, and public aid receipt. *Educational Evaluation and Policy Analysis*, *32*(1), 118-139.

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

Raffaele Mendez, L. M. (2003). Predictors of suspension and negative school outcomes: A longitudinal investigation. *New directions for youth development*, *2003*(99), 17-33.

Schwerdt, G., West, M. R., & 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.

Van Canegem, T., Van Houtte, M., & Demanet, J. (2022). Grade Retention: A Pathway to Solitude? A Cross-National Multilevel Analysis of the Effects of Being Retained on Students' Sense of Belonging. *Comparative Education Review*, *66*(4), 000-000.

Wu, W., West, S. G., & Hughes, J. N. (2010). Effect of grade retention in first grade on psychosocial outcomes. *Journal of educational psychology*, *102*(1), 135-152.

Table 1
Descriptive Statistics

	Full sample						Analytic Sample			
	All (*N*=390,853)		Retained (*N*=6,978)		Promoted (*N*=383,875)		Students who took Summer IREAD-3 (*N*=21,361)		Students who Failed Spring IREAD-3 (*N*=22,166)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
ELA achievement (Standardized)	0.010	0.994	-0.766	0.672	0.024	0.993	-1.011	0.668	-1.070	0.639
Math achievement (Standardized)	0.015	0.992	-0.577	0.767	0.026	0.992	-0.900	0.798	-0.945	0.784
Disciplinary action	0.032		0.110		0.031		0.086		0.090	
Unexcused absence	1.764	3.328	3.458	4.862	1.734	3.286	2.979	4.478	3.084	4.572
Female	0.492		0.456		0.493		0.472		0.469	
FRL	0.498		0.824		0.492		0.733		0.750	
ELL	0.080		0.074		0.080		0.000		0.000	
IEP	0.133		0.212		0.132		0.000		0.000	
Black	0.110		0.354		0.106		0.338		0.351	
Hispanic	0.113		0.125		0.113		0.067		0.066	
Other race/ethnicity	0.069		0.065		0.069		0.073		0.074	
White	0.706		0.454		0.710		0.520		0.506	

Note. These student characteristics are based on third grades. ELA = English Language Arts. FRL = eligibility for free or reduced-price lunch. ELL = English Language Learner. IEP = Individualized Education Program. Disciplinary action includes in-school suspension, out-of-school suspension, and expulsion.

Table 2
Regression Discontinuity Validation Tests

	(1)	(2)	(3)	(4)	(6)	(7)
	Female	FRL	Black	Hispanic	Other race/ethnicity	White
Failed IREAD-3 promotion cutoff	-0.005	-0.007	0.006	-0.001	0.009	-0.013
	(0.018)	(0.017)	(0.027)	(0.010)	(0.010)	(0.027)
N	9188	9188	9188	9188	9188	9188

Note. FRL = eligibility for free or reduced-price lunch. The sample is restricted to students who scored within 25 on Summer IREAD-3. Robust standard errors are reported in parentheses.

Table 3
First Stage Estimates of the Effects of Failing to Meet Summer IREAD-3 Cutoff on Third Grade Retention

	(1) Grade 4	(2) Grade 5	(3) Grade 6	(4) Grade 7
Panel A: ELA				
Retention (Grade 3)	0.352***	0.465***	0.526***	0.532***
	(0.022)	(0.024)	(0.028)	(0.036)
F-statistic	265.20	383.42	365.23	223.83
N	8376	7016	5079	3144
Panel B: Math				
Retention (Grade 3)	0.350***	0.470***	0.531***	0.537***
	(0.022)	(0.024)	(0.027)	(0.035)
F-statistic	262.76	390.85	377.91	237.47
N	8402	7034	5088	3155
Panel C: Discipline				
Retention (Grade 3)	0.352***	0.462***	0.526***	0.535***
	(0.021)	(0.024)	(0.027)	(0.035)
F-statistic	268.30	385.73	379.08	239.32
N	8542	7164	5200	3241
Panel D: Absence				
Retention (Grade 3)	0.355***	0.462***	0.530***	0.540***
	(0.022)	(0.023)	(0.027)	(0.035)
F-statistic	269.94	387.70	385.34	242.42
N	8519	7152	5192	3239

Note. This table reports the first stage estimates from analysis in Table 4. Standard errors in parentheses are clustered at the school level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4
The Effects of Third Grade Retention on Student Outcomes

