Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 26-1527

The Effect of Retention Threat on Student Learning Gains: Evidence from Ohio

Sarah Souders
American University

Robert Bifulco
Syracuse University

Grade retention policies create incentives for students to make adequate progress ahead of evaluation. The threat of retention thus has the potential to affect all students' academic outcomes, increasing achievement especially among students with test scores near the retention policy's threshold. By using unique data from Ohio which allows us to implement a regression discontinuity design, we find that students exposed to the threat of retention in Grade 3 experience larger learning gains than students who are not under retention threat. Gains in Grade 3 English Language Arts (ELA) are 0.077 SD and although effects partially fade over time, positive effects persist through Grade 7. Positive and lasting effects are detected across cohorts, student groups, and schools. We also examine non-test outcomes, finding reductions in student absences but limited evidence of effects on disciplinary incidents or the probability of retention beyond Grade 3. Exploration of possible mechanisms suggests that assignment to targeted literacy interventions contributes to the retention threat effect. We also show that retention threat effects are largest in schools facing stronger accountability pressures and with more capacity to respond to accountability incentives. This paper offers novel evidence of the broad effects of retention policy and provides results to inform early literacy policy initiatives, as many states debate the role of retention mandates in such initiatives.

VERSION: July 2026

Suggested citation: Souders, Sarah, and Robert Bifulco. (2026). The Effect of Retention Threat on Student Learning Gains: Evidence from Ohio. (EdWorkingPaper: 26-1527). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/ps8z-qv79

The Effect of Retention Threat on Student Learning Gains: Evidence from Ohio*

Sarah Souders†
American University

Robert Bifulco
Syracuse University

Abstract: Grade retention policies create incentives for students to make adequate progress ahead of evaluation. The threat of retention thus has the potential to affect all students' academic outcomes, increasing achievement especially among students with test scores near the retention policy's threshold. By using unique data from Ohio which allows us to implement a regression discontinuity design, we find that students exposed to the threat of retention in Grade 3 experience larger learning gains than students who are not under retention threat. Gains in Grade 3 English Language Arts (ELA) are 0.077 SD and although effects partially fade over time, positive effects persist through Grade 7. Positive and lasting effects are detected across cohorts, student groups, and schools. We also examine non-test outcomes, finding reductions in student absences but limited evidence of effects on disciplinary incidents or the probability of retention beyond Grade 3. Exploration of possible mechanisms suggests that assignment to targeted literacy interventions contributes to the retention threat effect. We also show that retention threat effects are largest in schools facing stronger accountability pressures and with more capacity to respond to accountability incentives. This paper offers novel evidence of the broad effects of retention policy and provides results to inform early literacy policy initiatives, as many states debate the role of retention mandates in such initiatives.

Keywords: Grade Retention, Accountability, Early Literacy, Policy Evaluation

* We thank Deven Carlson, David Figlio, Ying Shi, Vlad Kogan, Evan Riehl, Michah Rothbart, anonymous reviewers, and participants in the APPAM, AEFP, American University, and Syracuse University seminars for their helpful feedback. This study uses data provided by the Ohio Department of Education and Workforce. We thank Eben Dowell, Candi Patterson, and L.M. Clinton for their assistance. All findings and conclusions are solely the responsibility of the authors and do not represent the views of the Ohio Department of Education and Workforce. This study was approved by the Syracuse University Institutional Review Board (Protocol #23-067).

† Correspondence concerning this article should be addressed to Sarah Souders, 323 Kerwin Hall, American University Department of Public Administration and Policy, 4400 Massachusetts Ave. NW, Washington, DC 20016 (Email: ssouders@american.edu).

INTRODUCTION

In recent years states have adopted and repealed mandatory grade retention policies, renewing debates about the benefits and costs of mandatory retention.[3] Opponents argue that retention is costly[4], exacerbates inequality as parents unequally advocate for grade promotion (LiCalsi et al., 2019; Greene & Winters, 2009), and increases dropout rates and risky behaviors due to stigmatization of being held back. Advocates contend that retention policies help to improve educational outcomes by ensuring that students are academically prepared for grade-level work, providing accurate information to parents about their child's academic performance, and enabling teachers to create challenging learning environments for students.

Despite a large literature on grade retention, most empirical research focuses on the effects of grade repetition, examining effects for retained students or their affected peers. These studies consistently find positive short-run achievement gains from early-grade retention that fade over time (Hwang & Koedel, 2025; Hwang & Koedel, 2023; Mumma & Winters, 2023; Martorell & Mariano, 2018; Schwerdt et al., 2017; Greene & Winters, 2007, 2009; Mariano et al. 2013; Jacob & Lefgren, 2004, 2009; Manacorda, 2012; Özek, 2015; Gottfried 2013a; Gottfried 2013b). These estimates, while replicated across numerous settings, capture the effect of repeating a grade for a relatively small group of students, offering limited guidance on the

[3] In 1998, California became the first state to formally mandate test-based retention (Redding 2023). Eighteen states adopted mandatory retention in early elementary grades since the year 2000 (ExcelinEd 2024). Four more states—Nevada, West Virginia, Louisiana, and Colorado—will add retention policies in Grade 3 by 2029. Three states—Michigan, Ohio, and Oklahoma—have repealed test-based retention in Grade 3 since the early 2000s. Ohio's repeal began in the 2024-2025 school year, several years after the period of this study.

[4] Eren et al. (2017) estimate the direct cost to society of retaining around 450,000 students per year may exceed $5.6 billion (in 2012 dollars). Retained students also face costs as they delay labor market entry by a year. Winter (2023) provides updated estimates of the cost of retention and estimates the cost to taxpayers and to students retained by the policy. The respective costs are US$7,422 in 2024 dollars to taxpayers, and US$19,953 to high school dropouts and US$69,022 to eventual high school graduates.

desirability of retention policies for broader populations. One argument for grade retention policies is that the threat of retention creates incentives for parents, students, and teachers to ensure students make adequate progress prior to evaluation. A mandatory test-based retention policy may increase parent monitoring and involvement in a child's education, motivate interventions by school officials leading up to "gate" grades, and increase the motivation of students to achieve above the threshold level. Thus, retention policy likely influences far more than the 2–10% of students ultimately retained each year (Education Commission of the States, 2024).

Understanding whether retention threat improves and sustains learning is especially important in the current policy environment, where eighteen states use test-based retention as a culminating K–3 literacy intervention and 42 states have considered adding or removing retention components in recent years (ExcelinEd, 2024; Neumann et al., 2023). If effective, retention threat could help address large-scale early literacy challenges—roughly half of K–2 students across 43 states were "not on track" in recent years to meet Grade 3 literacy benchmarks (Arundel, 2024)—and, in turn, improve long-run achievement, attainment, and earnings (Heckman & Mosso, 2014).

This paper examines the impact of early-grade retention mandates by investigating how the threat of retention influences student learning in Grade 3 and whether these gains translate into sustained improvements in later outcomes. Exploiting the fact that students in Ohio can avoid retention at the end of Grade 3 by demonstrating proficiency on a test administered in the fall of Grade 3, we use a regression discontinuity design to estimate the effects of being subject to retention threat during third grade. Specifically, we compare students just below the retention

threshold in the fall, who face the risk of retention, with those just above the threshold, who do not.

Retention threat increases Grade 3 achievement in both ELA and math, with effects of 0.077 SD and 0.037 SD, respectively, for students just below the fall threshold. ELA effects persist through Grade 7 (~0.03–0.04 SD), while math effects fade earlier. Effects are larger and more sustained for students identified through other assessments as below grade-level reading expectations. These results are robust to various specifications of the regression discontinuity model. Persistence of ELA effects, as well as positive Grade 3 math effects, which do not factor into retention decisions, suggest that improved achievement reflects real learning rather than short-term accountability responses. Evaluation of non-academic outcomes indicates that students under retention threat are less likely to be absent from school. We do not find statistically detectable effects on disciplinary instances or the probability of retention beyond Grade 3.

While effects are similar across student groups defined by gender, race, and socioeconomic status, achievement effects are larger in districts with higher levels of pre-policy spending and in schools with larger shares of students predicted to be at risk of retention. The finding that achievement effects are larger in districts with higher levels of spending suggest that capacity for schools to respond to incentives created by the retention mandates and related school accountability measures may be important. Schools with the largest share of students predicted to be at risk of retention face the strongest school accountability pressures, which might at least partially explain why these schools do more to improve the achievement of students under threat of retention.

Finally, we find that being under threat of retention makes it more likely that a student is assigned to an individualized reading intervention, suggesting that such services may be a mechanism through which retention threat results in improved achievement. However, schools with higher spending and schools with the largest share of at-risk students, where we see the largest achievement effects, are not the schools where retention threat leads to the largest increases in assignment to individualized reading interventions. This last result suggests that assignment to individualized interventions is not the only mechanism through which effects are realized.

This study makes contributions to at least two existing literatures. First, it shifts the attention of the literature on retention mandates from the effects of grade repetition to the effects on broader sets of students. Because in most contexts it will effect a much larger set of students, the effects of retention threat are potentially more important than the direct effects of being retained. Evidence of effects of grade retention on non-retained students, however, is limited. A small set of studies show that retained students can influence the attendance and achievement of non-retained peers (Gottfried 2013a, 2013b) and younger siblings (Figlio et al. 2023). Another small set of studies focus on the effects on entire cohorts of students in schools with retention mandates, but these studies lack convincing causal research designs (Battistin and Schizzerotto 2019; Hong and Hong 2021; Roderick, Jacob, and Bryk 2002). The regression discontinuity design that we employ provide more compelling evidence. A recent study using data from Michigan employs a similar strategy to ours and finds that being flagged for retention increases next-year reading achievement by approximately 0.045 SD (Berne et al. 2026).

Second, this study examines retention mandates in the context of a larger early literacy initiative that includes school level accountability measures. Thus, the paper contributes to a

literature on responses to accountability pressure. Prior work shows that accountability systems can induce behavioral responses among both students and organizations by linking consequences to measured performance (Butts and Roman 2007; De Angelis 2016; Harris 2012; Mears and Butts 2008). Survey evidence suggests that students facing retention threat report greater attention to classwork and more time spent studying (Roderick and Engel 2001). However, these increases in effort may not translate into improved achievement, as high-stakes environments like those created by test-based retention can induce anxiety, stress, and disengagement that offset potential gains (Anderson et al. 2005; Heissel et al. 2021; Stanovich 2009; Walberg and Tsai 1983; Marsh et al. 2017). We present evidence that students increase both attendance and achievement in response to retention threat.

Prior work also shows that schools and teachers respond to performance incentives by reallocating effort and targeting supports (Jacob and Levitt, 2003; Jacob 2005; Figlio and Getzler, 2006; Reback 2008; Craig et al. 2013; Riehl and Welch, 2023). Our results suggest that school-level accountability may complement student level accountability. Particularly we find that the effect of retention threat is larger in the years after school accountability measures linked to early literacy were adopted. Also, our results on school level heterogeneity suggest that the effects of retention threat are larger in the schools likely to be subject to the greatest pressure from school level accountability measures.

The remainder of the paper proceeds as follows. First, we describe Ohio's early-grade retention policy and the broader literacy initiative that underpins it, situating the policy alongside similar reforms in other states. Next, we present the data and analytic methods before turning to the main results and an exploration of the underlying mechanisms. The paper concludes with a discussion of the findings' implications for policy and future research.

EARLY-GRADE RETENTION POLICIES IN OHIO AND OTHER STATES

In 2012-13, Ohio adopted the Third Grade Reading Guarantee (TGRG) policy. The policy included test-based promotion tied to English language arts (ELA) achievement beginning in the 2013-14 school year.[5] The Grade 3 ELA test is administered in the fall between late October and early November and in the spring between the end of March and middle of April. It is this Grade 3 ELA test that is used to determine promotion to fourth grade. The promotional standard and test used to assess proficiency have changed over time, but the policy of administering the Grade 3 ELA test to all students in the fall and spring of third grade was in place throughout the period of this study.[6] A student secures promotion to fourth grade if he or she passes the ELA test in either the fall or spring of third grade.

Our study period covers the years 2014 through 2017. The ELA test used in 2014 and 2015 was the Ohio Achievement Assessment (OAA). In 2016 and beyond, promotion was based on achievement on the Ohio State Test English Language Arts (OST ELA) exam. In 2016 and 2017 only a student's score on the reading subsection of the test was used to determine promotion. Scores required to avoid retention were increased incrementally overtime closer to levels identifying proficiency.[7]

Teachers and parents are provided a three-page report on the results of the Grade 3 ELA test administered in the fall via email, a secure learning management system message, or as

[5] Throughout the rest of the paper the year in spring is used to indicate school year. For example, the 2013-14 school year will be referred to as the 2014 year.

[6] Only about 2.7% of all first time third grade students test in the fall but not the spring and 94.3% of all first-time Grade 3 students test in the fall.

[7] To evaluate whether the change in cut score matters for estimates of the response to retention threat, Table 3 shows effects by each Grade 3 cohort.

printed pages. On the front page, there is a thermometer display of a student's test score where the retention threshold score is marked and a description of grade retention for students who fail to achieve at or above the retention threshold is provided (see Appendix Figure A1). Students at-risk of retention, and particularly those within a few points of the retention threshold, are easily identified.

In addition to grade retention, components of Ohio's Third Grade Reading Guarantee include a requirement that schools monitor and support student reading progress in grades K-3, school and district-level accountability measures, and financial incentives for districts tied to the number of students who achieve reading proficiency by the end of third grade. In grades K-3, schools must administer reading diagnostic tests before September 30th to determine whether students are "On Track" or "Not On-Track" to meet the Grade 3 retention threshold. The tests used for this purpose are different than the ELA OAA or OST that is used to determine promotion. "On Track" students are reading at the level of proficiency expected at the end of the previous grade. Students "Not On-Track" must receive an individualized reading improvement plan (RIMP) that is executed throughout Grade 3. The instructional services selected for a student with a RIMP rely on the judgment of the student's teacher and parents. Appendix Table A1 lists all 23 approved reading interventions.[8]

Ohio's school accountability system uses three measures to assign 1-5 stars on the Early Literacy Component of school and district-level report cards. These measures—"Improving K-3

[8] By law, the RIMP must include identification of the student's specific reading deficiency; a description of proposed supplemental instruction services that will target the student's identified reading deficiencies; opportunities for the student's parents or guardians to be involved in the instructional services; a process to monitor the implementation of the student's instructional services (progress monitoring); a reading curriculum during regular school hours that: (1) helps students read at grade level; (2) provides scientifically based and reliable assessment; and (3) provides initial and ongoing analysis of each student's reading progress; a statement that unless the student attains the appropriate level of reading competency by the end of Grade 3, the student will be retained unless otherwise exempt (Ohio Department of Education 2022).

Literacy," "Promotion to Fourth Grade," and "Proficiency in Third Grade Reading"—make up 25%, 35%, and 40%, respectively, of the Early Literacy Component. Students who score above the retention threshold and the proficiency cut score on either the fall or spring administration count positively toward "Promotion to Fourth Grade" and "Proficiency in Third Grade Reading." Only students identified as "Not On-Track" in the fall of Grade 3 who achieve proficient scores in the spring count positively toward the "Improving K-3 Literacy" measure. Appendix Figure A2 shows how the "Improving K-3 Literacy" measure is calculated. Scores on the Early Literacy Component make up 14.266% of a school or district's overall 1–5-star report card rating. The introduction of this measure in 2015-16 placed more accountability pressure on schools serving larger shares of lower-achieving students at risk of grade retention. Beginning in 2014–15, the state introduced bonus payments tied to Grade 3 ELA proficiency.[9] Figure 1 summarizes the timing of key policy components during the Grade 3 year, including identification, intervention, accountability, and financial incentives.

Ohio's approach is not unique. Eleven other states have implemented similar statewide early literacy policies, and four more are set to adopt them by 2029.[10] Like Ohio, policies in eight other states include test-based retention in Grade 3 and require monitoring of student reading progress in grades K-3, provide additional academic supports for students below proficiency benchmarks, involve parents and teachers in student reading plans, and include additional programming to increase student reading proficiency (ExcelinEd, 2024). Arizona and Colorado's

[9] Bonus payments are included in the state's foundation aid payments to districts and are allocated as follows: School % proficient X 0.075 X formula amount X # proficient students. For example, a district with 85% of Grade 3 students scoring proficient, $4,023.20 state aid allocation in 2016-17, and 145 proficient Grade 3 students would receive approximately $37,100 in the 2016-17 school year.

[10] In 2023, Louisiana, Arkansas, West Virginia, and Nevada approved policies to retain below proficient third grade students beginning in future years (Louisiana HB12, 2023; Arkansas LEARNS Act, 2023; West Virginia HB3035, 2023; Nevada Assembly Bill 400, 2023).

policies additionally require schools to distribute state aid dollars to incentivize K-3 reading proficiency.[11] Among these states, Colorado most closely mirrors Ohio, incorporating monitoring, retention, school report card accountability, and financial incentives (Colorado READ Act, 2012). Across states, the timing and frequency of the assessments linked to retention mandates varies. North Carolina, like Ohio, administers the high-stakes end-of-grade promotion assessment in both the fall and spring of Grade 3, allowing students to demonstrate proficiency earlier in the year (ExcelinEd, 2024). In most other states, the initial test is administered in the spring, with summer retakes offered after additional instructional support.

Whether at-risk students are ultimately retained depends on how policies are implemented, including the use of exemptions, alternative pathways to promotion, and the consistency of enforcement. For example, in Indiana, the probability of retention conditional on failing to meet the promotion threshold declined from nearly 60% in earlier cohorts to about 15% between 2017 and 2022.[12] In the first year of implementation in Tennessee, while nearly 60% of students were flagged for retention only about 1.2% were retained (Wegner 2024). In Ohio, roughly one-third of students are identified as at-risk of retention based on fall assessments, about 15% score below the threshold on the spring assessment, and approximately 3% are ultimately retained, a rate that has remained stable over time. Alternative routes are used relatively infrequently, with fewer than 2% of students promoted through a summer retake, roughly 2% meeting the standard through an approved alternative assessment, and about 3–5% promoted using good cause exemptions each year.[13]

[11] A.R.S. §15-701; Colorado READ Act, 2012.
[12] Authors' calculation based on Indiana students in third grade between 2011-12 and 2021-2022 with an IREAD-3 score in the summer following their third-grade year and whose score on the IREAD-3 test was within 50 points of the IREAD-3 446 cut score for retention.
[13] Authors' calculation using Ohio restricted data detailing student promotion reasons.

DATA, SAMPLE AND ANALYSIS

We use student-level data from the Ohio Department of Education and Workforce from the 2014–2022 school years. These data include information on student characteristics such as grade, race/ethnicity, gender, economically disadvantaged status, whether a student has an Individualized Education Plan (IEP) or participates in English language learning programs and student test scores. Records of student participation in approved reading programs in Grade 3, student disciplinary incidence involvement, attendance, and promotion reason also come from the Ohio Department of Education & Workforce.

We follow cohorts of students enrolled in Grade 3 for the first time in 2014 to 2017 through Grades 4-8.[14] Our Grade 3 sample is limited to first-time Grade 3 students, as students retained once cannot be retained again, and to students who have both fall and spring ELA test scores. We exclude students enrolled in joint vocational schools, education service center districts, or state schools, as well as students enrolled in a school or district for fewer than 100 hours. This sample includes 1,869 public schools and 503,783 students. All cohorts are followed through Grade 6 and the earliest cohort through Grade 8. However, we drop Spring 2020 test scores from the sample, as only a subset of schools and students participated in testing, and so not all students in our Grade 3 sample have test scores in each grade.

The first column of Table 1 provides summary statistics for all first-time Grade 3 students in our sample. About 34 percent of all students face retention threat based on their fall test scores and roughly half of these students also do not meet the threshold in the spring.[15] 3.1 percent of

[14] Among first-time Grade 3 students who test in the fall, 97.73 percent participate in the spring and characteristics of participants are similar to those of all students.

[15] Students who do not pass the retention threshold in the fall or the spring are offered an additional chance to meet the cut score on the Grade 3 ELA test with a summer administration of the test. Additionally, it is up to each district to provide students the opportunity to meet the retention threshold due to adequate performance on alternative assessments. The Ohio Department of Education has an approved cut score on the Iowa, NWEA – MAP, Terra

students in these cohorts are retained in third grade. Nearly 24 percent of students are determined as "Not On-Track" to exceed the Grade 3 threshold in the fall of Grade 3.[16] Student test scores are standardized to mean 0, SD of 1 within administration by subject by year among the full sample of test-takers. First-time Grade 3 test takers have spring test scores 0.020 and math test scores of 0.021. Mean values are larger than 0 because test scores are standardized among all test takers, including students retained Grade 3 the previous year and who are excluded from this analytic sample.

Primary Analysis

We use a sharp regression discontinuity design to examine the effect of retention threat on student outcomes. Specifically, the retention threshold on the fall test identifies who is subject to retention threat during third grade. Students who score just below the retention threshold may be retained unless their spring or summer test score exceeds the threshold, while students just above the threshold are not subject to the risk of retention.

To implement the design, we estimate the following regression model:

$$Y_{icg} = \alpha + \tau Threat_{ic} + f(StdFallTest_{ic}, Threat_{ic}) + \beta OnTrack_{ic} + \gamma_c + \varepsilon_{icg}$$

Y_{icg} indicates an outcome for student i in cohort c in grade g. Outcomes include standardized spring test scores in Math or ELA in Grades 3-8, or discipline, retention, or absence measures in Grades 3-8. Disciplinary incidences are measured as a dummy variable equal to 1 if a student is

Nova, STAR Reading, and the iReady assessment. Students with adequate performance on any one of these alternatives may be promoted to 4th Grade. Between 2014-2017, 1.3-1.5% of Grade 3 students were promoted using an alternative test option.

[16] The "On-Track" category overlaps but is different than that category of students at risk of retention. There are students at-risk of retention who are designated "On-Track", and students who are designated "On-Track" who are at-risk of retention. Importantly "on-track-status" is determined by test administered before whether the student is subject to retention risk during third grade is determined.

involved in an incident recorded in the EMIS system and 0 otherwise.[17] We identify students as retained, using a dummy variable equal to 1, if they are tested in the same grade in consecutive calendar years and 0 if they advance to the next grade. Attendance is measured as the share of total time in a school in a year a student is absent. $Threat_{ic}$ is a dummy variable that equals 1 if a student faces the threat of retention based on their fall Grade 3 test results and 0 otherwise. The running variable is a student's fall Grade 3 ELA or reading test score standardized to the retention threshold in each year. Across years, tests, and thresholds, all students who face the threat of retention in Grade 3 have a $StdFallTest_{ic} < 0$, and all students who avoid retention have a $StdFallTest_{ic} > 0$.

We pool across Grade 3 cohorts and include cohort fixed effects, (γ_t), so that comparisons are made within year (where the threshold and test is consistent). Our primary specification includes $OnTrack_{ic}$, which is an indicator equal to one if a student's fall Grade 3 diagnostic assessment classified them as reading at the expected Grade 2 proficiency level and zero otherwise. We include this measure because "On Track" status is determined using a separate diagnostic assessment, is not perfectly correlated with retention threat status, and may independently trigger literacy interventions during the Grade 3 year.[18] We also show that results are similar without conditioning on a student's $OnTrack_{ic}$ status. Appendix Table A2 presents a

[17] Disciplinary incident reason codes include truancy, fighting/violence, vandalism or damage to school property, use, possession, sale or distribution of firearms or any other weapon, explosive, incendiary, or poison gas, use, possession, sale or distribution of tobacco, alcohol, or any other drugs, bomb threats, disruptive behavior, harassment, unwelcome sexual conduct, or serious bodily injury. Most incident reports are for disruptive behavior.

[18] Among students identified as "Not On-Track," 28% are not under retention threat based on the fall ELA assessment, while 72% are under threat. Conversely, 21% of students identified as "On Track" are under retention threat. Under Ohio Revised Code Section 3313.608(B)(2)(a), students identified as "Not On-Track" must receive parent notification and a Reading Improvement and Monitoring Plan (RIMP) describing supplemental instructional services and supports. Because diagnostic assessments are administered by September 30th and the fall Grade 3 ELA assessment is administered in late October, these supports are unlikely to influence performance on the fall ELA assessment itself, though they may affect the services students receive throughout the remainder of the Grade 3 year. We therefore control for "On Track" status to account for an alternative pathway through which students may receive additional support and information about academic progress.

sequence of increasingly saturated specifications. We begin with models including only cohort fixed effects, followed by our preferred specification adding baseline "On Track" status. We then add baseline demographic controls and re-estimate the model using updated bandwidths. Results are substantively similar across specifications.

Estimates of τ from the above regression reflect two potential channels through which retention threat may affect later-grade outcomes: a direct effect of facing the threat of retention and an indirect effect operating through an increased likelihood of actual Grade 3 retention. However, because exposure to retention threat increases the probability of retention in our sample by only about 1 percentage point, the indirect component attributable to grade repetition is necessarily quite small. To gauge the possible magnitude of this channel, we draw on estimates from Indiana, which produces among the largest achievement effects in the retention literature, with Grade 4–6 ELA impacts of roughly 0.50, 0.35–0.40, and 0.25–0.30 SDs, respectively (Hwang & Koedel, 2023). Multiplying these estimates by the approximately 1 percentage point increase in retention induced by retention threat in our sample implies expected effects of only about 0.003–0.005 SD in later grades. These implied effects are substantially smaller than the persistent achievement gains we observe, suggesting that differences generated by retention are unlikely to explain our longer-run results. We therefore present the reduced-form estimates of retention threat that include the very small effect of the increased probability of retention on later-grade outcomes.[19]

[19] One approach to separating the direct effects of retention threat from effects operating through actual grade repetition is to combine the reduced-form RD estimates presented here with fuzzy RD estimates of the effect of actual retention among students participating in Ohio's summer testing option. Implementing such an approach requires additional assumptions regarding conditional independence and sequential ignorability (Baron and Kenny, 1986; Imai et al., 2010), as well as additional estimation procedures, including bootstrap methods to estimate standard errors for decomposed effects. Although we implemented this approach in supplementary analyses, decomposition of effects resulted in only limited differences from the reduced-form estimates, likely due to the relatively low rate of induced retention. We therefore focus on the reduced-form effects of retention threat throughout the paper.

We estimate the regression above within outcome-specific optimal bandwidths around the retention threshold chosen using the robust data-driven procedure outlined in Calonico et al. (2014). For each outcome, bandwidths are held constant across grades to ensure our examination of whether effects fade out over time is not confounded by changes in selected bandwidths. Separate bandwidths are selected for each model specification used. The second column of Table 1 describes students within the optimal Grade 3 ELA bandwidth estimated using a model with only cohort fixed effects. Compared to all first time Grade 3 students, students within the bandwidth are more likely to face retention threat in the fall and to be identified as "Not On-Track" than are the full sample of first time third graders. Also, a larger share of students in the bandwidth than in the full sample do achieve the scores above that required for promotion. Students close to the retention threshold resemble all students in the sample on race, and on economic disadvantage, English language learner, and disability status. Average spring test score outcomes for students close to the cut-off are lower in both ELA and Math than for all first time Grade 3 students. In line with evidence that higher-ordered polynomials of the running variable should not be used to estimate regression discontinuities within optimally selected bandwidths (Gelman & Imbens, 2019), we specify the running variable linearly in our preferred specification and cluster standard errors at the level of the running variable (Kolesár and Rothe 2018).

Student-Level Heterogeneity

Certain groups of students might respond differently to the threat of retention than others. The literature on stereotype threat suggests that individuals from groups subject to negative stereotypes might experience greater anxiety while taking achievement tests than individuals from other groups (Steele & Aronson, 1995). Also, low-income parents might be less able to provide additional supports outside of school for students at risk of retention. This might affect

the ability of students from low-income families to improve their achievement in response to the retention threat. Also, some students are more likely to qualify for good-cause exemptions from retention and thus be subject to less pressure from retention threat. We separately assess effects for groups defined by sex, race, socioeconomic status, disability status, and participation in English learner programs. Within each group, we interact indicators of student type with all variables in our RD model to evaluate differential responses to retention threat. We use an F-test to detect differential responses to retention threat among White, Black, Hispanic, and Asian students in a fully interacted model.

School-Level Heterogeneity

Curriculum and programs to improve literacy are at the discretion of teachers and principals. Thus, there is likely variation in how schools respond to a student's risk of retention. Some schools might emphasize school- or grade-wide responses that impact all students and not only those under threat of retention. These efforts may include changes to curriculum or pedagogy or shifting the highest quality teachers to early grades (Chingos & West, 2011; Cohen-Vogel, 2011; Fuller & Ladd, 2013; Grissom et al., 2017). Others might increase individualized instruction for those under threat of being retained. In either case, resource constraints might limit the ability of schools to effectively improve student learning. Also, given the design of school-based accountability measures in Ohio, schools serving the largest shares of students under retention threat may face the strongest accountability and financial incentives to improve outcomes for these students.

We evaluate whether resource capacity and concentration of student need influence the effect of retention threat using two pre-policy measures: district per-pupil operational expenditures and predicted school-level retention risk among first-time Grade 3 students. To

avoid endogeneity concerns, both expenditures and predicted retention risk are measured prior to policy implementation using 2012–2013 data. Specifically, we estimate the school-level share of students under retention threat using student demographic, attendance, and disciplinary characteristics, including interaction and nonlinear terms, observed before the policy took effect. The model provided strong fit, with an R^2 of 0.915 and a 0.96 correlation between predicted and observed school-level threat rates.[20]

For both of these school level measures, we split schools into quintiles. Quintile 1 is a dummy variable set equal to 1 if a school has the lowest per pupil expenditures (or school-level retention risk) and 0 otherwise. Quintile 5 is set equal to 1 if a school is in the top 20 percent of per pupil expenditures (or school-level retention risk), indicating that the school had the greatest amount of pre-policy spending or largest predicted share of Grade 3 students under threat of retention. We estimate effects for students in quintiles defined by spending and share under threat separately and estimate a model that includes both spending- and threat-based quintiles to assess the response to retention threat given each school's combination of resources and need. Specifically, we interact quintile indicators with the treatment and running variables in our main RD model. Robust standard errors are clustered at the school level.[21]

RESULTS

We begin our presentation of results with evidence supporting the validity of our RD design. Next, we present the main effect estimates and robustness analyses, followed by an

[20] We also estimated an alternative pooled model using data from 2009–2013. However, the 2012-2013-only specification demonstrated superior predictive performance.

[21] We also estimated models including school-by-year fixed effects. Findings for heterogeneity by district spending and the predicted share of students at risk of retention are substantively unchanged.

analysis of effect heterogeneity and then an exploration of a possible mechanism for the estimated effects.

Validity of RD Design

The key assumption of the RD design is that students cannot precisely control their outcome on the fall Grade 3 tests which puts them under threat of retention (Lee and Lemieux, 2010). If the risk of retention among students scoring near the cut point is indeed as good as random, the density of the running variable should be smooth around the cut point. Appendix Figure A3 presents the distribution of the running variable. Although the distribution exhibits several sharp changes in density, including near the cutoff, these patterns are present throughout the distribution rather than concentrated at the threshold. Consistent with this visual evidence, a density test similar to McCrary's (2008) using a local polynomial estimator fails to reject the null hypothesis of no manipulation around the cutoff ($t=0.78$, $p=0.40$). To further assess whether local changes in density influence the results, we estimate regression discontinuity donut specifications that sequentially exclude observations within 0 to 0.1 SDs of the threshold from the preferred bandwidth for each outcome. Estimates for outcomes with statistically significant effects are shown in Appendix Figure A4.

If students are indeed randomly distributed around the retention threshold, then students on either side of the threshold should be observationally similar. Table 2 presents tests of covariate balance analogous to the main specification, where predetermined student characteristics measured at the beginning of Grade 3 replace spring achievement outcomes as the dependent variables. Results show no significant differences between the treated and untreated groups around the retention threshold for demographic characteristics, predicted spring ELA and math achievement, or attrition. The only discontinuity appears in students' fall Grade 3 "On

Track" status, where students under retention threat are less likely to be identified as "On-Track" based on the earlier diagnostic assessment. Figure 2 presents these results graphically. We additionally regress treatment status on student covariates and conduct joint F-tests of covariate significance. These tests similarly indicate no systematic differences between treated and untreated students around the cutoff and are reported in Appendix Table A3.

Attrition is not an issue for our analysis of third grade outcomes. Ninety-six-point five percent of students in the entire study sample test in both the fall and spring of Grade 3, and within the optimal bandwidth for our ELA outcome, only 1.9 percent of students who do not face retention threat based on their fall test score and 2 percent of those who do face retention threat do not have spring Grade 3 test score outcomes observed. Attrition is greater for later grade outcomes. Appendix Table A4 reports tests of differences around the threshold using indicators for missing test scores by grade and subject within the subject-specific bandwidths. Results in this table show that differences in attrition between students just below and just above the promotion cutoff are tiny—generally on the order of ±0.5 percentage points across ELA, math, absence, and disciplinary outcomes through Grade 8. Although a few disciplinary attrition estimates are marginally statistically significant, the pattern provides little reason for concern differential attrition biases the effect estimates.

Main Effects

Table 3, Panel A presents estimates of retention threat on ELA spring test score outcomes in Grades 3-8, pooled across all Grade 3 cohorts. Students who face retention threat in the fall score 0.077 SD higher on the spring Grade 3 ELA test compared to their peers who narrowly met the promotion threshold in the fall of Grade 3. Effects persist from fourth through seventh grade, with students under the threat of retention achieving 0.026 SD, 0.031 SD, 0.039 SD, and 0.026

SD more than their peers who did not face retention threat. Results are statistically significant at the 0.01 level through Grade 6 and at the 0.05 level in Grade 7. The Grade 8 estimate (0.023 SD) remains positive but is not statistically significant at conventional levels. [22] Figure 3 presents a visual representation of ELA estimates by grade. In each subfigure, gray dots represent the mean grade-level outcome within quantile-spaced bins of the running variable, i.e. bins that contain an equal number of observations, and the fitted relationship corresponds to the primary specification.[23] Across grades, Figure 3 shows that students just below the threshold who faced retention threat in Grade 3 achieved higher ELA scores than students just above the threshold who did not face retention threat.[24] If retention threat primarily affected effort on the high-stakes Grade 3 ELA exam, estimated effects would be unlikely to extend into later grades. The persistence of positive effects through Grade 7 suggests that retention threat influenced more than short-run test-taking effort alone.

A separate issue regards changes in student effort to learn during the school year. Of course, differences in student effort during the year are an important mechanism through which retention threat is expected to influence student achievement, and thus, is not a potential source of bias in our impact estimates. However, it is difficult empirically to distinguish whether our impact estimates reflect changes in effort by students subject to retention threat or changes in effort by those not subject to threat, relative to a counterfactual without a retention mandate. We believe that it is much more plausible that our estimated effects represent changes in effort

[22] Students in all cohorts are observed through the end of sixth grade. Only students in the 2014 cohort are observed up to Grade 8.

[23] See Appendix Figure A4 Panel A for results estimated within bins defined by equal increments of the running variable and Appendix Figure A4 Panel B for results estimated within a limited number of bins (10) on each side of the cutoff.

[24] Appendix Figure A5 presents results the estimated effect of retention threat on Grade 3 ELA spring outcomes by cohort showing results using both the raw scale test score and the standardized test score as the running variable.

among students subject to retention threat. The retention mandate changes incentives only for those subject to the threat of retention. Those not subject to threat under the policy have the same near-certainty of being promoted in the absence of the retention policy as they have under the retention policy. Additionally, accountability metrics and financial incentives at the school-level are tied specifically to improvements among students identified as at-risk of retention in the fall of Grade 3. Thus, there is not a strong reason to believe that the policy affects the effort level of those not subject to retention threat. Nonetheless, in Appendix B we provide evidence consistent with the interpretation that our impact estimates reflect changes in effort among students subject to retention.

Table 3 Panel B examines variation in responses to retention threat across cohorts. The 2013–2014 school year marked the first cohort subject to mandatory retention under the policy. Estimated effects for early cohorts are comparatively modest, particularly in later grades. Beginning in 2015–2016, districts became eligible for bonus payments tied to the share of at-risk students promoted to Grade 4, increasing the financial stakes associated with student performance. In 2016, Ohio also began publicly reporting the Improving K-3 Literacy Measure on school report cards, further increasing the visibility and accountability pressure surrounding early literacy outcomes. Consistent with these policy changes, estimated Grade 3 effects increase from 0.056 SD in 2014 and 0.034 SD in 2015 to 0.101 SD in 2016 and 0.078 SD 2017. Positive effects remain evident for cohorts exposed during this later accountability period, including students who entered kindergarten after Ohio expanded required K–2 literacy supports, suggesting that retention threat continued to influence achievement even as earlier-grade interventions became more widespread.

Table 4 reports pooled estimates of the effect of retention threat on math achievement, absence rates, disciplinary incidents, and the probability of later-grade retention.[25] The control group means of these measures within the bandwidth are included underneath each coefficient. Because retention decisions are based solely on ELA performance, estimated effects on math achievement are less susceptible to concerns that results reflect increased effort on the high-stakes Grade 3 ELA exam rather than broader changes in learning. Results show positive effects on math achievement in Grades 3 through 6, ranging from 0.017 SD to 0.037 SD, though estimates attenuate and become statistically insignificant in later grades.

Beyond test score effects, Table 4 examines several nonacademic outcomes, including absences, disciplinary incidents, and later-grade retention. Results show modest reductions in absenteeism among students facing retention threat. Estimated effects range from 0.1 to 0.6 percentage points, or roughly 2 to 7 percent relative to the control group mean, with statistically significant reductions in Grades 3, 5, and 8 at the 0.01 level. These results suggest some improvement in school attendance, though the magnitudes are small and are unlikely to explain a large share of the achievement effects on their own. In contrast, estimated effects on disciplinary incidents are generally small and statistically insignificant across grades. Finally, retention threat modestly increases the probability of retention in Grade 3 by 0.7 percentage points but has no detectable effect on retention in later grades.

Robustness

We perform a battery of robustness checks including specifying the running variable as a quadratic; using alternative bandwidths around the threshold; using different forms of clustering

[25] See Appendix Figures A6-A9 for alternative estimates of these outcomes. For each outcome, Panel A shows results estimated within bins defined by equal increments of the running variable and Panel B shows results estimated within a limited number of bins (10) on each side of the cutoff.

in computation of standard errors; excluding On-Track status from the specification; using nonparametric regression based on kernel or local polynomial estimators; and excluding observations nearest the retention cutoff. Importantly, none of the results are substantively sensitive to these alternative specifications. We also show alternate results using the bias-corrected RDHonest estimation procedure to account for potential specification error and provide robust confidence intervals. These results are presented in Appendix Tables A2, A5, and Figures A10-A12. Across specifications, estimated effects on ELA achievement in Grades 3–8 remain substantively and statistically similar to the main results, reinforcing evidence that retention threat produces sustained increases in ELA achievement. Although estimated effects on math achievement, absenteeism, disciplinary incidents, and later-grade retention are smaller and less precisely estimated, resulting in some variation in statistical significance across specifications, the overall pattern of results remains similar in magnitude and direction.

Heterogeneity in Effects by Student Characteristics

Table 5 presents heterogeneity in the effects of retention threat on Grade 3 and Grade 6 ELA and math outcomes by student characteristics.[26] Overall, estimated effects are broadly similar across demographic groups, with relatively few statistically significant differences by gender, race, economic disadvantage, English learner status, or disability status.[27] Instead, the clearest and most consistent source of heterogeneity is whether students were identified as On-

[26] We present results for Grades 3 and 6 because Grade 6 is the latest grade observed for all cohorts in the sample (2014–2017).

[27] Given that English language learners and students with a disability may be eligible for Good-Cause Exemptions from retention, it is perhaps surprising that the estimated impacts of failing the fall exam are as large for these students as for non-ELL and non-SWD students. In Ohio's accountability system, however, all students achieving proficiency, including ELL and SWD, improve a district's metrics on report card measures, and generate additional state funds. Furthermore, treatment of ELL and SWD students as typical students under the policy is consistent with Ohio's data which show that only a small share of students, 3-5% depending on the year, are promoted to Grade 4 with use of Good Cause Exemptions.

Track or Not On-Track in Grade 3. This pattern is notable given that Ohio's accountability system and associated financial incentives place particular value on improving outcomes for students identified as Not On-Track who are near the promotion threshold, as gains among these students can contribute to literacy improvement, proficiency, and promotion metrics. Consistent with these incentives, Not On-Track students experience larger and more persistent achievement effects than On-Track students. In Grade 3, estimated ELA effects are 0.102 SD for Not On-Track students compared to 0.063 SD for On-Track students, and the difference between groups is statistically significant at the 5 percent level. By Grade 6, this pattern becomes even more pronounced: Not On-Track students experience a 0.081 SD increase in ELA achievement, while the corresponding estimate for On-Track students declines to 0.016 SD and is no longer statistically distinguishable from zero; the difference between groups is statistically significant at the 1 percent level.

Similar, though less pronounced, patterns emerge for math achievement. Math scores suggest that Black students do benefit, and perhaps more than non-Black students, from the threat of retention. Effects are not distinguishable for students categorized by economic disadvantage, English language learner participation, and disability identification. There is little evidence that effects on absences, disciplinary outcomes or retention vary across individual student characteristics. Corresponding estimates for nonacademic outcomes in Grades 3 and 6 are reported in Appendix Table A6.

Appendix Tables A7-A10 present analogous heterogeneity estimates for all other observed grades. These results show patterns broadly consistent with those discussed above, with the largest and most persistent effects concentrated among students identified as Not On-Track in Grade 3 and comparatively limited heterogeneity across other demographic groups.

Heterogeneity in Effects by School Characteristics

Column (1) of Table 6 shows that the effects of retention threat on student third-grade ELA achievement increase with district spending. Relative to the Quintile 1 reference category, districts in Quintiles 4 and 5 experience an additional 0.109 SD to 0.120 SD increase in achievement, compared with 0.062 SD to 0.079 SD in Quintiles 2 and 3. The estimated effect in Quintile 1 is small and statistically insignificant. Column (1) of Table 7 shows a similar pattern for Grade 6. Relative to Quintile 1, estimated effects are largest in Quintiles 4 and 5, with increases in achievement of 0.067 SD and 0.089 SD, respectively, while estimates in Quintiles 2 and 3 are smaller and not statistically distinguishable from zero. These results suggest that higher-capacity districts may be better positioned to respond effectively to retention threat.