	(1) Grade 4	(2) Grade 5	(3) Grade 6	(4) Grade 7
Panel A: ELA				
Retention (Grade 3)	0.528***	0.274***	0.315***	0.245**
	(0.073)	(0.064)	(0.071)	(0.088)
N	8376	7016	5079	3144
Panel B: Math				
Retention (Grade 3)	0.522***	0.378***	0.188**	0.154
	(0.082)	(0.077)	(0.071)	(0.094)
N	8402	7034	5088	3155
Panel C: Discipline				
Retention (Grade 3)	-0.070	-0.004	0.030	-0.069
	(0.038)	(0.041)	(0.048)	(0.066)
N	8542	7164	5200	3241
Panel D: Absence				
Retention (Grade 3)	-0.271	0.210	-0.401	0.692
	(0.572)	(0.507)	(0.577)	(0.843)
N	8519	7152	5192	3239

Note. Our models include student level controls, such achievement/school discipline/absence, sex, race/ethnicity, and FRL in the third grade. The sample is restricted to students who scored within 25 points of the policy cutoff for retention on the summer IREAD-3 test. Standard errors in parentheses are clustered at the school level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5
The Effects of Third Grade Retention on Student Outcomes across Subgroups

Panel A: Student Achievement	ELA				Math			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Grade 4	Grade 5	Grade 6	Grade 7	Grade 4	Grade 5	Grade 6	Grade 7
Female	0.461***	0.277**	0.349***	0.277*	0.441***	0.279**	0.085	-0.014
	(0.092)	(0.088)	(0.095)	(0.132)	(0.105)	(0.104)	(0.101)	(0.136)
Male	0.592***	0.271**	0.270**	0.209	0.596***	0.480***	0.299**	0.314*
	(0.110)	(0.093)	(0.103)	(0.119)	(0.122)	(0.108)	(0.094)	(0.130)
FRL	0.528***	0.261***	0.236**	0.214*	0.481***	0.352***	0.154*	0.156
	(0.076)	(0.071)	(0.078)	(0.094)	(0.086)	(0.085)	(0.077)	(0.105)
Non-FRL	0.514*	0.298*	0.623***	0.330	0.729***	0.473**	0.318	0.166
	(0.212)	(0.147)	(0.174)	(0.203)	(0.214)	(0.164)	(0.169)	(0.220)
Black	0.470***	0.158	0.282*	0.143	0.383*	0.286*	0.228	0.092
	(0.127)	(0.107)	(0.112)	(0.153)	(0.151)	(0.140)	(0.119)	(0.162)
Hispanic	0.855**	0.628*	0.488	0.482	0.740*	0.548	-0.235	0.638
	(0.275)	(0.287)	(0.328)	(0.475)	(0.352)	(0.301)	(0.310)	(0.446)
White	0.499***	0.284**	0.323**	0.275*	0.465***	0.350***	0.183	0.136
	(0.097)	(0.092)	(0.104)	(0.116)	(0.107)	(0.097)	(0.103)	(0.136)
Other race/ethnicity	0.750**	0.538**	0.459	0.328	1.140***	0.814**	0.277	0.108
	(0.234)	(0.204)	(0.245)	(0.243)	(0.288)	(0.262)	(0.205)	(0.290)

Panel: Behavioral Outcome	Discipline				Absence			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Grade 4	Grade 5	Grade 6	Grade 7	Grade 4	Grade 5	Grade 6	Grade 7
Female	-0.087*	-0.038	0.035	-0.016	0.119	-0.254	-0.344	0.413
	(0.040)	(0.046)	(0.060)	(0.088)	(0.719)	(0.712)	(0.860)	(1.284)
Male	-0.049	0.031	0.026	-0.121	-0.647	0.671	-0.451	1.130
	(0.068)	(0.064)	(0.072)	(0.093)	(0.888)	(0.699)	(0.747)	(1.126)
FRL	-0.066	0.005	0.062	-0.032	-0.332	0.473	-0.323	0.973
	(0.044)	(0.047)	(0.056)	(0.077)	(0.664)	(0.602)	(0.688)	(1.028)
Non-FRL	-0.099	-0.028	-0.097	-0.183	0.103	-0.924	-0.749	-0.208
	(0.066)	(0.074)	(0.083)	(0.121)	(1.148)	(0.694)	(0.813)	(1.104)
Black	-0.021	0.059	-0.030	-0.083	-1.139	-0.474	-2.556*	-1.096
	(0.094)	(0.078)	(0.084)	(0.128)	(1.258)	(0.833)	(1.151)	(1.673)
Hispanic	0.028	0.061	0.290	-0.232	3.021	1.642	0.458	4.064
	(0.168)	(0.185)	(0.173)	(0.276)	(2.834)	(2.830)	(1.817)	(4.830)
White	-0.098*	-0.022	0.041	-0.048	0.196	0.545	1.142	2.090
	(0.039)	(0.044)	(0.061)	(0.082)	(0.629)	(0.659)	(0.719)	(1.075)
Other race/ethnicity	-0.146	-0.274*	0.007	-0.076	-1.142	0.689	0.321	-1.452
	(0.098)	(0.132)	(0.171)	(0.178)	(1.865)	(1.925)	(1.831)	(2.079)