Column (2) of Table 6 shows that retention threat improves Grade 3 ELA achievement by an additional 0.073 SD in Quintile 5 schools, where about 64 percent of students are predicted to face retention threat, relative to Quintile 1 where about 13 percent of students are predicted to face retention threat. Estimated effects are largest in Quintile 4, with an additional 0.096 SD increase in achievement above the 0.035 SD effect for students in Quintile 1 schools. Effects in Quintiles 2 and 3 are smaller and not statistically different from that for Quintile 1. Similar patterns, shown in Column (2) of Table 7, emerge in Grade 6, where schools serving larger shares of students under threat continue to exhibit larger estimated responses to retention threat, although differences across quintiles are less precisely estimated than in Grade 3.

Column (3) of Tables 6 and 7 presents models that allow the effect of retention threat to vary simultaneously by district spending quintile and by the predicted share of students under threat of retention within a school. For Grade 3, retention threat continues to show larger effects in schools with higher per pupil spending and higher shares of students at risk of retention. In

Grade 6, estimated effects remain largest in higher-spending districts and in schools serving the largest shares of students predicted to face retention threat.[28]

Assignment to Reading Interventions

Retention threat might affect student outcomes by influencing their assignment to additional reading interventions. Students identified as Not On-Track to meet the Grade 3 ELA promotion threshold must be assigned to one or more of the 23 approved program codes for improving early literacy (see Appendix Table A1 for a list of all program codes). Students who are not identified as "Not On-Track" by the fall diagnostic assessment may be assigned to these services at the discretion of their teacher or principal. Districts and community schools are required to report reading improvement interventions provided during the school year in the Education Management and Information System (EMIS) at three points throughout the school year. During our sample period, literacy intervention assignment and related program participation codes were only available in Grade 3, limiting analysis of these mechanisms in later grades.

We estimate our main empirical specification replacing our academic and nonacademic outcome measures with five measures of student participation in reading programs. $AnyInt_{ic}$ is set equal to 1 if a student receives any of the 23 RIMP programs and 0 otherwise. The other four measures indicate interventions by program type as follows: programs that assign extra time for reading instruction throughout the school day ($ExtraTime_{ic}$), small group, peer, or one-on-one tutoring ($SmallGroup_{ic}$), reading programs including explicit instruction in reading interventions ($ReadingInt_{ic}$), and writing interventions ($WritingInt_{ic}$).

[28] Appendix Tables A11–A14 report corresponding estimates for the remaining observed grades. Across grades, the results suggest stronger responses to retention threat in higher-spending districts and, especially in the earlier grades, in schools serving relatively larger shares of students predicted to face retention threat.

Results in Figure 4 show increased assignment to literacy interventions for students facing retention threat. Students under threat of retention are 6.5 percentage points more likely to be assigned to any literacy intervention than students who met the promotion threshold in the fall. Relative to the control group mean of 0.485, this represents approximately a 13 percent increase in intervention assignment.[29] Students facing retention threat are also 5.1 percentage points more likely to receive additional instructional time for literacy support, 3.6 percentage points more likely to be assigned small-group, peer-learning, or one-on-one instructional supports and 1.8 percent more like to be assigned to interventions focused on vocabulary, fluency, comprehension, or sight-word reading. All of these estimates are statistically significant at conventional levels. In contrast, assignment to writing interventions is small and not statistically distinguishable from zero. Together, these results suggest that students facing retention threat are more likely to receive targeted literacy supports that may contribute to subsequent gains in ELA achievement. Although these effects are not small, it is unclear whether they are large enough to explain a significant portion of the effect of retention threat.[30]

Teachers and school officials' ability to assign interventions to support students under retention threat may depend on the level of resources and need in a school. Table 9 shows how the effect of retention threat on the probability of assignment to reading or writing interventions varies across school contexts. Panel A uses quintiles to categorize schools based on their

[29] Appendix Table A15 shows these estimates, and the control group mean for each outcome.

[30] Assignment to various interventions is likely influenced by student characteristics, and the research design in this study does not allow us to estimate of the effect of being assigned to various interventions and determine what portion of the estimated retention threat effects might be due to those interventions. If assignment to an intervention has an average effect on Grade 3 ELA scores of 0.25 standard deviations, then the results in Table 9 would imply that the increased probability of being assigned to intervention accounts for a 0.25 * .065 = 0.01625 increase in ELA achievement, which is just over 20 percent of the effect estimated in Table 3. The average effect of these interventions would need to be 0.077/0.065 = 1.18 standard deviations to account for all of the estimate effect of retention threat on Grade 3 ELA scores reported in Table 3.

district's spending level. Panel B uses quintiles to categorize schools based on the proportion of first-time Grade 3 students facing the threat of retention. The effect of retention threat on a student's probability of program assignment is estimated within each quintile. Column (1) shows limited evidence that the effect of retention threat on intervention assignment differs systematically across district spending quintiles.

Column (2) shows that the effect of retention threat on intervention assignment differs more clearly across schools serving different predicted shares of students at risk of retention. Schools that serve the smallest predicted shares of students under threat increases the probability of intervention assignment by 12 percentage points. In contrast, schools that serve the largest predicted shares of students under threat experience smaller increases in probability of intervention assignment, ranging from 4.6 to 10.9 percentage points below those in Quintile 1 schools. Overall, the effect of retention threat on intervention assignment tends to increase as the concentration of at-risk students in a school declines, suggesting that schools serving fewer at-risk students may be better able to provide individualized literacy supports.

The fact that the effect of retention threat on achievement is highest in schools with high shares of students at-risk of retention, but these schools do not show especially large effects of retention threat on assignment of students to individualized reading intervention suggests that individualized interventions are not the only mechanism through which the effects of retention threat are realized. Also, the fact that these schools are likely to be subject to the strongest school accountability pressures suggests that individual student retention policies and school-level accountability for achieving early literacy benchmarks are complementary measures for achieving the goals of early literacy initiatives.

CONCLUSION

This paper provides new evidence on the broad effects of early-grade retention mandates by examining how the threat of retention influences student learning prior to retention decisions. Exploiting Ohio's policy context, in which students are intentionally identified for retention risk and can avoid end-of-Grade 3 retention by demonstrating proficiency on a fall ELA assessment, we compare students just below the promotion threshold, who face retention risk, with students just above it, who do not. This setting allows us to observe how students and schools respond over the course of the "gate" grade and whether these responses translate into sustained improvements in later outcomes.

We find that retention threat increases achievement in both ELA and math during Grade 3, with ELA effects persisting through middle school, particularly among students identified as "Not On-Track." Students facing retention threat are also less likely to be absent and more likely to be assigned to literacy interventions, though we do not detect effects on disciplinary incidents or later-grade retention. The persistence of ELA gains, together with positive effects in math, a subject not tied to retention decisions, suggests that these responses reflect meaningful learning rather than short-term accountability responses alone.

Several implications follow from these findings. First, the results suggest that evaluations of retention mandates focused solely on retained students likely understate the policies' broader effects. A large literature on grade retention consistently finds short-run achievement gains for retained students that often fade over time (Jacob & Lefgren, 2004, 2009; Greene & Winters, 2007, 2009; Mariano et al., 2013; Schwerdt et al., 2017; Hwang & Koedel, 2023), but offers limited evidence on how retention mandates shape outcomes for students exposed to retention risk beyond those ultimately retained. Our findings show that retention policies influence more

than the 2–10% of students retained each year by changing student, family, and school behavior during the "gate" grade.

Second, the persistence of ELA gains, together with positive effects in math, a subject not tied to retention decisions, suggests that responses to retention threat reflect meaningful learning rather than short-term behavioral responses alone. This result is particularly notable given concerns that high-stakes environments may induce anxiety, disengagement, or strategic effort without improving underlying learning (Anderson et al., 2005; Heissel et al., 2021).

Third, our findings highlight the importance of resource capacity and policy design in shaping responses to retention threat. Stronger achievement effects in higher-spending districts, as well as larger effects of retention threat on assignment to individualized literacy interventions in schools serving fewer students at-risk of retention, suggest that resource availability might affect schools' ability to help students subject to retention threat. Yet achievement gains remain largest in higher-need settings indicating that broader accountability pressure may itself motivate school and student responses to retention risk. Ohio's combination of retention threat, district accountability, and financial incentives tied to student proficiency further underscores that the broader design of early literacy accountability systems likely matters for whether retention mandates improve student outcomes.

These findings are particularly relevant as states continue to adopt, revise, and repeal retention mandates as part of broader K–3 literacy reforms. Our results show that retention threat can improve student outcomes by changing student, family, or school behavior prior to retention decisions. More broadly, the findings suggest that the effectiveness of retention mandates may depend not only on whether students face retention risk, but also on how that risk is communicated and reinforced through broader early literacy systems, including accountability

incentives, intervention supports, and repeated signals to students, families, and schools during the "gate" grade.

Nonetheless, the effects documented here represent only one dimension of a broader debate on retention mandates. Although retention threat may generate broader benefits than previously recognized, retention in early grades may still increase inequality (LiCalsi et al., 2019; Greene & Winters, 2009) and impose costs on students and society (Winters, 2023; Manacorda, 2012; Eide & Showalter, 2001). Policy decisions should therefore weigh the broader benefits documented here alongside these potential costs.

References

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.

Anderson, K. P., Ritter, G. W., & Zamarro, G. (2019). Understanding a vicious cycle: The relationship between student discipline and student academic outcomes. Educational Researcher, 48(5), 251-262.

Arkansas General Assembly. (2023). *LEARNS Act*, S.B. 294, 94th Gen. Assemb., Reg. Sess. (Ark. 2023).

Arundel, K. (2024, March 26). *Half of K-2 students on track for reading, report finds*. K-12 Dive.

Battistin, E., & Schizzerotto, A. (2019). Threat of grade retention, remedial education and student achievement: evidence from upper secondary schools in Italy. *Empirical Economics*, *56*, 651-678.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.

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

Bloom H. S., Hill C. J., Black A. R., Lipsey M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, 1(4), 289–328.

Booher-Jennings, J. (2005). Below the bubble:"Educational triage" and the Texas accountability system. *American educational research journal*, *42*(2), 231-268.

Butts, J., & Roman, J. (2007). Changing Systems: Outcomes from the RWFJ Reclaiming Futures Initiative on Juvenile Justice and Substance Abuse.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust data-driven inference in the regression-discontinuity design. *The Stata Journal*, *14*(4), 909-946.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2015). Optimal data-driven regression discontinuity plots. *Journal of the American Statistical Association*, *110*(512), 1753-1769.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, *17*(2), 372-404.

Cattaneo, M. D., & Vazquez-Bare, G. (2017). The choice of neighborhood in regression discontinuity designs. *Observational Studies*, *3*(2), 134-146.

Chingos, M. M., & West, M. R. (2011). Promotion and reassignment in public school districts: How do schools respond to differences in teacher effectiveness? *Economics of Education Review*, 30(3), 419–433.

Cohen-Vogel, L. (2011). "Staffing to the Test": Are Today's School Personnel Practices Evidence Based? Educational Evaluation and Policy Analysis, 33(4), 483–505.

Colorado General Assembly. (2012). *READ Act*, H.B. 12-1238, 68th Gen. Assemb., 2nd Reg. Sess. (Colo. 2012).

Craig, S. G., Imberman, S. A., & Perdue, A. (2013). Does it pay to get an A? School resource allocations in response to accountability ratings. *Journal of Urban Economics*, *73*(1), 30-42.

De Angelis, J. (2016). What do citizens think about police accountability measures? Lessons from community attitudinal surveys. Criminal Justice Policy Review, 27(5), 520-536.

Downey, D. B., Von Hippel, P. T., & Broh, B. A. (2004). Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review*, 69(5), 613–635. https://doi.org/10.1177/000312240406900501

Education Commission of the States (ECS). 2023. *50-State Comparison: State K-3 Policies*.

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

Eren, O., Lovenheim, M. F., & Mocan, H. N. (2022). The effect of grade retention on adult crime: Evidence from a test-based promotion policy. *Journal of Labor Economics*, *40*(2), 361-395.

ExcelinEd. (2024). Comprehensive Early Literacy Policy: A State-by-State Analysis of Fundamental Principles. *ExcelinEd Policy Toolkit.*

Figlio, D. N., & Getzler, L. S. (2006). Accountability, ability and disability: Gaming the system?. In *Improving school accountability* (Vol. 14, pp. 35-49). Emerald Group Publishing Limited.

Figlio, D. N., Karbownik, K., & Özek, U. (2023). Sibling spillovers may enhance the efficacy of targeted school policies (No. w31406). National Bureau of Economic Research.

Fitzpatrick, M. D., Grissmer, D., & Hastedt, S. (2011). What a difference a day makes: Estimating daily learning gains during kindergarten and first grade using a natural experiment. *Economics of Education Review*, *30* (2), 269–279.

Fuller, S. C., & Ladd, H. F. (2013). School-Based Accountability and the Distribution of Teacher Quality Across Grades in Elementary School. *Education Finance and Policy*, *8*(4), 528–559.

Gelman, A., & Imbens, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. Journal of Business & Economic Statistics, 37(3), 447-456.

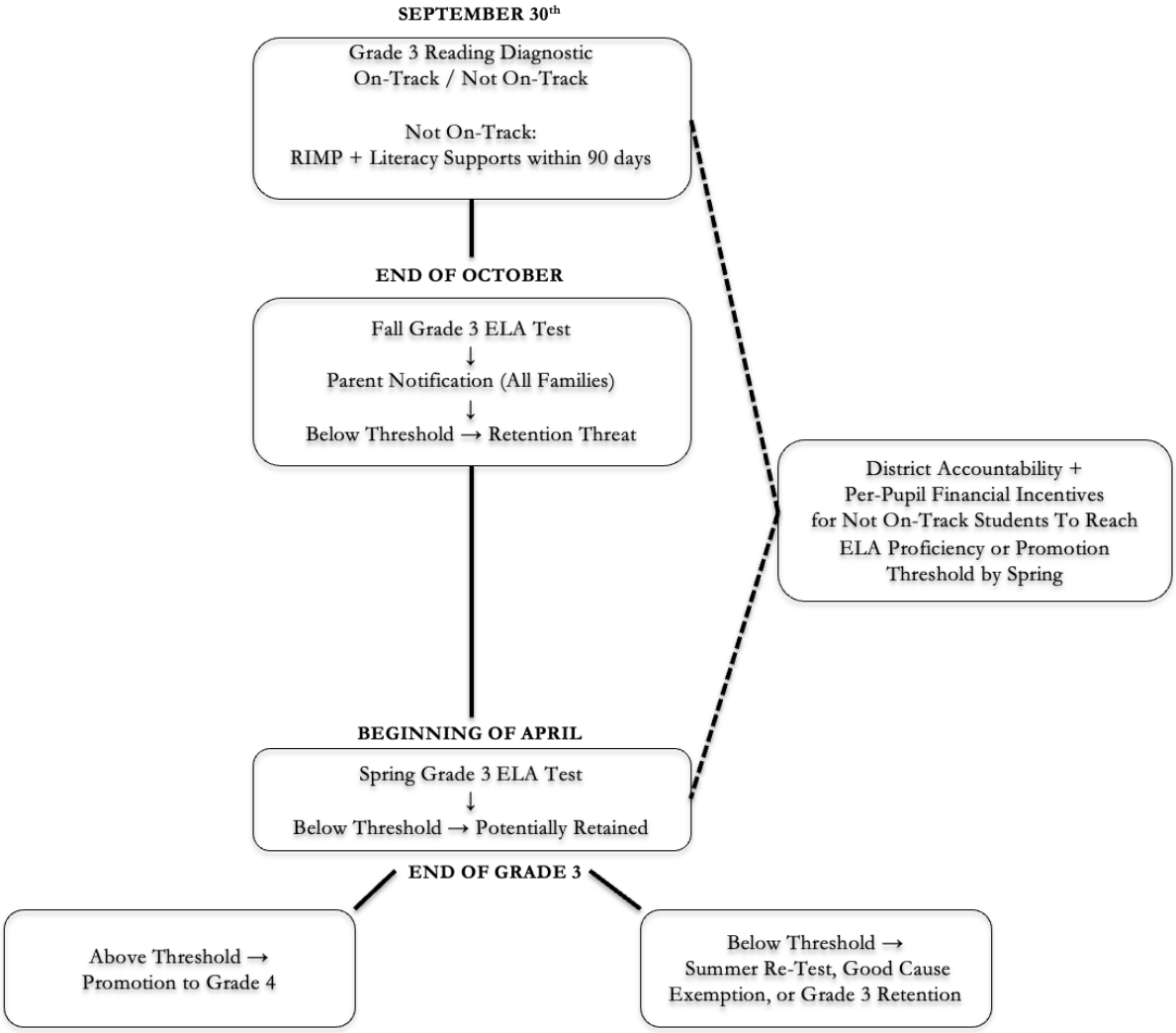
Gottfried, M. A. (2010). Evaluating the relationship between student attendance and achievement in urban elementary and middle schools: An instrumental variables approach. American Educational Research Journal, 47(2), 434-465.

Gottfried, M. A. (2011). The detrimental effects of missing school: Evidence from urban siblings. American Journal of Education, 117(2), 147-182.
Gottfried, M. A. (2013a). The spillover effects of grade-retained classmates: Evidence from urban elementary schools. American Journal of Education, 119(3), 405-444.
Gottfried, M. A. (2013b). Retained students and classmates' absences in urban schools. American Educational Research Journal, 50(6), 1392-1423.
Gottfried, M. A., & Kirksey, J. J. (2017). "When" students miss school: The role of timing of absenteeism on students' test performance. Educational Researcher, 46(3), 119-130.
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.
Greene, J. P., & 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.
Grissom, J. A., Kalogrides, D., & Loeb, S. (2017). Strategic Staffing? How Performance Pressures Affect the Distribution of Teachers Within Schools and Resulting Student Achievement. American Educational Research Journal, 54(6), 1079–1116.
Harris, F. (2012). Holding police accountability theory to account. Policing: A Journal of Policy and Practice, 6(3), 240-249.
Heckman, J. J., & Mosso, S. (2014). The economics of human development and social mobility. *Annu. Rev. Econ.*, *6*(1), 689-733.
Heissel, J. A., Adam, E. K., Doleac, J. L., Figlio, D. N., & Meer, J. (2021). Testing, stress, and performance: How students respond physiologically to high-stakes testing.
Hong, Y., & Hong, G. (2021). Schools with test-based promotion: Effects on instructional time allocation and student learning in grade 3. *AERA Open*, *7*, 2332858420979167.
Hwang, N., & Koedel, C. (2023). Helping or Hurting: The Effects of Retention in the Third Grade on Student Outcomes. *Educational Evaluation and Policy Analysis*, 01623737231197639.
Hwang, N., & Koedel, C. (2025). The Peer Effects of Grade Retention. EdWorkingPaper No. 25-1146. *Annenberg Institute for School Reform at Brown University*.
Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods, 15*(4), 309-334.
Imbens, G. W., and K. Kalyanaraman. 2009. *Optimal Bandwidth Choice for the Regression Discontinuity Estimator*. (Unpublished working paper).
Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools. *Journal of public Economics*, *89*(5-6), 761-796.
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.
Kolesár, M., & Rothe, C. (2018). Inference in regression discontinuity designs with a discrete running variable. *American Economic Review*, *108*(8), 2277-2304.
Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241-253.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of economic literature*, *48*(2), 281-355.
Louisiana State Legislature. (2023). *H.B. 12*, Reg. Sess. (La. 2023).
Manacorda, M. (2012). The cost of grade retention. *Review of Economics and Statistics*, *94*(2), 596-606.
Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Lichtenfeld, S. (2017). Long-term positive effects of repeating a year in school: Six-year longitudinal study of self-beliefs, anxiety, social relations, school grades, and test scores. *Journal of Educational Psychology*, *109*(3), 425.
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.
Mears, D. P., & Butts, J. A. (2008). Using performance monitoring to improve the accountability, operations, and effectiveness of juvenile justice. Criminal Justice Policy Review, 19(3), 264-284.
Mumma, K., & Winters, M. (2023). *The effect of retention under Mississippi's test-based promotion policy."* working paper 2023-1, Wheelock Educational Policy Center at Boston University, Winter 2023), https://doi. org/10.26300/hq2t-7x64.
Neuman, S. B., Quintero, E., Reist, K. (2023). READING REFORM ACROSS AMERICA: A Survey of State Legislation. Albert Shanker Institute.
Nevada Legislature. (2023). *Assembly Bill 400*, 82nd Leg. (Nev. 2023).
Noltemeyer, A. L., Ward, R. M., & Mcloughlin, C. (2015). Relationship between school suspension and student outcomes: A meta-analysis. School Psychology Review, 44(2), 224-240.
Ohio Department of Education. (2022). *Short Explanation of Improving At-Risk K-2 Readers.* Report Card Technical Document.
Papay, J. P., & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*, *130*, 105-119.
Quinn, D. M., Cooc, N., McIntyre, J., & Gomez, C. J. (2016). Seasonal dynamics of academic achievement inequality by socioeconomic status and race/ethnicity: Updating and extending past research with new national data. *Educational Researcher*, *45*(8), 443–453.
Reback, R. (2008). Teaching to the rating: School accountability and the distribution of student achievement. *Journal of public economics*, *92*(5-6), 1394-1415.
Redding, C., & Carlo, S. M. (2023). The Adoption of Test-Based Grade Retention Policies: An Event History Analysis. *Educational Policy*, 08959048231174886.
Riehl, E., & Welch, M. (2023). Accountability, test prep incentives, and the design of math and English exams. *Journal of Policy Analysis and Management*, *42*(1), 60-96.
Roderick, M., & Engel, M. (2001). The grasshopper and the ant: Motivational responses of low-achieving students to high-stakes testing. *Educational Evaluation and Policy Analysis*, *23*(3), 197-227.
Roderick, M., Jacob, B. A., & Bryk, A. S. (2002). The impact of high-stakes testing in Chicago on student achievement in promotional gate grades. *Educational evaluation and policy analysis*, *24*(4), 333-357.
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.

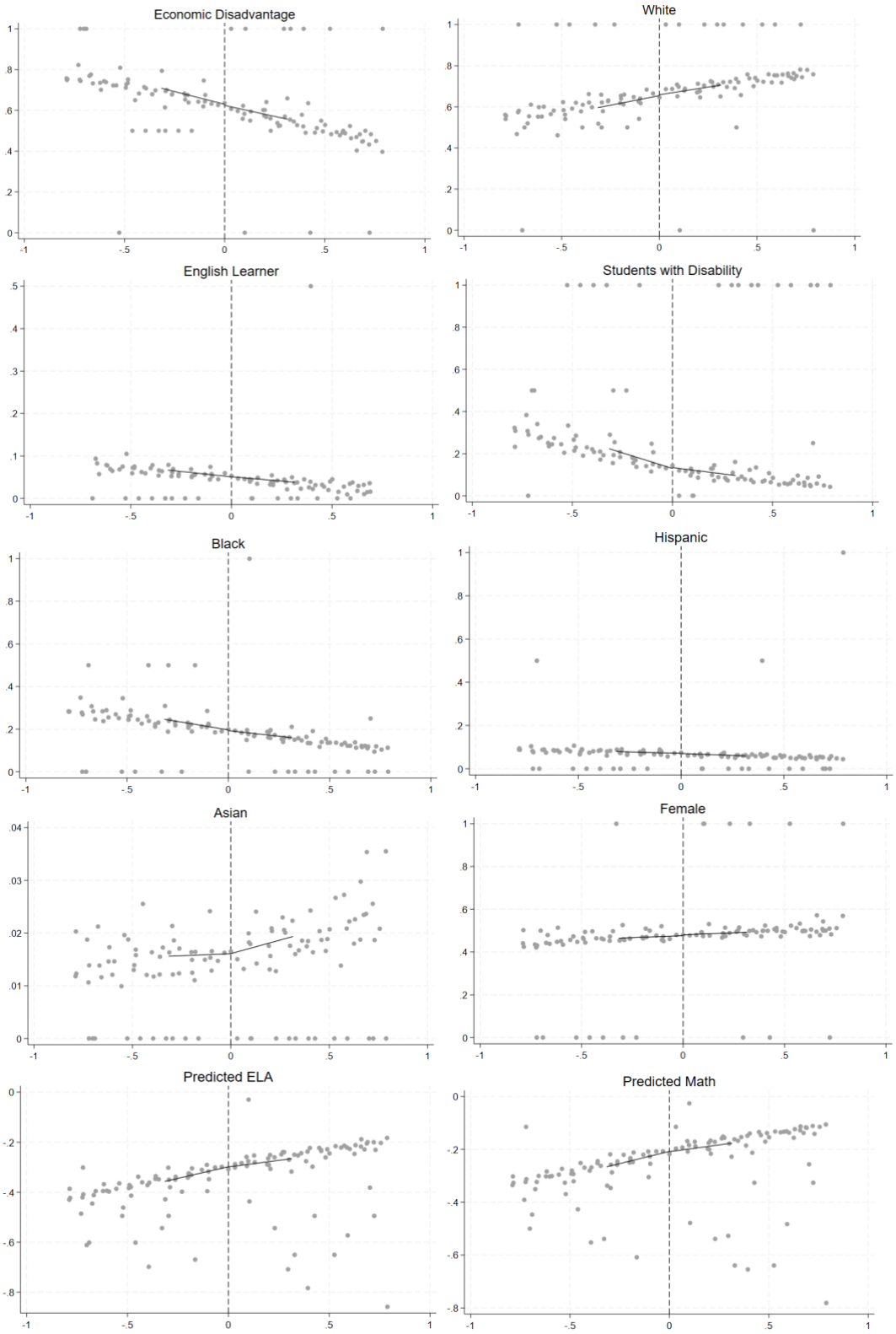
Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. Journal of personality and social psychology, 69(5), 797.
Walberg, H. J., & Tsai, S. L. (1983). Matthew effects in education. *American educational research Journal*, *20*(3), 359-373.
Wegner R. What to know as TCAPS start and Tennessee third, fourth graders face being held back. The Tennessean (March 2024).
West Virginia Legislature. (2023). *H.B. 3035*, Reg. Sess. (W. Va. 2023).

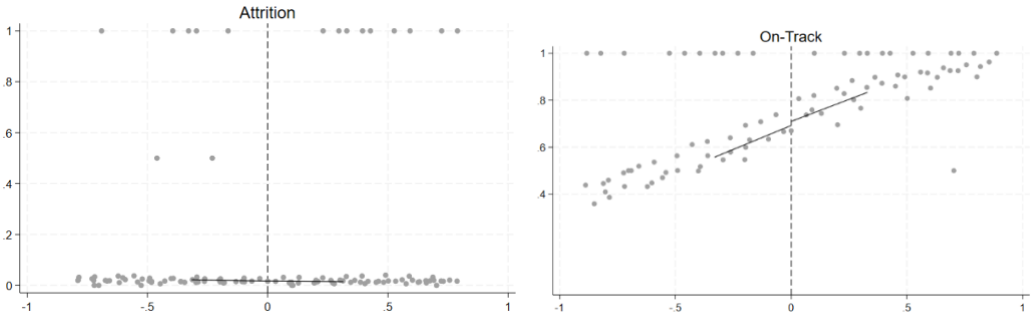
Figure 1. Student identification for accountability and promotion under Ohio's TGRG



Note: This figure illustrates Ohio's Grade 3 reading assessment, retention threat, and accountability timeline during the study period. By September 30, students complete a Grade 3 reading diagnostic and are classified as On-Track or Not On-Track. Students identified as Not On-Track are assigned a Reading Improvement and Monitoring Plan (RIMP) and literacy supports. Following the Fall Grade 3 ELA assessment administered in October, families are notified of student performance, and students scoring below the promotion threshold face the threat of Grade 3 retention. Students are reassessed on the Spring Grade 3 ELA assessment in April. Students scoring above the proficiency threshold are promoted to Grade 4, while students scoring below the threshold may be retained absent a summer retest or Good Cause Exemption. During the study period, districts also faced accountability and financial incentives tied to the spring proficiency or promotion outcomes of students identified as Not On-Track. Additional detail on Ohio's Improving K–3 Literacy accountability metric is provided in Appendix A2.

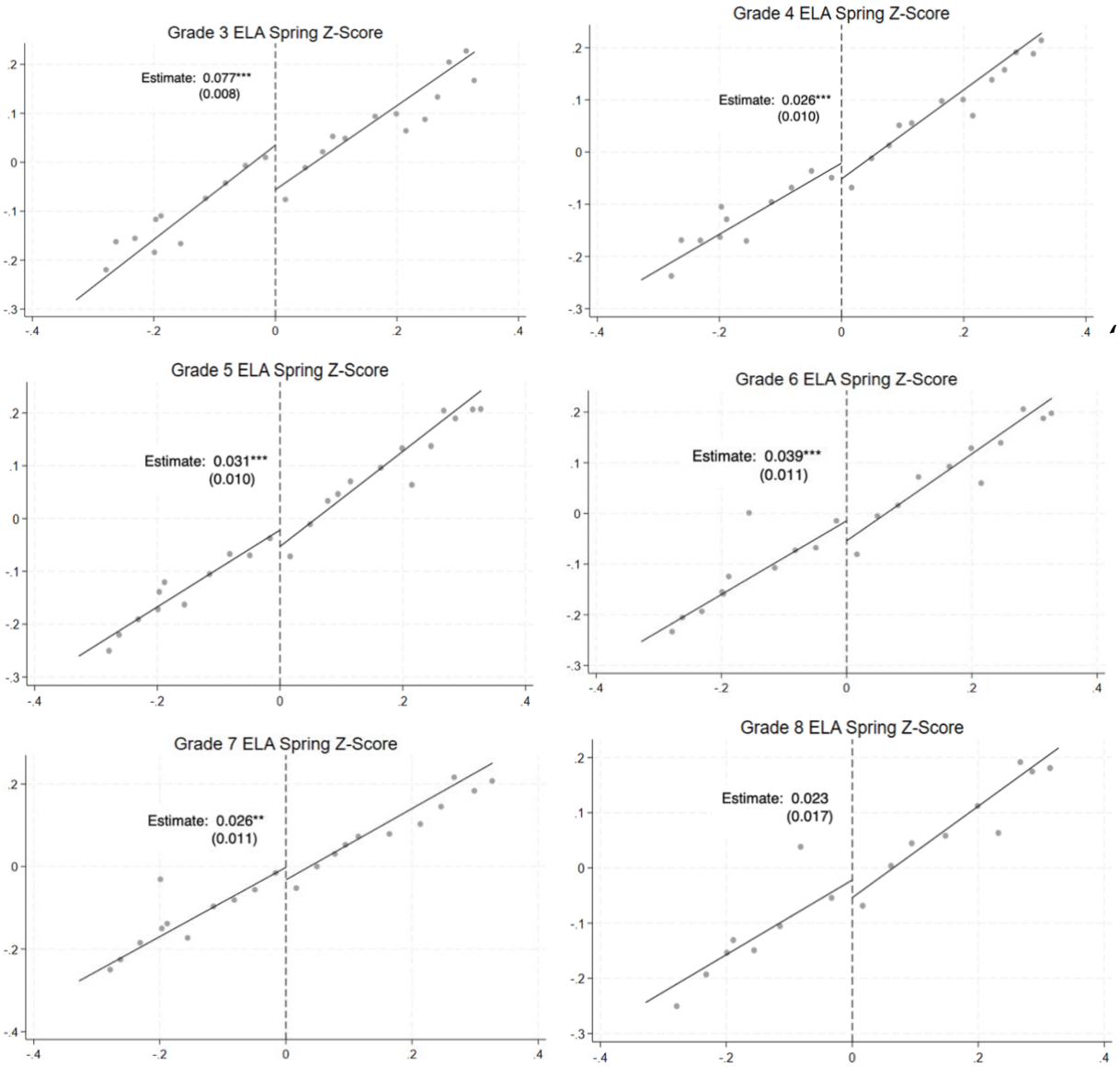
Figure 2. RD validity: continuity of student characteristics at cutoff.





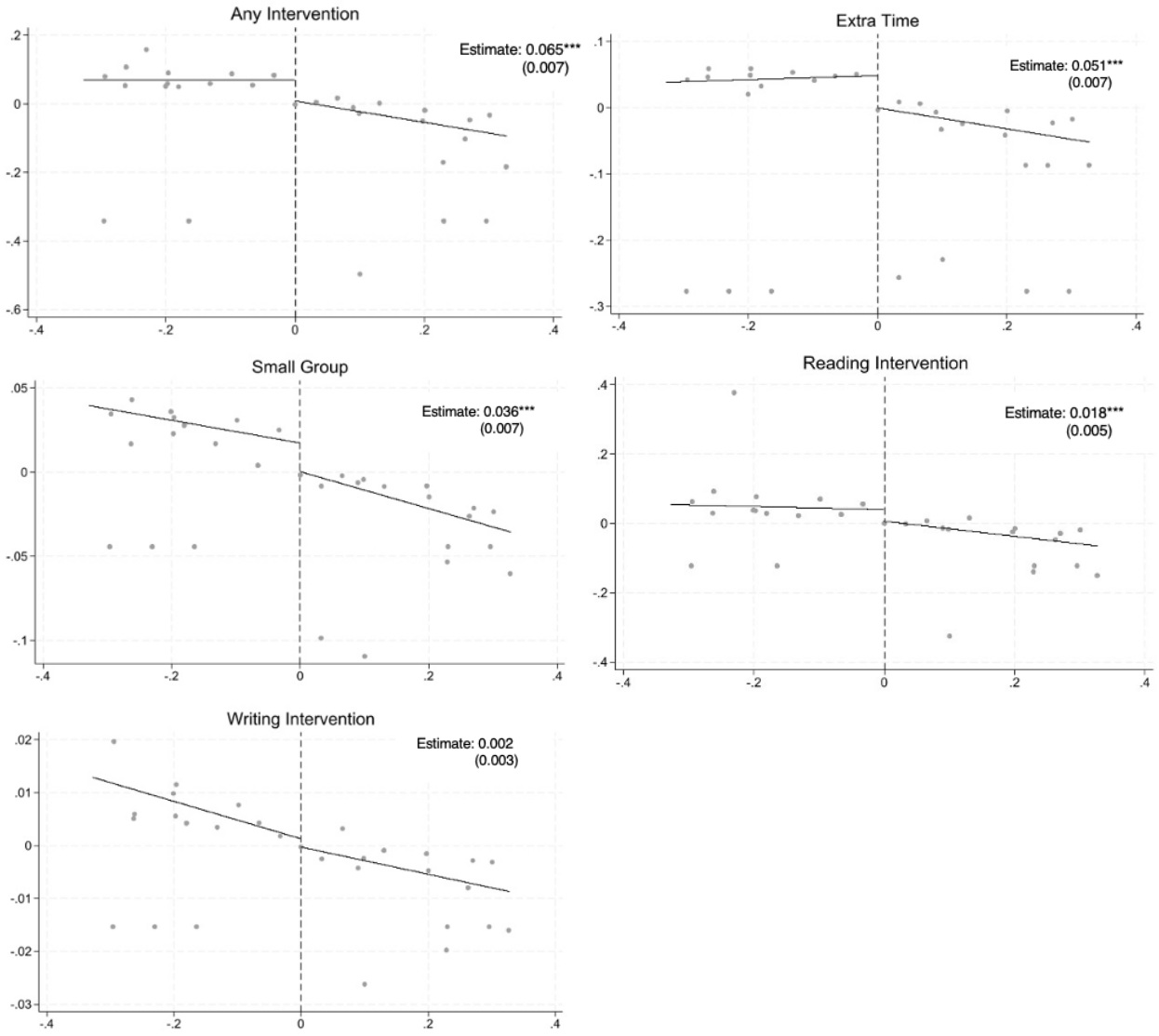
Note: The gray dots represent the mean outcome within quantile-spaced bins, bins that contain an equal number of observations, of the running variable, and the fitted relationship corresponds to a regression using the variable in the figure heading as the dependent variable which include controls for year and a student's On-Track status in Grade 3. Effects are estimated within the main ELA bandwidth of 0.328 SD. Bandwidth selected using Calonico et al. (2014) rdbwselect. The x-axis in each graph represents the running variable centered on the retention threat threshold in each year for each test.

Figure 3. Estimated effect of retention threat on ELA spring test score outcomes.



Notes: The gray dots represent the mean outcome within quantile-spaced bins, bins that contain an equal number of observations, of the running variable, and the fitted relationship corresponds to a regression using the variable in the figure heading as the dependent variable which include controls for year and a student's On-Track status in Grade 3. Effects are estimated within the main ELA bandwidth of 0.328 SD. Bandwidth selected using Calonico et al. (2014) rdbwselect. The x-axis in each graph represents the running variable centered on the retention threat threshold in each year for each test. Effect estimates and robust standard errors clustered on the running variable are shown in the text box, with standard errors reported in parentheses. *p < 0.10; **p < 0.05; ***p < 0.01.

Figure 4. Estimated effect of retention threat on probability of assignment to reading interventions in Grade 3.



Notes: The gray dots represent the mean outcome within quantile-spaced bins of the running variable, where each bin contains an equal number of observations. The fitted lines correspond to separate regressions using the outcome in each figure heading as the dependent variable and include controls for cohort fixed effects and a student's On-Track status in Grade 3. Effects are estimated within the Grade 3 ELA bandwidth of 0.328 SD, calculated for a model that includes cohort fixed effects and the On-Track variable. Effect estimates and robust standard errors clustered on the running variable are shown in the text box, with standard errors reported in parentheses. *p < 0.10; **p < 0.05; ***p < 0.01.

Table 1. Descriptive statistics for study sample.

	(1) All First-Time Grade 3 Students	(2) Within Bandwidth
Panel A: Accountability		
Threat (Fall)	0.336	0.373
	(0.472)	(0.484)
Threat (Spring)	0.162	0.145
	(0.368)	(0.352)
Not On-Track – Fall	0.241	0.294
	(0.428)	(0.455)
Panel B: Student Characteristics		
Female	0.489	0.479
	(0.500)	(0.500)
Asian	0.022	0.016
	(0.147)	(0.125)
Black	0.164	0.187
	(0.370)	(0.390)
Hispanic	0.058	0.066
	(0.234)	(0.249)
White	0.702	0.673
	(0.457)	(0.469)
Free- or Reduced-Price Lunch	0.536	0.611
Eligible (FRPL)	(0.499)	(0.487)
English Language Learner (ELL)	0.046	0.130
	(0.209)	(0.336)
Student with Disability (SWD)	0.133	0.052
	(0.340)	(0.222)
Panel C: Grade 3 Outcomes		
Spring ELA	0.020	-0.251
	(0.995)	(0.646)
Spring Math	0.021	-0.168
	(0.966)	(0.796)
Disciplinary Instances	0.042	0.045
	(0.201)	(0.207)
Absence Ratio	0.044	0.046
	(0.058)	(0.061)
Retained	0.032	0.017
	(0.175)	(0.129)
No. of Students	503,783	118,894

Note: Mean values are shown with standard deviations (SD) in parentheses. The sample is limited to first time Grade 3 students with fall test scores. Spring ELA test scores are observed for 486,366 students. Spring Math test scores observed for 465,527 students. Students observed in state supported schools (schools for the deaf and blind), joint vocational schools, and education service center districts are excluded (138 students). N in the second column is shown for students with spring ELA test scores.

Table 2. Continuity of student characteristics at cutoff.

	(1) Female	(2) Asian	(3) Black	(4) Hispanic	(5) White	(6) FRPL	(7) ELL	(8) SWD	(10) Predicted ELA	(11) Predicted Math	(12) Attrition	(13) On-Track
Threat (Fall)	-0.000	0.001	-0.007	-0.000	0.005	-0.002	0.004	-0.005	0.002	0.003	-0.003	-0.016**
	(0.007)	(0.002)	(0.006)	(0.004)	(0.007)	(0.007)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.007)
N	123,082	123,082	123,082	123,082	123,082	123,082	123,082	123,082	123,082	123,082	123,082	123,082

Note: Coefficient estimates in each column are from a separate regression using the variable in the row heading as the dependent variable. All estimates are from models which include the running score and cohort fixed effects. Effects are estimated within the Grade 3 ELA bandwidth of 0.392 SD, calculated for a model that only includes cohort fixed effects. Bandwidth selected using Calonico et al. (2014) rdbwselect. Robust standard errors clustered on the running variable are shown in parentheses. *p< 0.10; **p< 0.05; ***p< 0.01.

Table 3. Estimated effect of retention threat on ELA outcomes.

		(1) Grade 3	(2) Grade 4	(3) Grade 5	(4) Grade 6	(5) Grade 7	(6) Grade 8
Panel A: Pooled Estimates							
	Threat (Fall)	0.077***	0.026***	0.031***	0.039***	0.026**	0.023
		(0.008)	(0.010)	(0.010)	(0.011)	(0.011)	(0.017)
	N	120,368	111,848	110,090	80,733	75,580	76,441
	Control Mean	-0.107	-0.184	-0.176	-0.217	-0.149	-0.167
Panel B: Estimates by Cohort							
2014	Threat (Fall)	0.056***	0.024	0.017	0.007	0.033	0.012
		(0.015)	(0.022)	(0.019)	(0.020)	(0.020)	(0.021)
	N	26,687	24,348	23,728	23,728	23,409	22,528
2015	Threat (Fall)	0.034**	0.002	-0.001	0.021	0.019	
		(0.013)	(0.016)	(0.016)	(0.016)	(0.018)	COVID
	N	29,664	27,804	27,509	27,000	26,091	
2016	Threat (Fall)	0.101***	0.090***	0.099***	0.107***		0.091***
		(0.015)	(0.016)	(0.017)	(0.017)	COVID	(0.018)
	N	33,550	31,256	30,716	29,731		27,677
2017	Threat (Fall)	0.078***	0.042***	0.055***		0.025	0.056***
		(0.012)	(0.015)	(0.016)	COVID	(0.017)	(0.018)
	N	30,467	28,440	27,731		25,428	25,684

Note: Coefficient estimates in each column are from a separate regression using the variable in the row heading as the dependent variable. All estimates are from models which include the running score, On-Track, and cohort fixed effects. Effects are estimated within the Grade 3 ELA bandwidth of 0.328 SD, calculated for a model that includes cohort fixed effects and the On-Track variable. Bandwidth selected using Calonico et al. (2014) rdbwselect. Robust standard errors clustered on the running variable are shown in parentheses. *p< 0.10; **p< 0.05; ***p< 0.01.

Table 4. Estimated effect of retention threat on Math and non-academic outcomes.