Note. The sample is restricted to students who scored within 25 on Summer IREAD-3. Standard errors in parentheses are clustered at the school level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6
The Effects of Third Grade Retention on Student Achievement and Behavior (Age-based Comparisons)

	(1)	(2)	(3)	(4)	(5)
	1 Year	2 Years	3 Years	4 Years	5 Years
Panel A: ELA					
Retention (Grade 3)	9.972***	0.297	9.737**	10.998**	16.630*
	(2.742)	(2.935)	(3.516)	(3.890)	(7.029)
N	10145	6602	4661	3126	1646
Panel B: Math					
Retention (Grade 3)	3.903	0.825	2.363	-1.624	6.126
	(3.511)	(3.828)	(3.475)	(3.515)	(5.648)
N	10202	6660	4677	3137	1646
Panel C: Discipline					
Retention (Grade 3)	-0.038	-0.055	-0.021	-0.053	-0.034
	(0.026)	(0.035)	(0.043)	(0.048)	(0.071)
N	10448	6821	4814	3253	1724
Panel D: Absence					
Retention (Grade 3)	0.004	0.380	-0.124	0.151	0.631
	(0.416)	(0.462)	(0.537)	(0.642)	(1.079)
N	10411	6804	4804	3251	1721

Note. Our models include student level controls, such achievement/school discipline/absence, sex, race/ethnicity, and FRL in the third grade. The sample is restricted to students who scored within 25 on Summer IREAD-3. Standard errors in parentheses are clustered at the school level.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 1
Distribution of summer IREAD-3 scores, pooled across cohorts, centered on the retention cutoff (446)