	(1) Grade 3	(2) Grade 4	(3) Grade 5	(4) Grade 6	(5) Grade 7	(6) Grade 8
Math						
Threat (Fall)	0.037***	0.018*	0.017*	0.024**	0.002	0.005
	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.013)
Bandwidth	0.368	0.368	0.368	0.368	0.368	0.368
N	128,915	124,105	121,851	85,893	86,918	87,519
Control Mean	-0.068	-0.146	-0.178	-0.215	-0.174	-0.185
Absence						
Threat (Fall)	-0.002***	-0.001	-0.002**	-0.002	-0.002	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Bandwidth	0.347	0.347	0.347	0.347	0.347	0.347
N	115,896	114,280	114,223	113,807	113,529	111,679
Control Mean	0.044	0.048	0.052	0.057	0.070	0.083
Discipline						
Threat (Fall)	-0.002	-0.003	0.003	0.000	-0.002	0.006*
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Bandwidth	0.617	0.617	0.617	0.617	0.617	0.617
N	220,199	213,678	209,491	206,082	203,230	199,549
Control Mean	0.033	0.047	0.063	0.089	0.090	0.091
Probability of Retention						
Threat (Fall)	0.007***	-0.001	0.000	-0.000	-0.001	-0.001
	(0.0016)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Bandwidth	0.465	0.465	0.465	0.465	0.465	0.465
N	167,572	163,201	163,025	157,393	155,171	152,486
Control Mean	0.008	0.008	0.008	0.008	0.007	0.007

Note: Coefficient estimates in each column are from a separate regression using the variable in the row heading as the dependent variable. All estimates are from models which include the running score, On-Track, and cohort fixed effects. Effects are estimated within each outcome-specific bandwidth, calculated for a model that includes cohort fixed effects and the On-Track variable. Bandwidth selected using Calonico et al. (2014) rdbwselect. Robust standard errors clustered on the running variable are shown in parentheses. *p< 0.10; **p< 0.05; ***p< 0.01.

Table 5. Heterogeneity in effect of retention threat on Grade 3 and Grade 6 ELA and math outcomes by student characteristics.

	(1) On Track	(2) Not On Track	(3) Female	(4) Male	(5) Asian	(6) Black	(7) Hispanic	(8) White	(9) FRPL	(10) Non-FRPL	(11) ELL	(12) Non-ELL	(13) SWD	(14) Non-SWD
Grade 3														
	0.063***	0.102***	0.067***	0.089***	0.143*	0.136***	0.153***	0.136***	0.143***	0.157***	0.072**	0.077***	0.074***	0.074***
ELA	(0.010)	(0.015)	(0.011)	(0.012)	(0.082)	(0.022)	(0.038)	(0.042)	(0.012)	(0.016)	(0.036)	(0.008)	(0.023)	(0.009)
P-Value of Test	0.028		0.177			0.909			0.509		0.892		0.993	
N	85,539	35,134	57,889	62,784	1,937	22,223	7,920	81,649	73,277	47,396	6,162	114,511	15,425	105,248
	0.033***	0.029*	0.021	0.050***	-0.000	0.166***	0.134***	0.050***	0.074***	0.094***	0.018	0.037***	-0.000	0.037***
Math	(0.012)	(0.018)	(0.014)	(0.014)	(0.083)	(0.021)	(0.044)	(0.044)	(0.014)	(0.024)	(0.035)	(0.011)	(0.031)	(0.011)
P-Value of Test	0.847		0.058			0.000			0.495		0.360		0.248	
N	89,540	39,666	61,831	67,375	1,956	23,508	8,324	88,061	77,817	51,389	6,360	122,846	16,720	112,486
Grade 6														
	0.016	0.081***	0.044***	0.033***	0.198*	0.136***	0.087*	0.135***	0.135***	0.129***	0.047	0.038***	0.064*	0.027**
ELA	(0.013)	(0.020)	(0.015)	(0.016)	(0.120)	(0.029)	(0.051)	(0.015)	(0.017)	(0.022)	(0.049)	(0.011)	(0.034)	(0.011)
P-Value of Test	0.008		0.616			0.771			0.843		0.853		0.303	
N	57,112	23,621	38,532	42,201	1,160	14,466	5,041	55,567	48,558	32,175	4,009	76,724	10,684	70,049
	0.017	0.032	0.039*	0.012	0.203	0.065***	0.087*	0.114***	0.098***	0.115***	-0.005	0.038***	0.030	0.017
Math	(0.013)	(0.020)	(0.015)	(0.016)	(0.132)	(0.025)	(0.052)	(0.015)	(0.016)	(0.022)	(0.049)	(0.011)	(0.032)	(0.012)
P-Value of Test	0.520		0.223			0.255			0.534		0.853		0.708	
N	61,150	24,743	41,018	44,875	1,230	15,379	5,397	58,280	51,516	34,377	4,243	81,650	11,313	74,580

Note: Coefficient estimates for each group: sex, race, FRPL, ELL, and SWD come from separate regressions where the variable in the row heading is the dependent variable. The coefficient on the dummy variable Threat or Threat plus student type is displayed to show the estimated effect of retention threat for students of all types within a group. All estimates displayed are from models which control for and interacts student type indicator variables with the running variable. The model also controls for year and a student's On-Track status in Grade 3. Effects are estimated within outcome-specific bandwidths determined in Grade 3 that include controls for cohort and On-Track. Robust standard errors clustered on the running variable are shown in parentheses. The p-value of test indicates the statistical significance of a test that student types within the same group are equivalent. The reported Ns correspond to each analytic sample; however, each pair of estimates within a group comes from the same regression. *p< 0.10; **p< 0.05; ***p< 0.01.

Table 6. Estimated effect of retention threat on Grade 3 ELA by school and district characteristics.

	(1) Level of District Financing	(2) Predicted Share Subject to Retention Threat^	(3) Level of District Financing Controlling for Retention Threat^
Panel A: District Finance	0.109***		0.086***
Quintile 5 # Threat	(0.029)		(0.029)
Quintile 4 # Threat	0.120***		0.111***
	(0.031)		(0.031)
Quintile 3 # Threat	0.079***		0.072**
	(0.030)		(0.029)
Quintile 2 # Threat	0.062**		0.063**
	(0.030)		(0.030)
Threat (Fall)	-0.007		-0.015
	(0.025)		(0.030)
Panel B: Share Subject to Threat			
Quintile 5 # Threat		0.073***	0.082***
		(0.028)	(0.028)
Quintile 4 # Threat		0.096***	0.095***
		(0.029)	(0.026)
Quintile 3 # Threat		-0.005	-0.001
		(0.027)	(0.027)
Quintile 2 # Threat		0.023	0.031
		(0.027)	(0.027)
Threat (Fall)		0.035*	-0.015
		(0.021)	(0.030)
P-value of test:			Finance; Risk
5 = 4	0.661	0.396	0.326 ; 0.611
5 = 3	0.196	0.002	0.577 ; 0.001
5 = 2	0.051	0.044	0.367 ; 0.052
5 = 1	0.000	0.008	0.003 ; 0.004
4 = 3	0.105	0.000	0.127 ; 0.000
4 = 2	0.026	0.005	0.065 ; 0.018
4 = 1	0.000	0.001	0.000 ; 0.001
3 = 2	0.484	0.242	0.707 ; 0.182
3 = 1	0.008	0.848	0.013 ; 0.984
2 = 1	0.041	0.390	0.033 ; 0.251
All	0.001	0.000	0.007 ; 0.000
5 = 4 + 3 + 2	0.012	0.460	0.007 ; 0.425
4 = 3 = 2	0.075	0.000	0.154 ; 0.001

Note: Coefficients in each column are from a separate regression using Grade 3 ELA as the dependent variable and our main RD model fully interacted with each quintile and the running variable in a model that includes On-Track and cohort fixed effects. The coefficients shown are the interaction of Threat and a school's quintile. P-values indicate t-tests of differences in outcomes across coefficients. *p< 0.10; **p< 0.05; ***p< 0.01. Robust standard errors are clustered at the school level.

Table 7. Estimated effect of retention threat on Grade 6 ELA by school and district characteristics.

	(1) Level of District Financing	(2) Predicted Share Subject to Retention Threat^	(3) Level of District Financing Controlling for Retention Threat^
Panel A: District Finance	0.067**		0.046
Quintile 5 # Threat	(0.034)		(0.035)
Quintile 4 # Threat	0.089**		0.078**
	(0.031)		(0.035)
Quintile 3 # Threat	0.052		0.048
	(0.036)		(0.036)
Quintile 2 # Threat	0.057		0.057
	(0.037)		(0.036)
Threat	-0.019		-0.042
	(0.027)		(0.035)
Panel B: Share Subject to Threat			
Quintile 5 # Threat		0.063*	0.072**
		(0.035)	(0.036)
Quintile 4 # Threat		0.101***	0.092***
		(0.035)	(0.036)
Quintile 3 # Threat		0.037	0.033
		(0.034)	(0.034)
Quintile 2 # Threat		0.042	0.040
		(0.034)	(0.033)
Threat		-0.015	-0.042
		(0.026)	(0.035)
P-value of test:			Finance; Risk
5 = 4	0.470	0.266	0.310 ; 0.552
5 = 3	0.620	0.436	0.968 ; 0.238
5 = 2	0.760	0.531	0.750 ; 0.338
5 = 1	0.053	0.073	0.182 ; 0.045
4 = 3	0.247	0.055	0.355 ; 0.078
4 = 2	0.338	0.076	0.532 ; 0.124
4 = 1	0.013	0.005	0.028 ; 0.011
3 = 2	0.869	0.870	0.782 ; 0.814
3 = 1	0.155	0.270	0.181 ; 0.334
2 = 1	0.124	0.205	0.118 ; 0.235
All	0.165	0.067	0.287 ; 0.105
5 = 4 + 3 + 2	0.069	0.086	0.055 ; 0.178
4 = 3 = 2	0.460	0.111	0.639 ; 0.175

Notes: Coefficients in each column are from a separate regression using Grade 6 ELA as the dependent variable and our main RD model fully interacted with each quintile and the running variable in a model that includes On-Track and cohort fixed effects. The coefficients shown are the interaction of Threat and a school's quintile. P-values indicate t-tests of differences in outcomes across coefficients. *p< 0.10; **p< 0.05; ***p< 0.01. Robust standard errors are clustered at the school level.

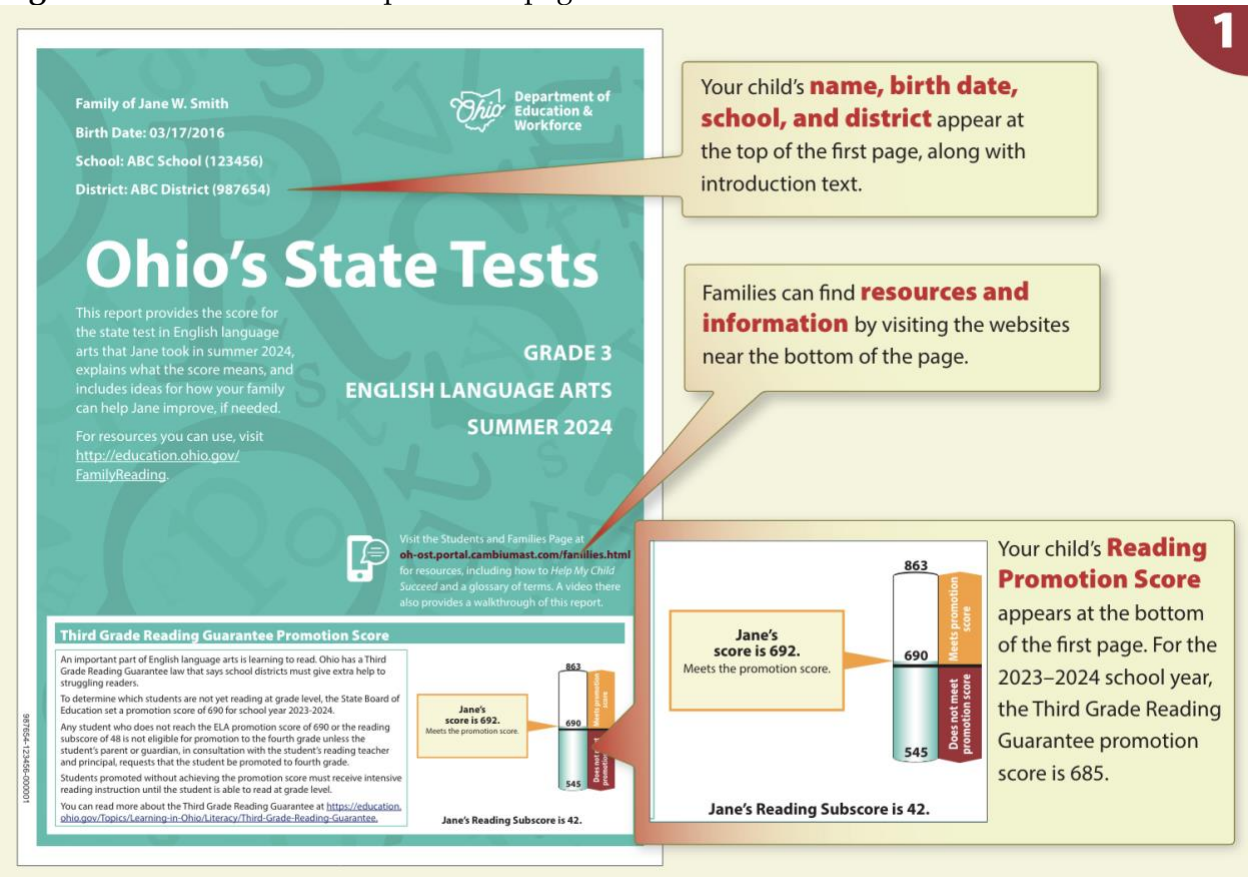
Table 8. Estimated effect of retention threat on the probability of assignment to reading interventions in Grade 3 by school and district characteristics.

	(1) Level of District Financing	(2) Predicted Share Subject to Retention Threat^	(3) Level of District Financing Controlling for Retention Threat^
Panel A: District Finance	-0.020		-0.009
Quintile 5 # Threat	(0.029)		(0.029)
Quintile 4 # Threat	-0.007		0.004
	(0.030)		(0.030)
Quintile 3 # Threat	0.028		0.029
	(0.031)		(0.031)
Quintile 2 # Threat	-0.060*		-0.061
	(0.032)		(0.032)
Threat	0.079***		0.043
	(0.025)		(0.032)
Panel B: Share Subject to Threat			
Quintile 5 # Threat		-0.058**	-0.044
		(0.027)	(0.028)
Quintile 4 # Threat		-0.109***	-0.099***
		(0.027)	(0.027)
Quintile 3 # Threat		-0.046*	-0.046*
		(0.026)	(0.026)
Quintile 2 # Threat		-0.039	-0.030
		(0.027)	(0.027)
Threat		0.120***	0.043
		(0.020)	(0.032)
P-value of test:			Finance; Risk
5 = 4	0.551	0.041	0.574 ; 0.033
5 = 3	0.0389	0.988	0.122 ; 0.951
5 = 2	0.105	0.469	0.043 ; 0.586
5 = 1	0.492	0.034	0.770 ; 0.110
4 = 3	0.170	0.036	0.331 ; 0.035
4 = 2	0.0445	0.005	0.013 ; 0.008
4 = 1	0.830	0.000	0.882 ; 0.000
3 = 2	0.00130	0.446	0.001 ; 0.513
3 = 1	0.371	0.028	0.349 ; 0.078
2 = 1	0.0629	0.142	0.058 ; 0.267
All	0.0266	0.001	0.020 ; 0.006
5 = 4 + 3 + 2	0.753	0.005	0.757 ; 0.015
4 = 3 = 2	0.00550	0.014	0.003 ; 0.021

Note: Coefficients in each column are from a separate regression using probability of assignment to any literacy intervention in Grade 3 as the dependent variable and our main RD model fully interacted with each quintile and the running variable in a model that includes On-Track and cohort fixed effects. The coefficients shown are the interaction of Threat and a school's quintile. P-values indicate t-tests of differences in outcomes across coefficients. *p< 0.10; **p< 0.05; ***p< 0.01. Robust standard errors are clustered at the school level.

APPENDIX A: ADDITIONAL FIGURES AND TABLES

Figure A1. Ohio test score report: front page.



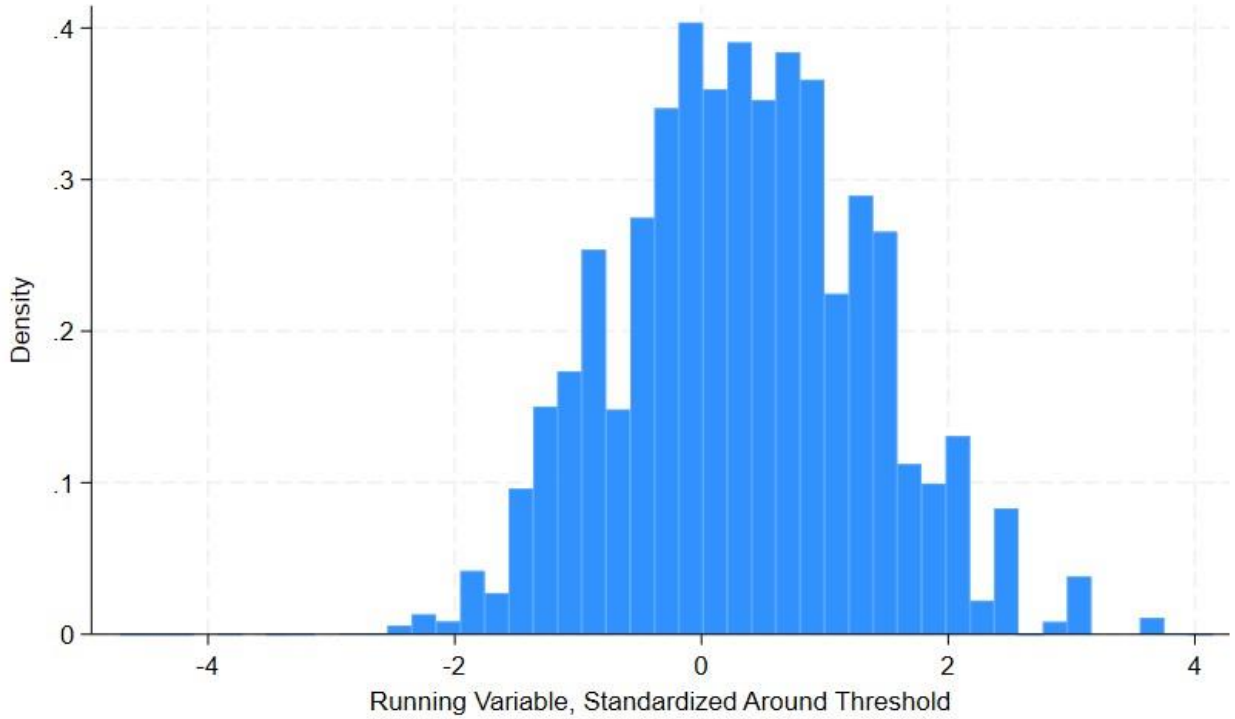
Notes: Figure displays the front page of the Ohio student test score report sent to families in the fall, spring, and summer, including student performance information and whether the student met the Grade 3 reading promotion threshold.

Figure A2. Calculating "Improving At-Risk K-2 Readers" measure.

Not On-Track in 2018-2019			Improving to On-Track In 2019-2020			
Kindergarten Reading Diagnostic Fall 2018-2019 school year	**91**	to	1st Grade Reading Diagnostic Fall 2019-2020 school year	**74**		Your district had 91 Kindergarten students who were not on-track on the fall reading diagnostic in the 2018-2019 school year. As first-graders, 74 of the 91 improved to on-track on the 2019-2020 fall reading diagnostic.
1st Grade Reading Diagnostic Fall 2018-2019 school year	**72**	to	2nd Grade Reading Diagnostic Fall 2019-2020 school year	**56**		Your district had 72 first-grade students who were not on-track on the fall reading diagnostic in the 2018-2019 school year. As second-graders, 56 of the 72 improved to on-track on the 2019-2020 fall reading diagnostic.
2nd Grade Reading Diagnostic Fall 2018-2019 school year	**62**	to	3rd Grade Reading Diagnostic Fall 2019-2020 school year	**48**		Your district had 62 second-graders who were not on-track on the fall reading diagnostic in the 2018-2019 school year. As third-graders, 48 of the 62 improved to on-track on the 2019-2020 fall reading diagnostic.
3rd Grade Reading Diagnostic Fall 2019-2020 school year	**79**	to	3rd Grade Reading ELA Fall or Spring 2019-2020 school year	**62**		Your district had 79 third-graders who were not on-track on the fall reading diagnostic in the 2019-2020 school year. Of the 79 third-graders who were not on-track at the beginning of third grade, 62 of them scored proficient (700) or higher on the Fall or Spring ELA assessment.
Totals before deduction	**304**			**240**		Your district had 178 students improve from not on-track in the 2018-2019 school year to on-track in the 2019-2020 school year and 62 students improve from not on-track on the 2019-2020 3rd grade diagnostic to proficient on the 3rd grade ELA test.
Deduction/304* for 3rd graders who did not pass 3rd grade ELA test and were not on a Reading Improvement and Monitoring Plan				**-10**		Your district received a deduction of ten (10) students because ten third grade students were not on a Reading Improvement and Monitoring Plan and did not score proficient on the third grade ELA assessment (700)
Totals after deductions	**304**			**230**		Your district's Improving At-Risk K-3 Readers Measure grade is 230 out of 304 or 75.7%, an "B" based on the grade scale shown to the left.

Source: Ohio Department of Education. (2022). *Short Explanation of Improving At-Risk K-2 Readers.* Report Card Technical Document.

Figure A3. Frequency distribution of the running variable.

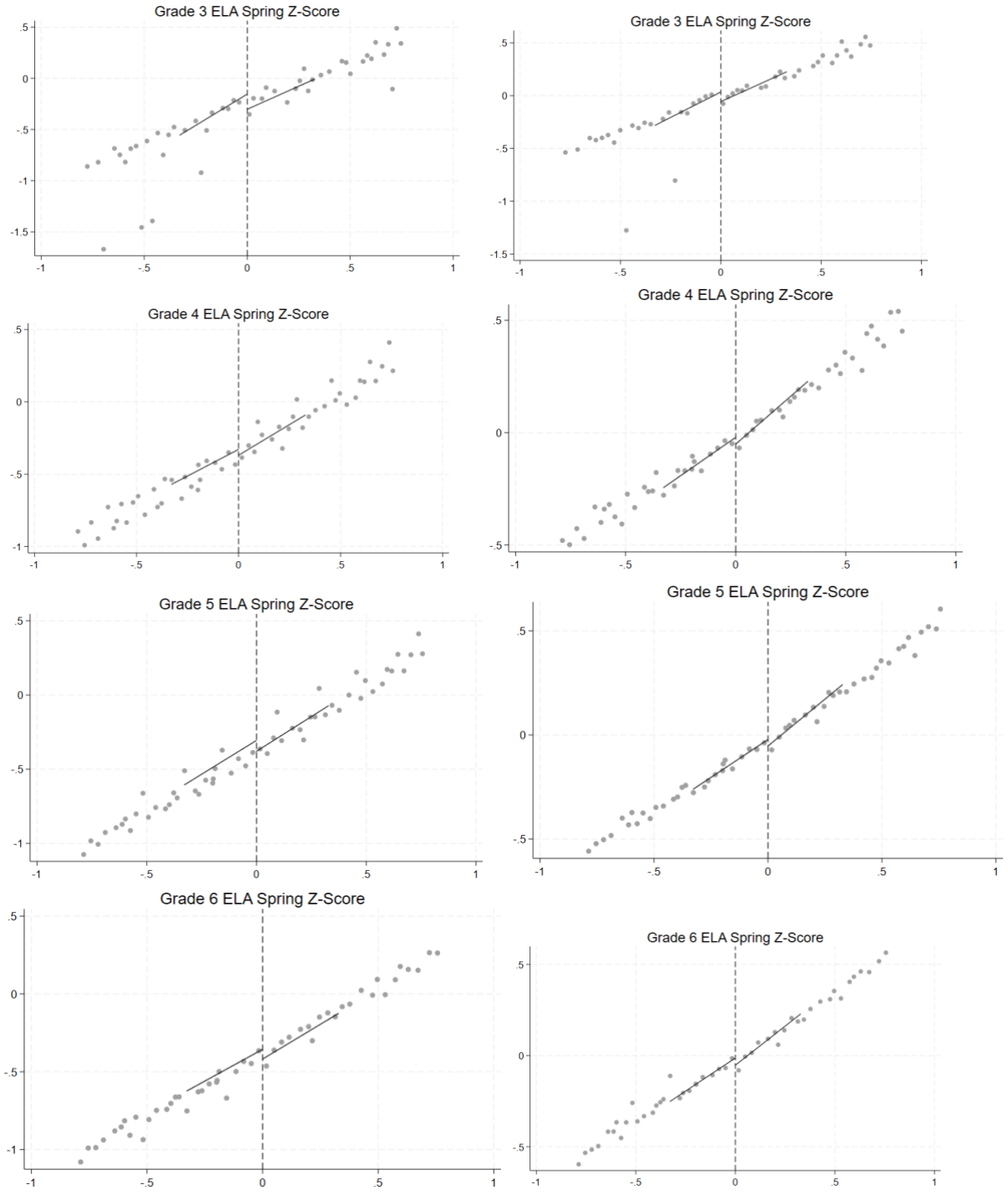


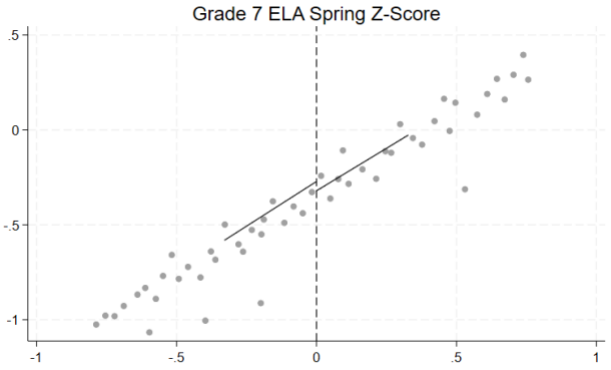
Notes: This figure displays the density of the standardized running variable centered at the promotion threshold. The running variable is normalized such that zero corresponds to the threshold determining eligibility for retention threat. Consistent with this visual evidence, a density test similar to McCrary (2008) using a local polynomial estimator fails to reject the null hypothesis of no manipulation around the cutoff ((t = 0.78), (p = 0.40)).

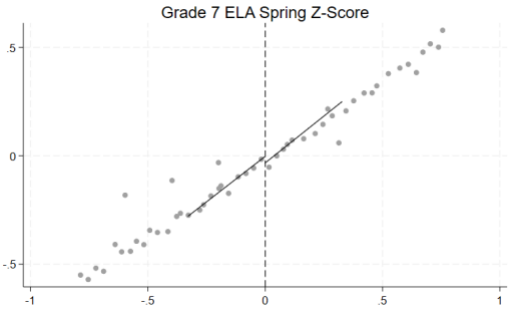
Figure A4. Robustness of ELA effects: Estimated effect of retention threat on ELA spring test score outcomes using alternative bins.

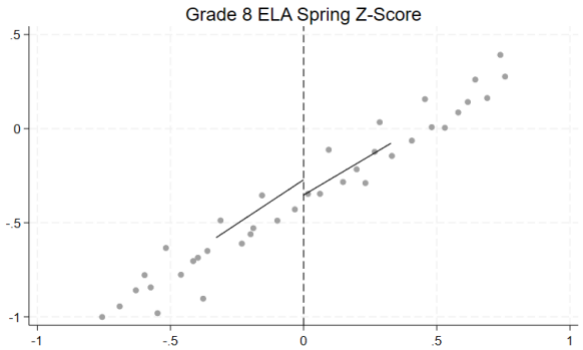
Panel A. Evenly Spaced Bins.

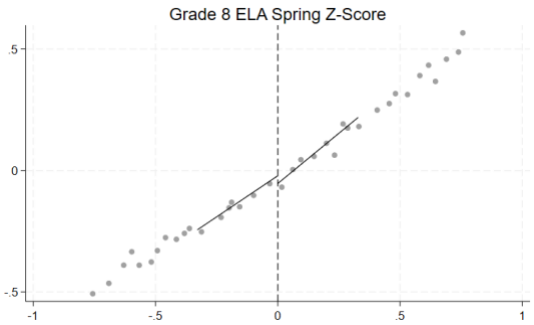
RAW RESIDUALIZED





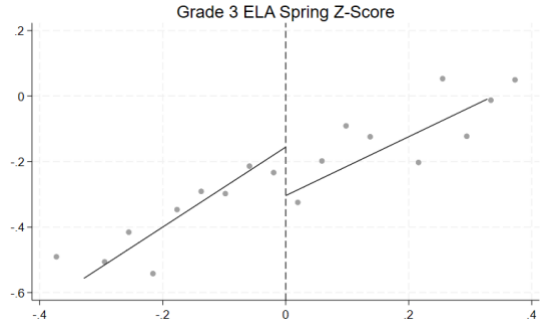


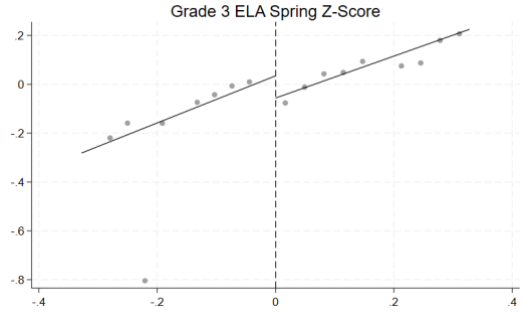


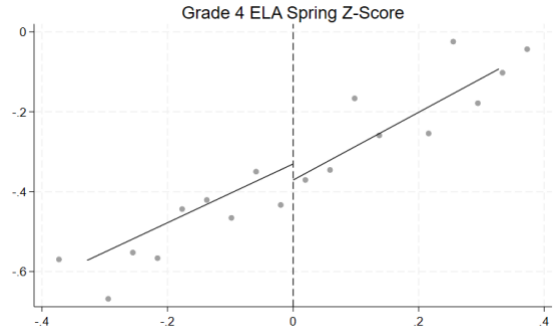


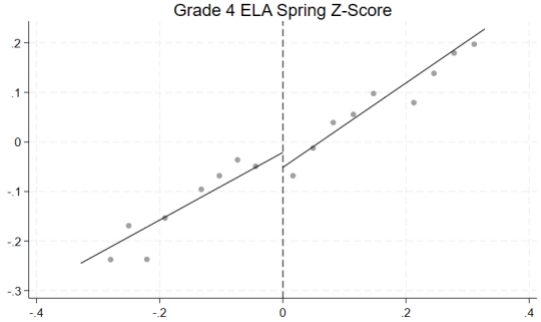
Panel B. 10 Bins per Side of the Threshold.

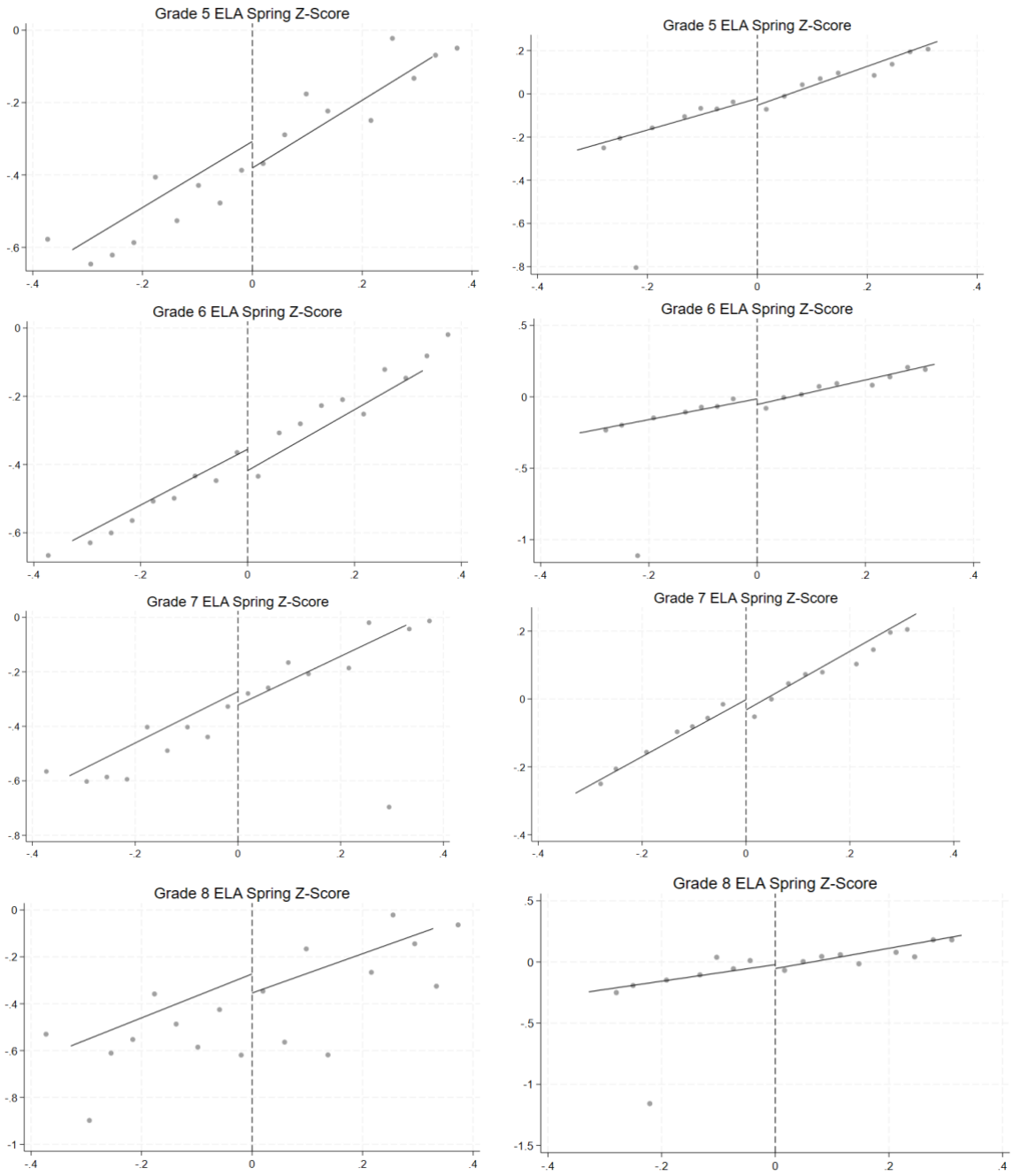
RAW RESIDUALIZED











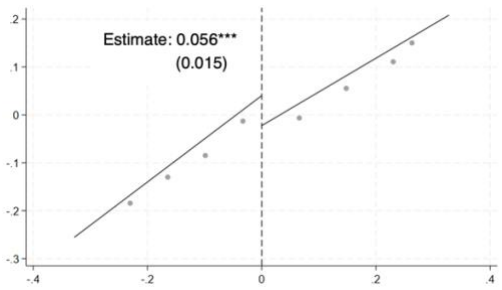
Notes: In both panels, the Raw column displays mean outcomes using unadjusted data, while the Residualized column displays mean outcomes after adjusting for year and students' Grade 3 On-Track status. In Panel A, gray dots represent the mean outcome within evenly spaced bins of the running variable. In Panel B, gray dots represent the mean outcome within 10 bins on each side of the retention threat threshold. Fitted lines correspond to regressions using the variable listed in the figure heading as the dependent variable and include controls for year and students' Grade 3 On-Track status. Effects are estimated within the main ELA bandwidth of 0.328 SD. Bandwidth selected using Calonico et al. (2014) rdbwselect. The x-axis in each graph represents the running variable centered on the retention threat threshold in each year for each test.

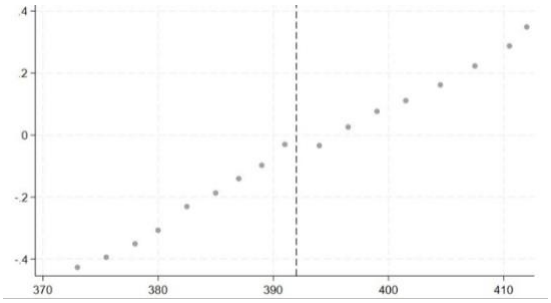
Figure A5. Estimated effect of retention threat on Grade 3 ELA spring test score outcomes by cohort, comparing the standardized vs. raw running variable.

Standardized Running Variable

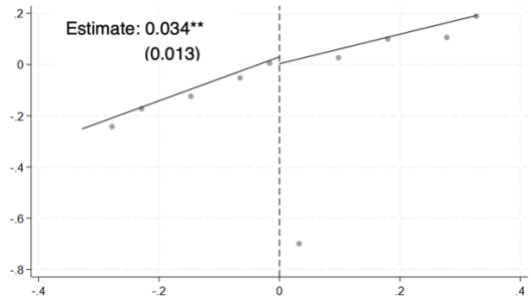
Raw Test Score Running Variable

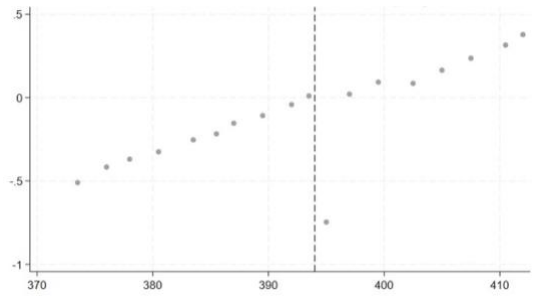
2014



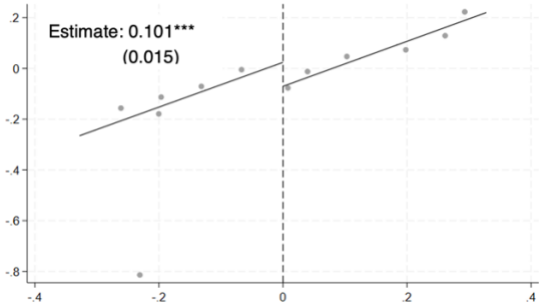


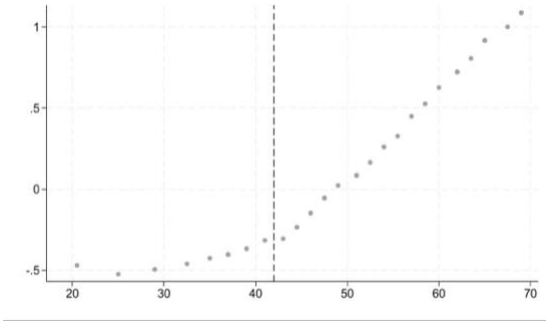
2015



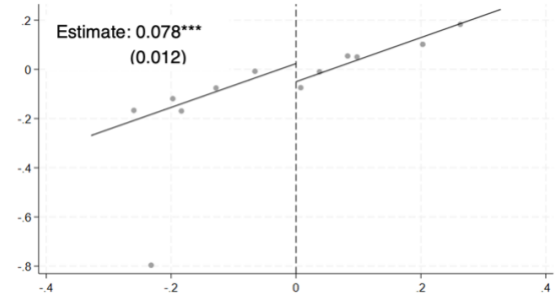


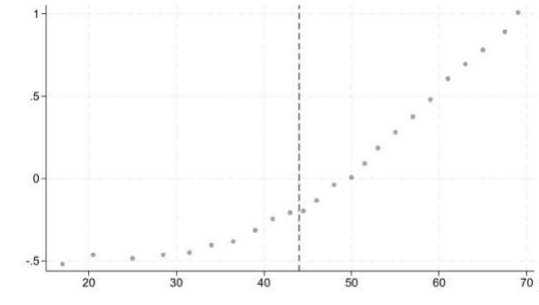
2016





2017



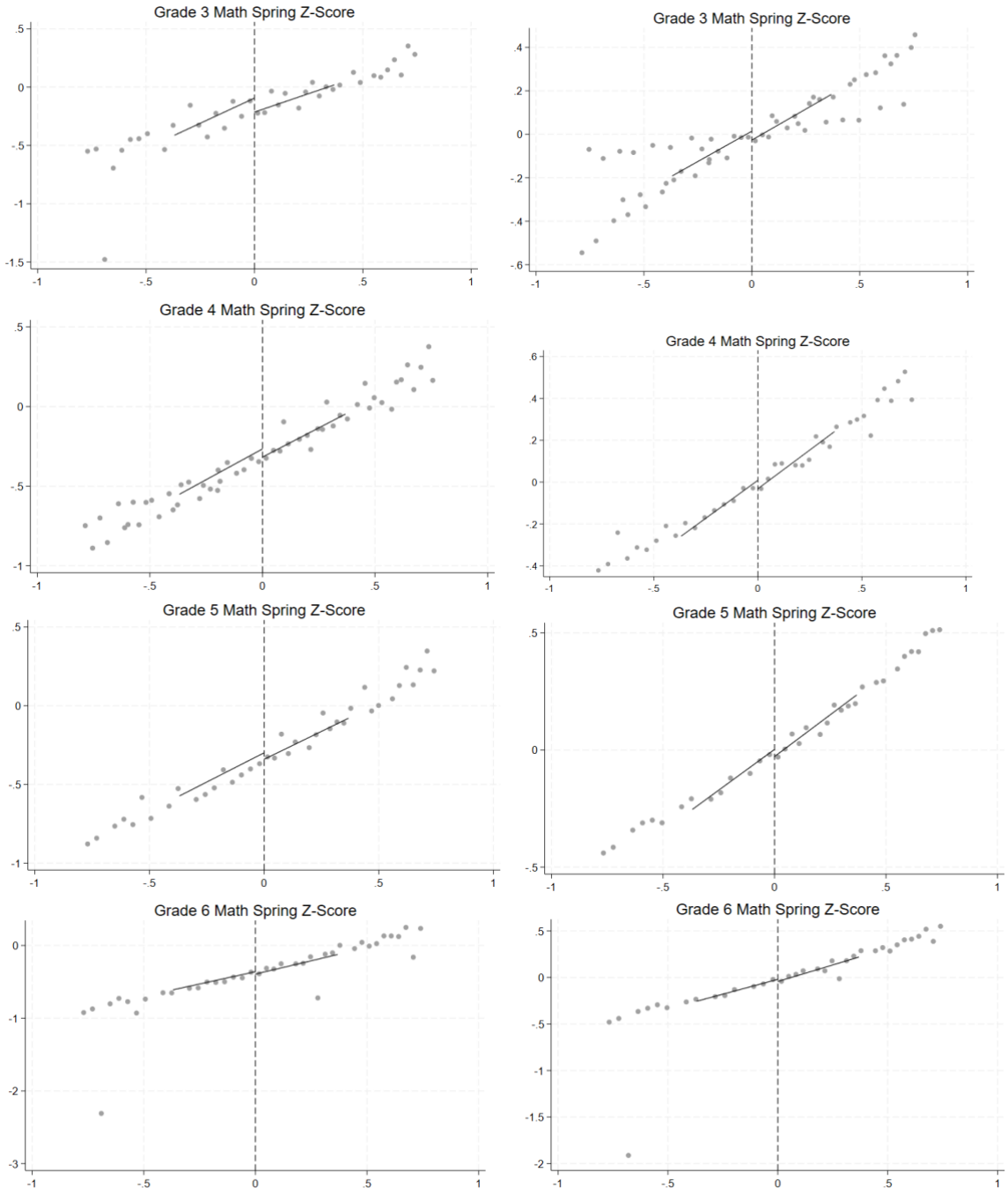


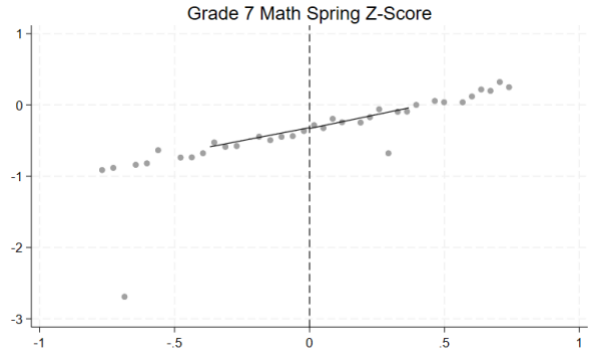
Notes: The figures display estimated effects of retention threat on Grade 3 ELA achievement separately by cohort. The left column presents estimates using the standardized running variable, centered at the promotion threshold and measured in standard deviations, while the right column presents estimates using the raw fall assessment score running variable. Gray dots represent mean outcomes within quantile-spaced bins of the running variable.