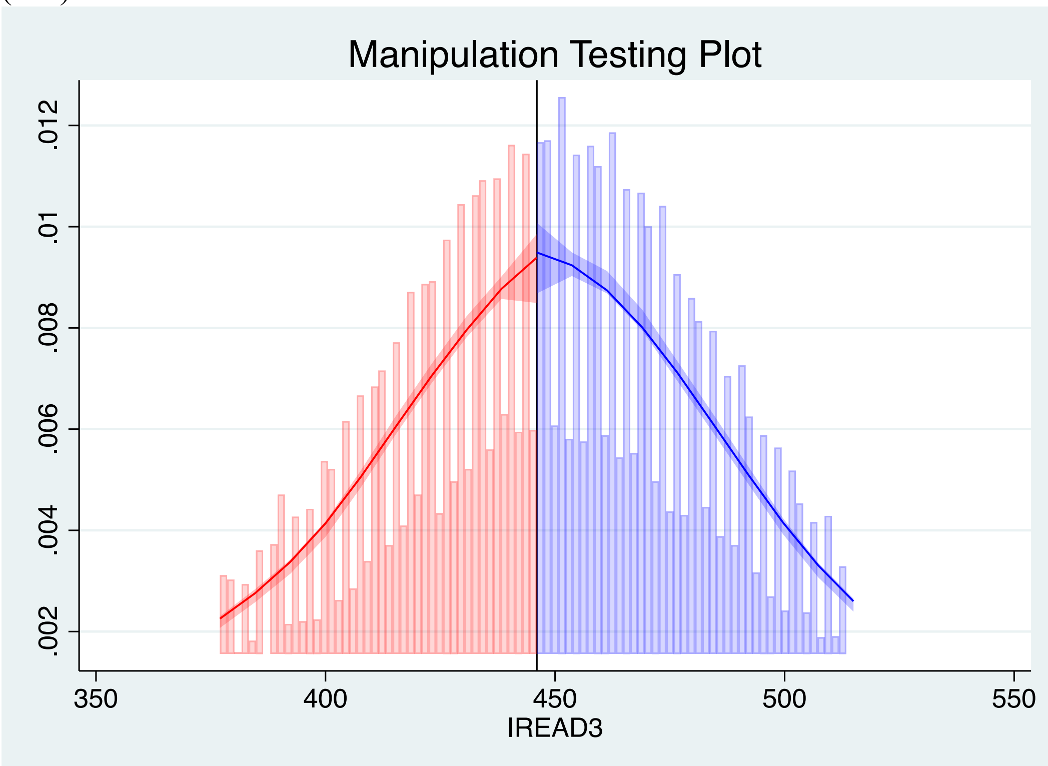


Figure 2
The Effects of Third Grade Retention on ELA and math achievement in Fourth through Seventh Grades