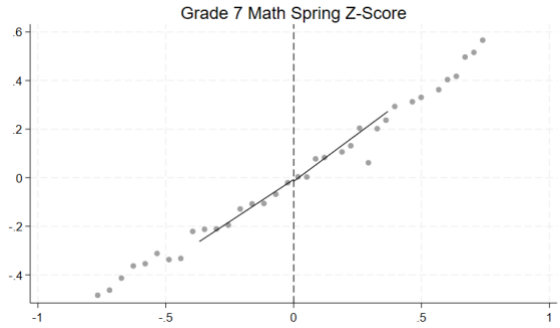
Figure A6. Robustness of Math effects: estimated effect of retention threat on Math spring test score outcomes using alternative bins.

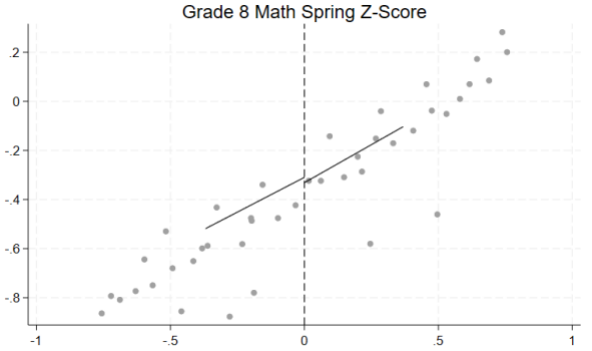
Panel A. Evenly Spaced Bins.

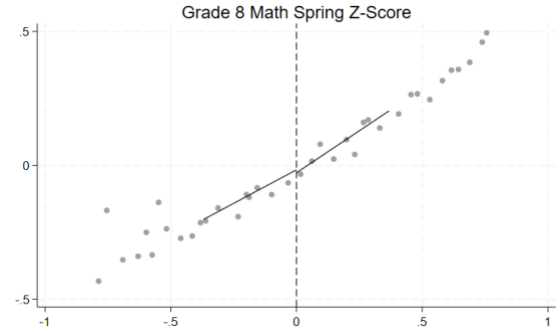
RAW RESIDUALIZED





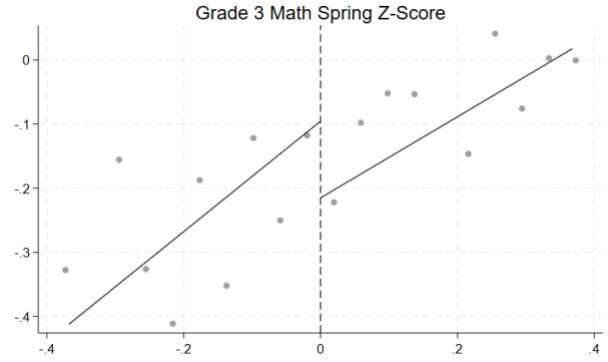


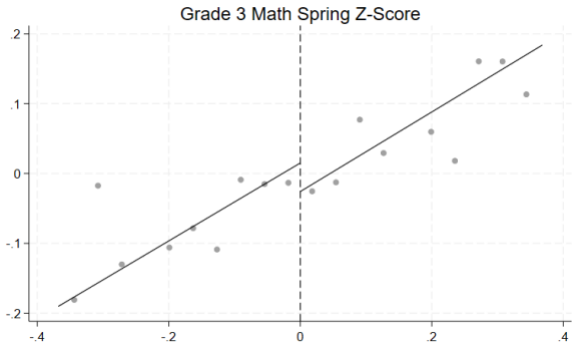


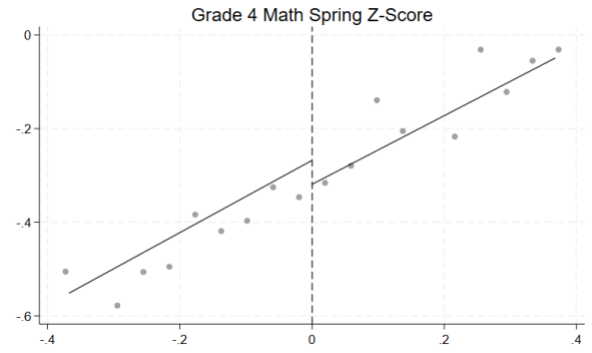


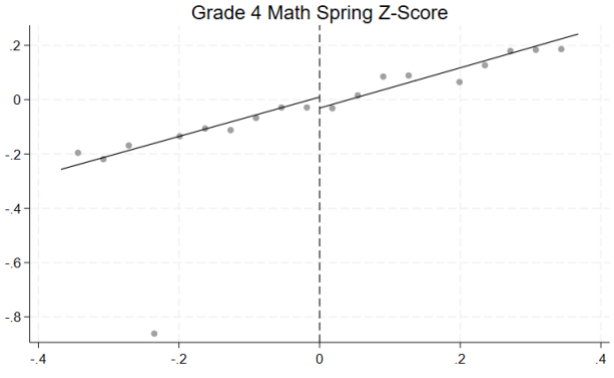
Panel B. 10 Bins per Side of the Threshold.

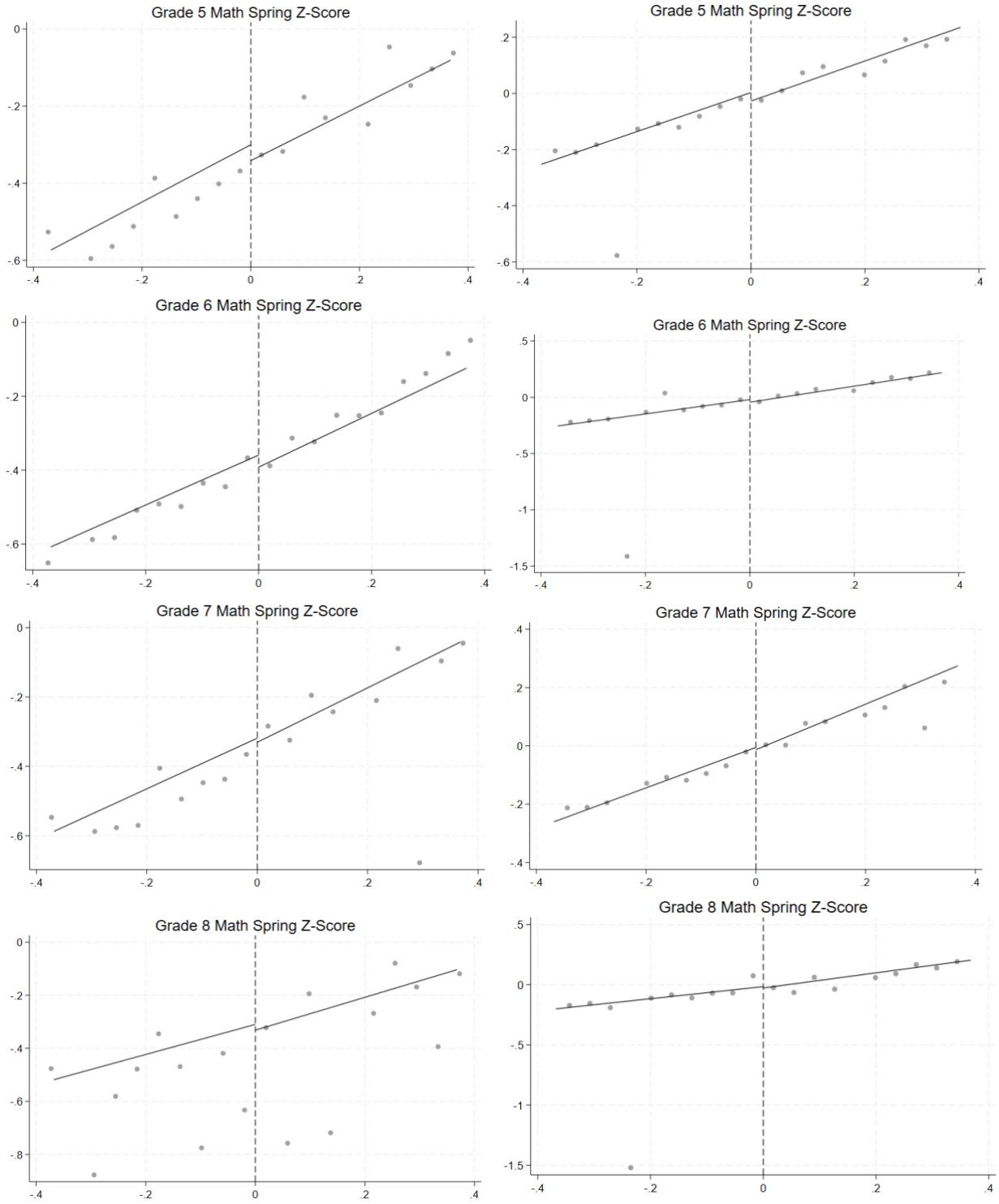
RAW RESIDUALIZED









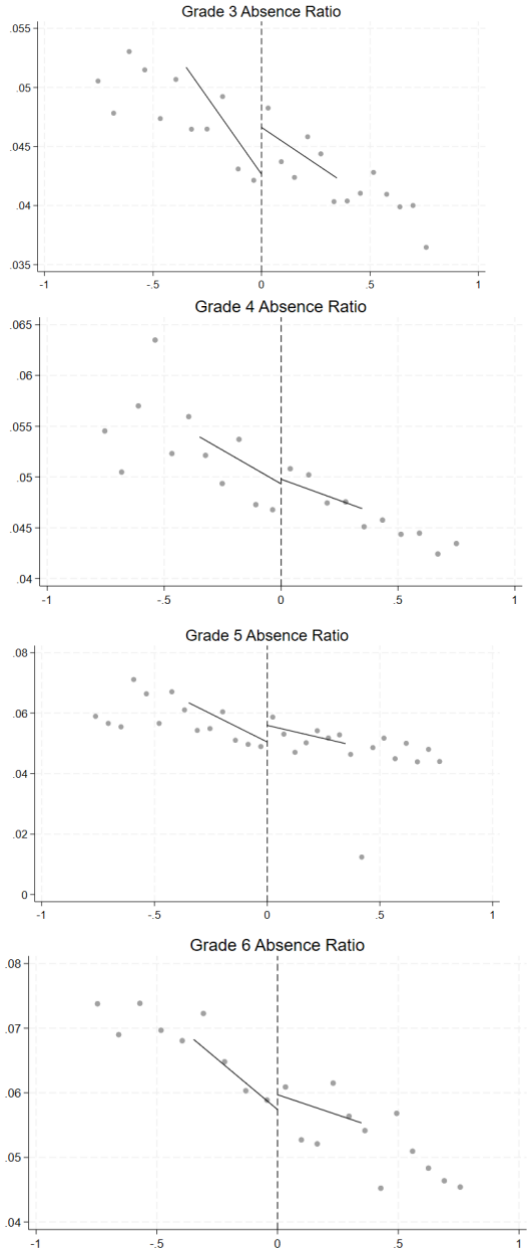


Notes: In both panels, the Raw column displays mean outcomes using unadjusted data, while the Residualized column displays mean outcomes after adjusting for year and students' Grade 3 On-Track status. In Panel A, gray dots represent the mean outcome within evenly spaced bins of the running variable. In Panel B, gray dots represent the mean outcome within 10 bins on each side of the retention threat threshold. Raw figures show observed outcomes; residualized figures remove variation from covariates. Fitted lines correspond to regressions using the variable listed in the heading as the dependent variable and control for cohort and students' Grade 3 On-Track status. Effects are estimated within the Math bandwidth of 0.368 SD. Bandwidth selected using Calonico et al. (2014) rdbwselect. The x-axis in each graph represents the running variable centered on the retention threat threshold in each year.

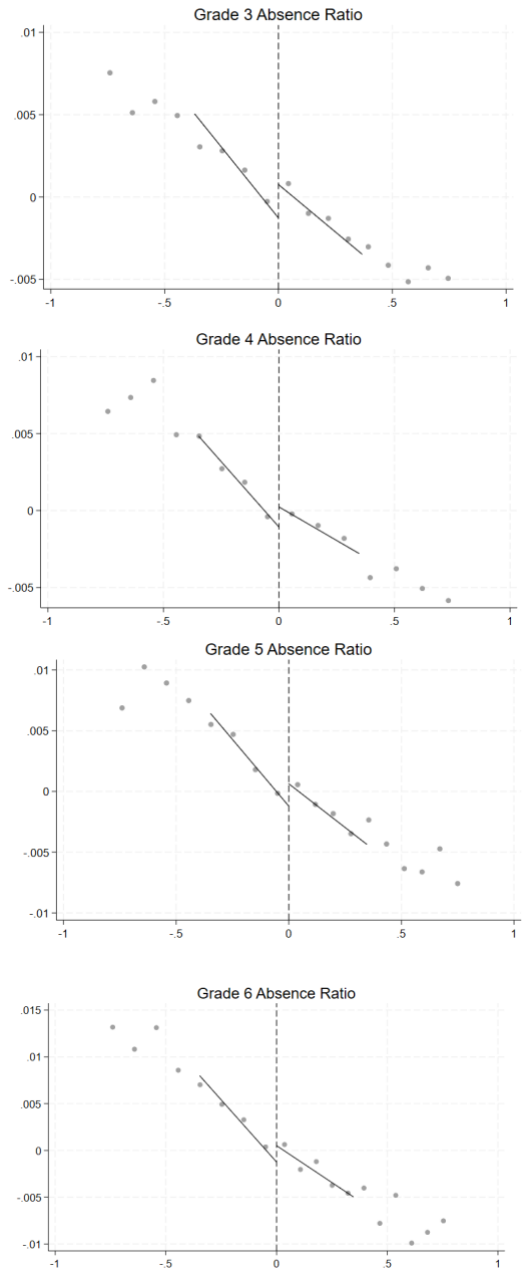
Figure A7. Robustness of absence effects: estimated effect of retention threat on absences using alternative bins.

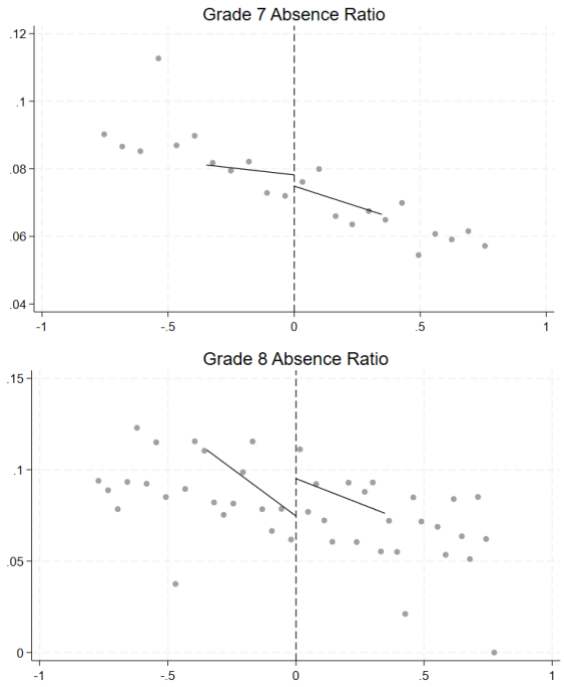
Panel A. Evenly Spaced Bins.

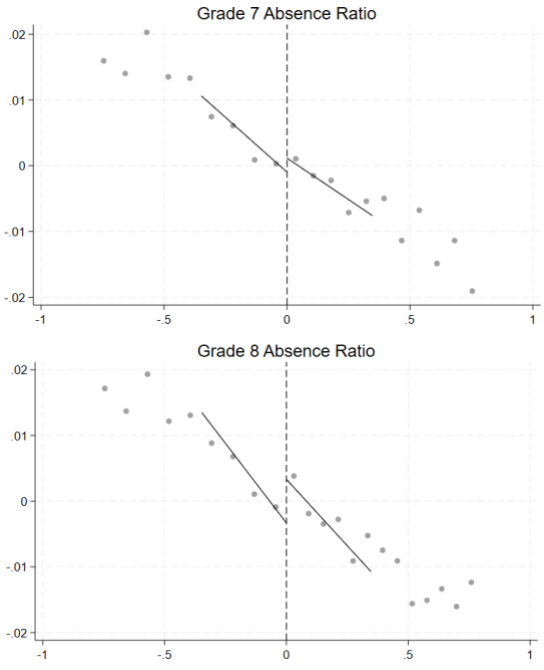
RAW



RESIDUALIZED

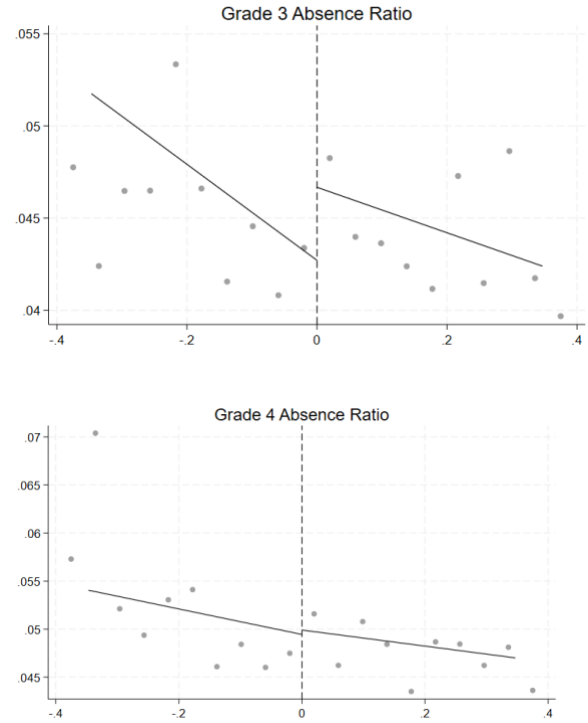




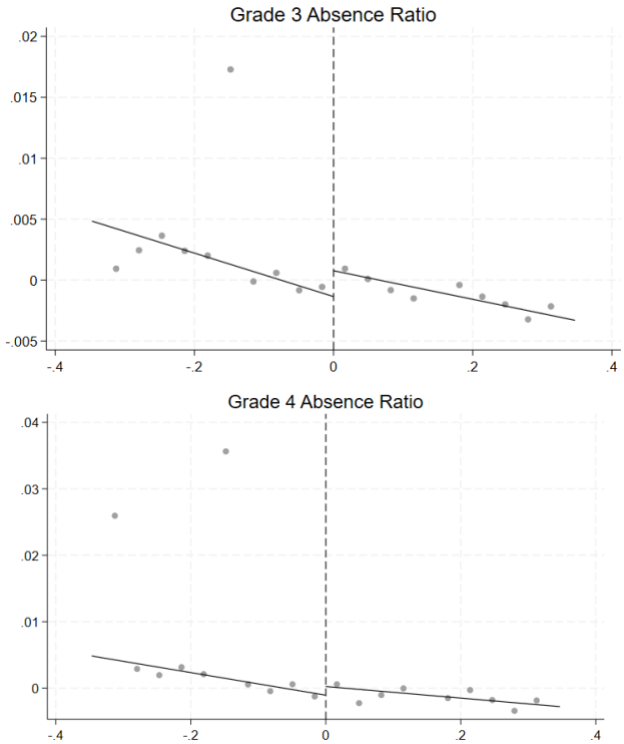


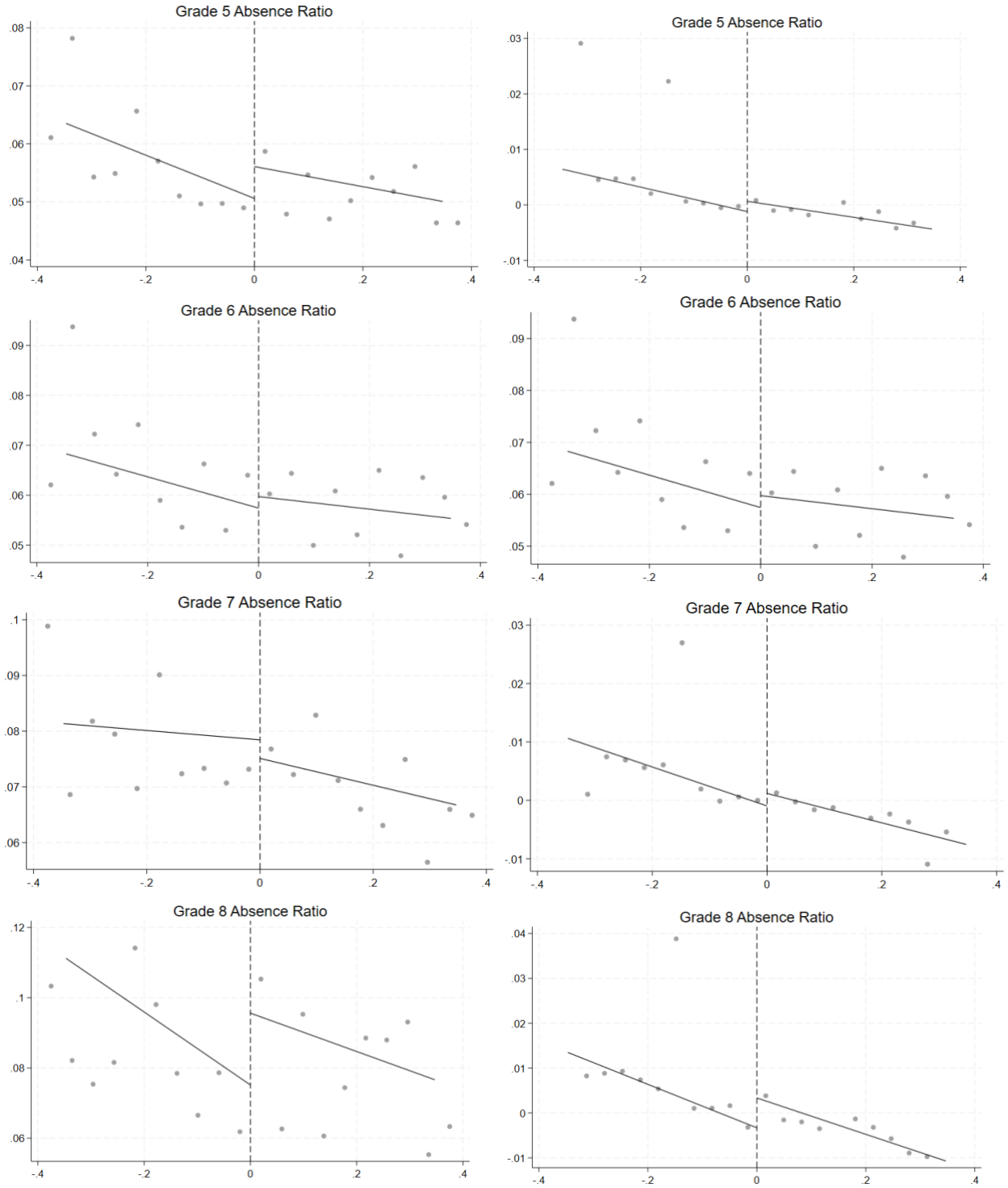
Panel B. 10 Bins per Side of the Threhold.