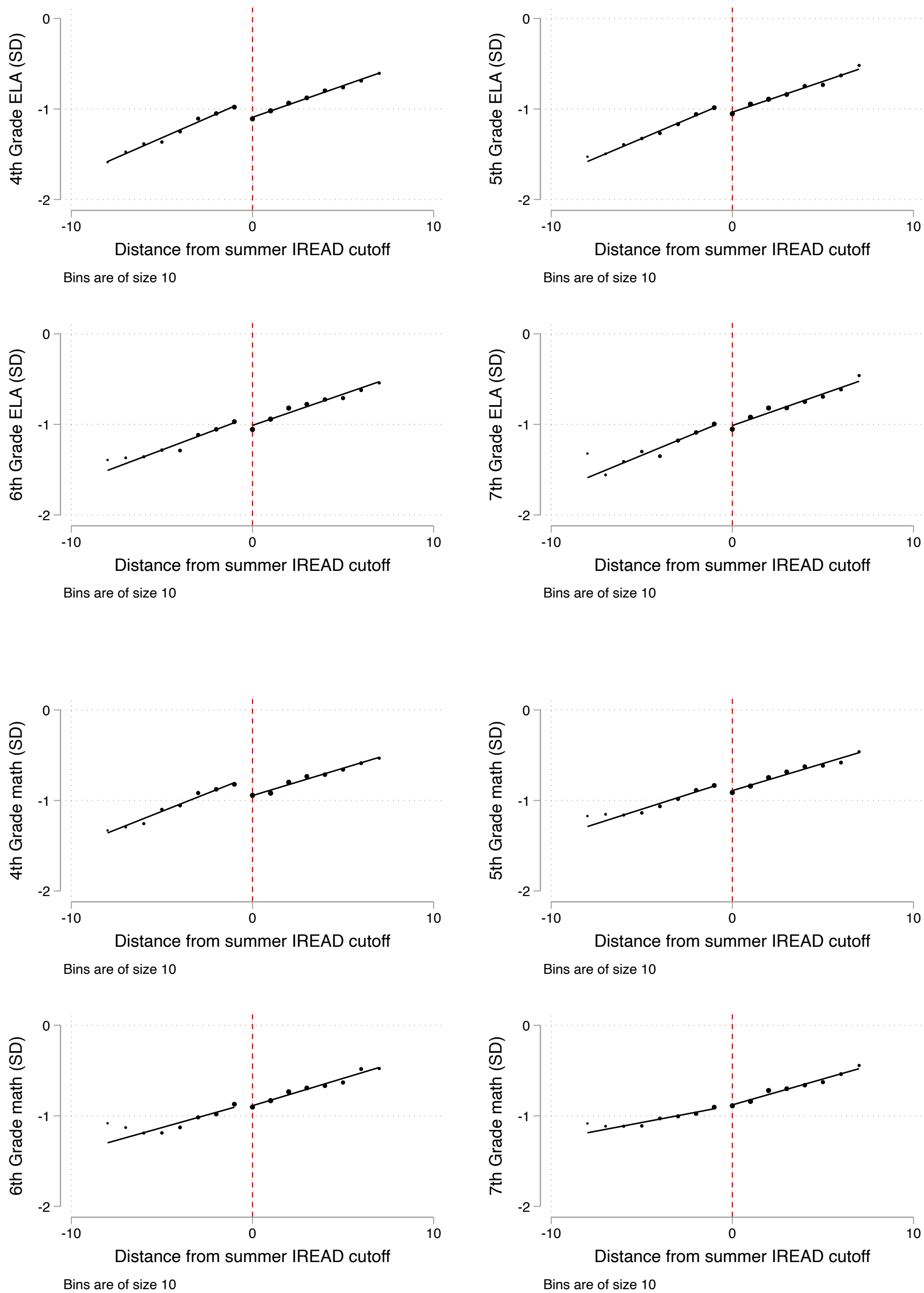
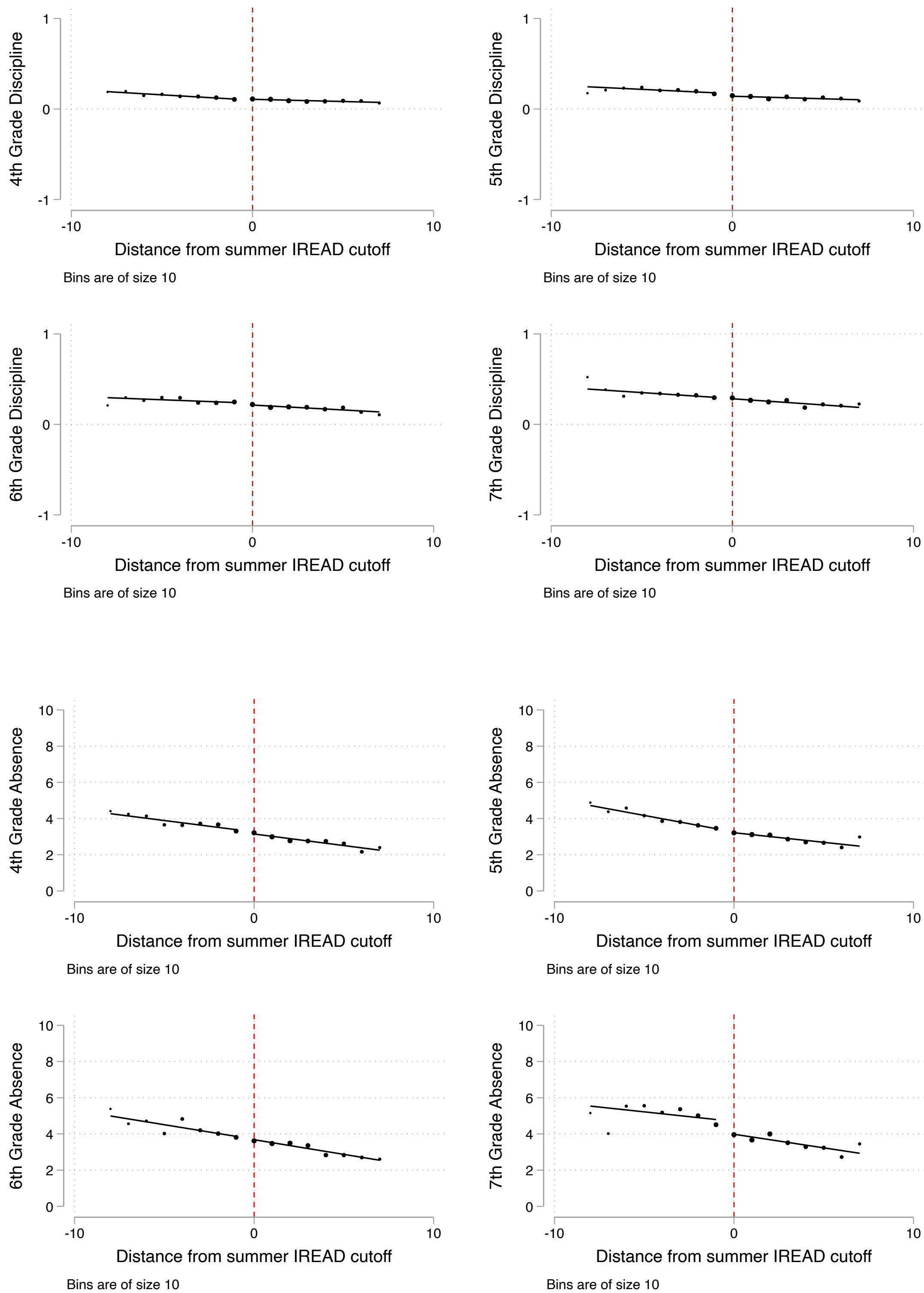


Figure 3
The Effects of Third Grade Retention on School Discipline and Absence in Fourth through Seventh Grades



Appendix Table 1
The Effects of Third Grade Retention on Student Outcomes (Quadratic Functions f(.) and g(.))

	(1) Grade 4	(2) Grade 5	(3) Grade 6	(4) Grade 7
Panel A: ELA				
Retention (Grade 3)	0.467***	0.198	0.248***	0.182
	(0.132)	(0.108)	(0.109)	(0.129)
N	8376	7016	5079	3144
Panel B: Math				
Retention (Grade 3)	0.466**	0.406**	0.124	0.140
	(0.142)	(0.124)	(0.107)	(0.135)
N	8402	7034	5088	3155
Panel C: Discipline				
Retention (Grade 3)	-0.072	0.054	0.036	-0.050
	(0.066)	(0.066)	(0.072)	(0.091)
N	8542	7164	5200	3241
Panel D: Absence				
Retention (Grade 3)	-0.655	-0.123	-1.290	0.378
	(0.973)	(0.786)	(0.908)	(1.185)
N	8519	7152	5192	3239

Note. Our models include student level controls, such achievement/school discipline/absence, sex, race/ethnicity, and FRL in the third grade. The running variable functions in equations (1) and (2), f(.) and g(.), are specified as quadratic. The sample is restricted to students who scored within 25 (between -0.52 SD and 0.37 SD) on Summer IREAD-3. Standard errors in parentheses are clustered at the school level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix Table 2
The Effects of Third Grade Retention on Student Outcomes with Different Bandwidth.

	Grade 4			Grade 5			Grade 6			Grade 7		
	(1) BW= ± 15	(2) BW=± 35	(3) BW= ± 45	(4) BW= ± 15	(5) BW=± 35	(6) BW= ± 45	(7) BW= ± 15	(8) BW=± 35	(9) BW= ± 45	(10) BW= ± 15	(11) BW=± 35	(12) BW= ± 45
Panel A: ELA												
Retention (Grade 3)	0.475***	0.521***	0.531***	0.219*	0.275***	0.270***	0.264**	0.310***	0.309***	0.206	0.256***	0.244***
	(0.108)	(0.059)	(0.052)	(0.090)	(0.053)	(0.047)	(0.094)	(0.059)	(0.052)	(0.112)	(0.072)	(0.065)
N	5131	11074	13321	4289	9199	10953	3106	6564	7728	1889	4078	4832
Panel B: Math												
Retention (Grade 3)	0.485***	0.549***	0.578***	0.396***	0.367***	0.364***	0.148	0.189**	0.185***	0.150	0.146	0.129
	(0.117)	(0.067)	(0.059)	(0.105)	(0.064)	(0.057)	(0.092)	(0.059)	(0.052)	(0.117)	(0.078)	(0.070)
N	5152	11113	13365	4302	9225	10984	3113	6575	7737	1899	4085	4840
Panel C: Discipline												
Retention (Grade 3)	-0.078	-0.067*	-0.064*	0.036	-0.004	-0.002	0.035	0.032	0.017	-0.052	-0.056	-0.052
	(0.054)	(0.032)	(0.029)	(0.055)	(0.034)	(0.031)	(0.063)	(0.040)	(0.036)	(0.079)	(0.054)	(0.048)
N	5242	11292	13570	4387	9405	11193	3189	6730	7918	1956	4208	4985
Panel D: Absence												
Retention (Grade 3)	-0.417	-0.090	0.074	-0.005	0.254	0.184	-0.960	-0.167	-0.156	0.670	0.886	0.840
	(0.800)	(0.452)	(0.404)	(0.667)	(0.427)	(0.382)	(0.774)	(0.500)	(0.457)	(1.020)	(0.705)	(0.647)
N	5229	11260	13525	4379	9393	11180	3184	6720	7906	1954	4206	4983

Note. Our models include student level controls, such achievement/school discipline/absence, sex, race/ethnicity, and FRL in the third grade. The sample is restricted to students who scored within 15, 35, and 45 Summer IREAD-3. Standard errors in parentheses are clustered at the school level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$