RAW



RESIDUALIZED



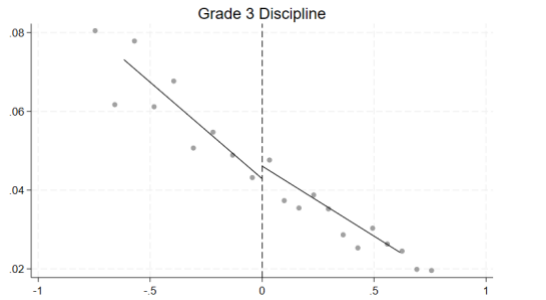


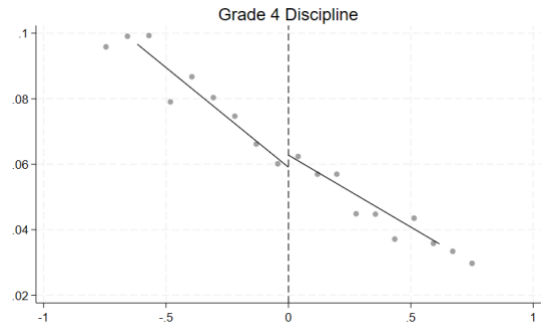
Notes: In both panels, the Raw column displays mean outcomes using unadjusted data, while the Residualized column displays mean outcomes after adjusting for year and students' Grade 3 On-Track status. In Panel A, gray dots represent the mean outcome within evenly spaced bins of the running variable. In Panel B, gray dots represent the mean outcome within 10 bins on each side of the retention threat threshold. Fitted lines correspond to regressions using the variable listed in the figure heading as the dependent variable and include controls for year and students' Grade 3 On-Track status. Effects are estimated within the absence ratio bandwidth of 0.347 SD. Bandwidth selected using Calonico et al. (2014) rdbwselect. The x-axis in each graph represents the running variable centered on the retention threat threshold in each year for each test.

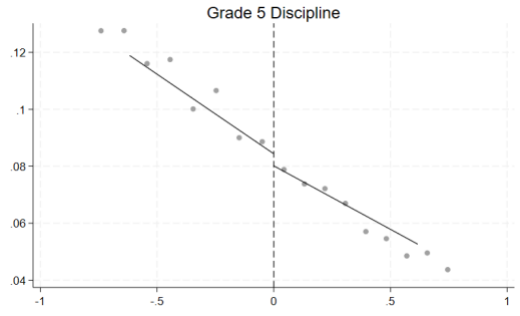
Figure A8. Robustness of discipline effects: estimated effect of retention threat on disciplinary effects using alternative bins.

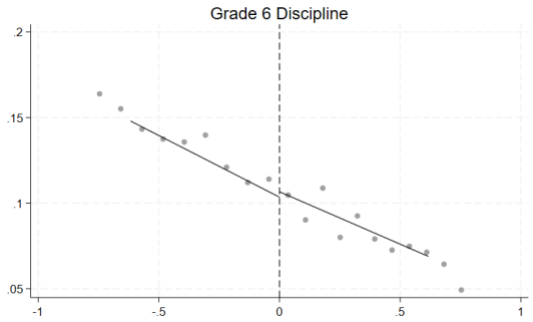
Panel A. Evenly Spaced Bins.

RAW

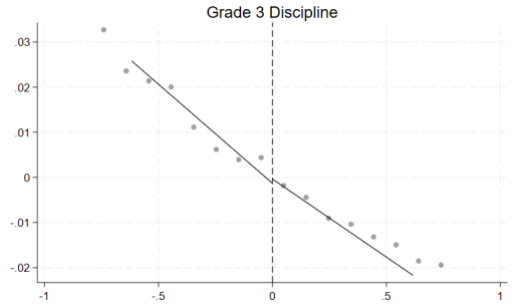


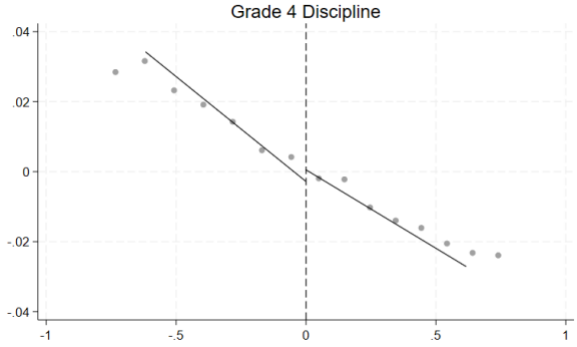


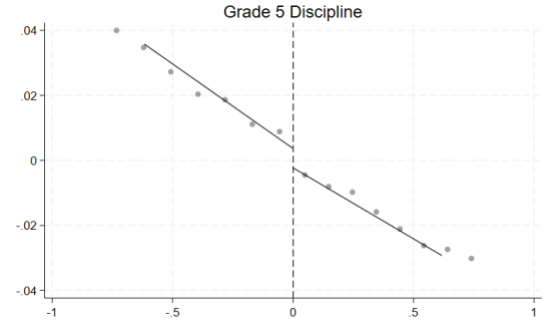


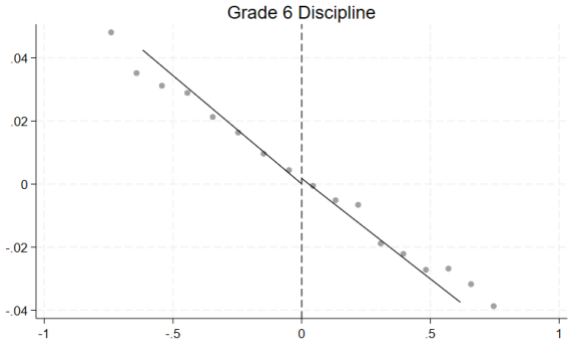


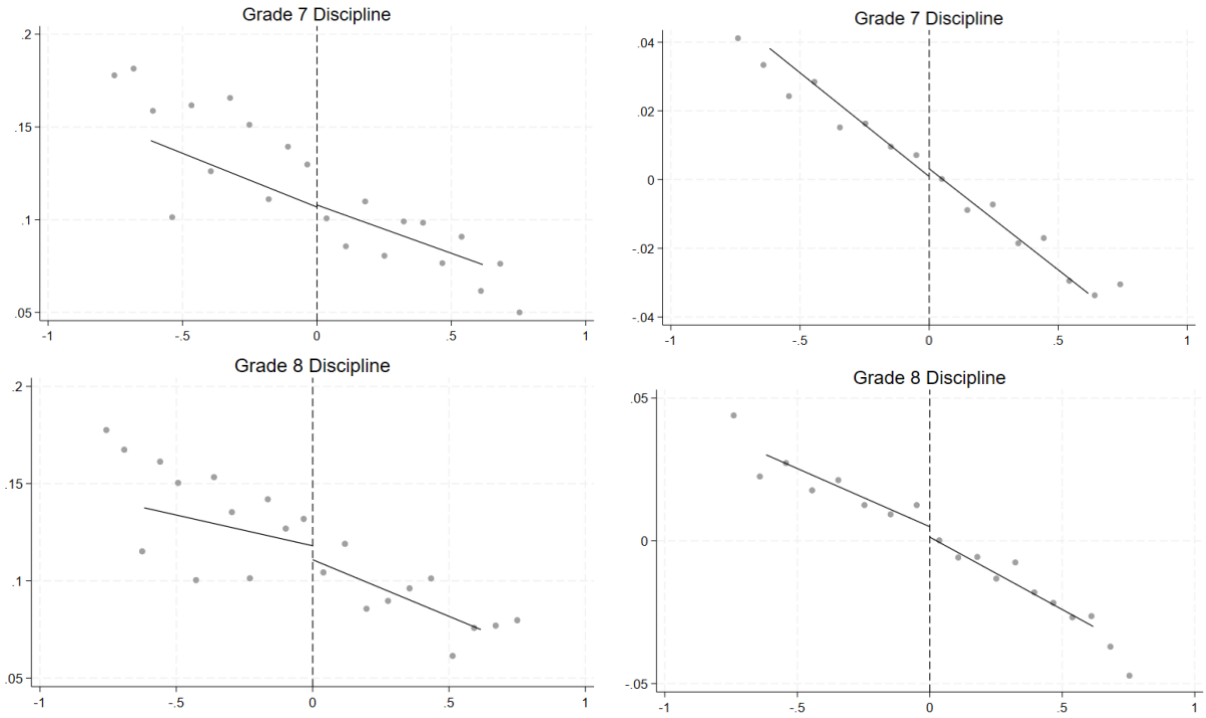
RESIDUALIZED





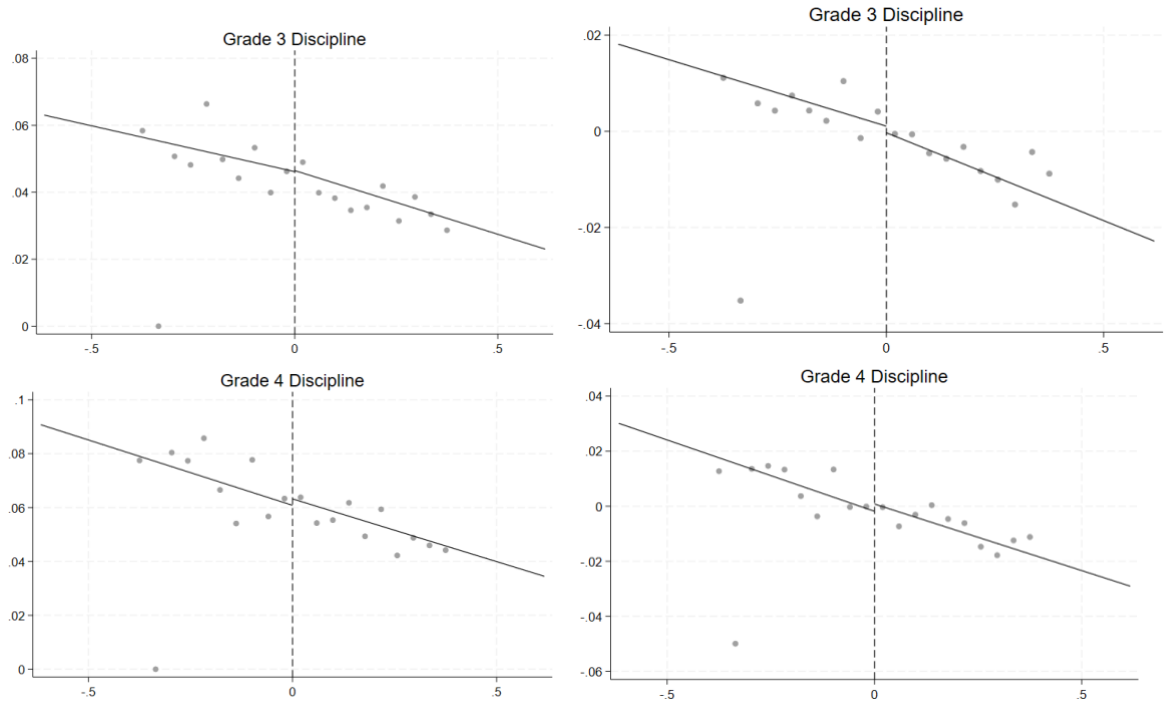


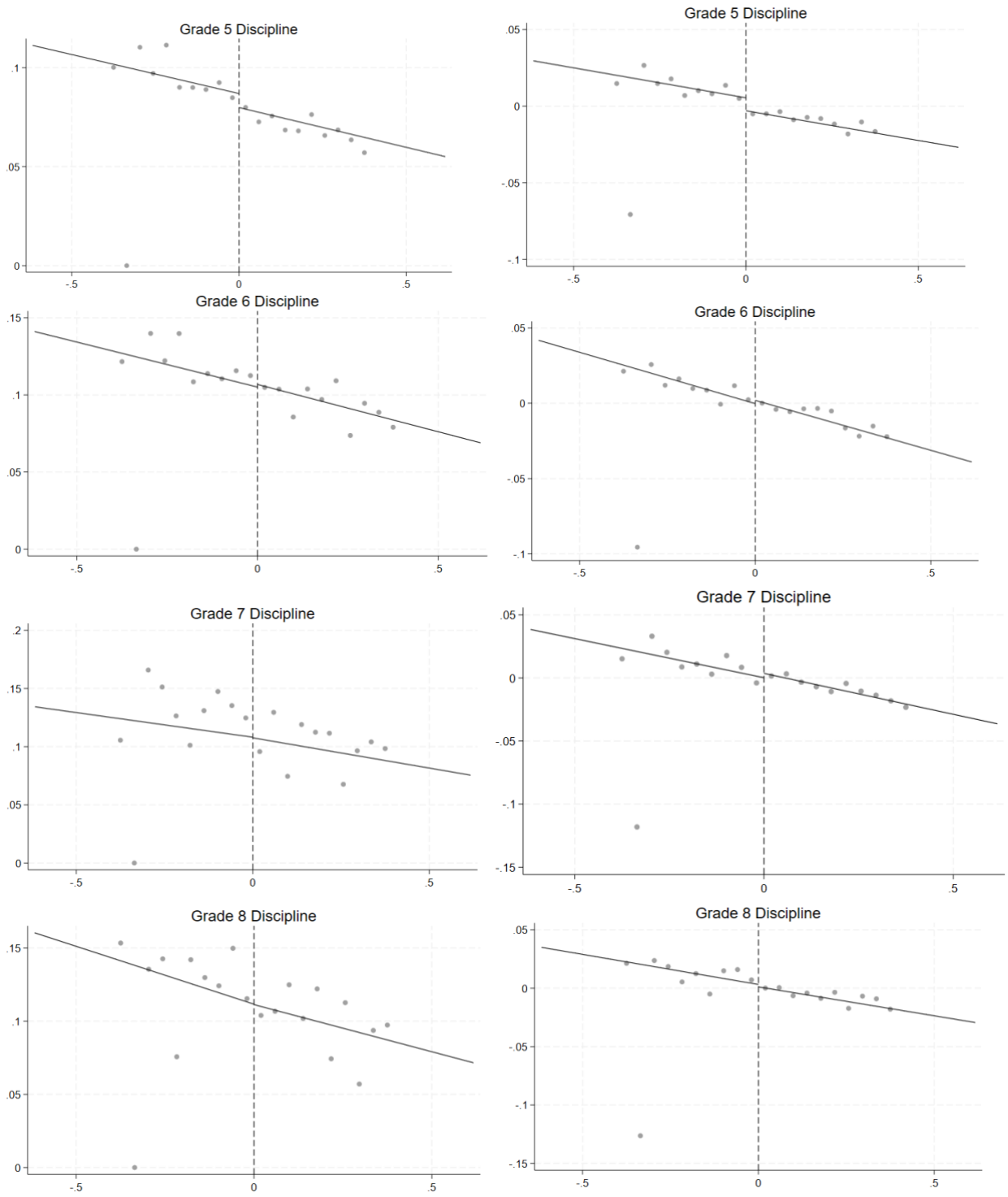




Panel B. 10 Bins per Side of the Threshold.

RAW RESIDUALIZED



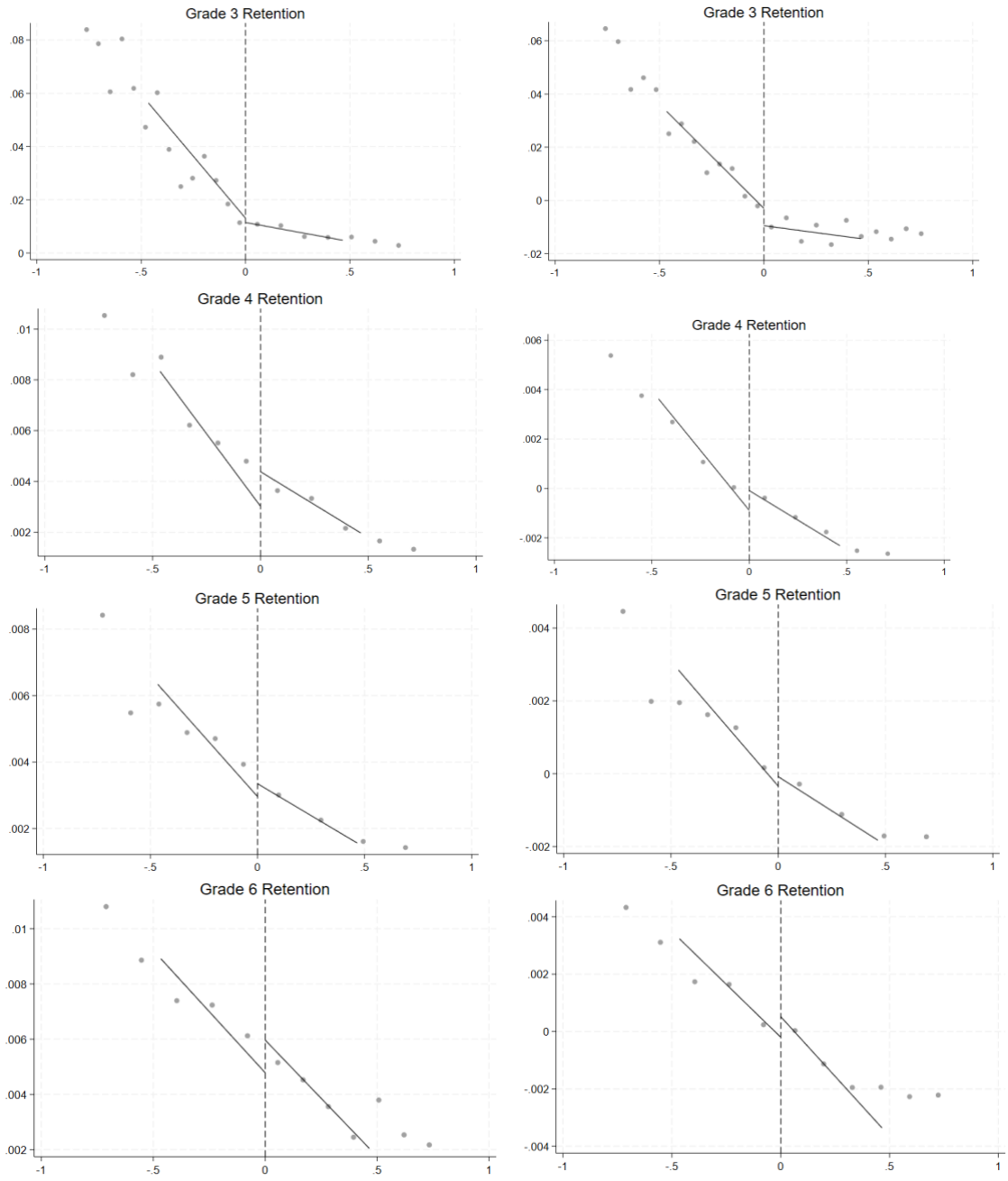


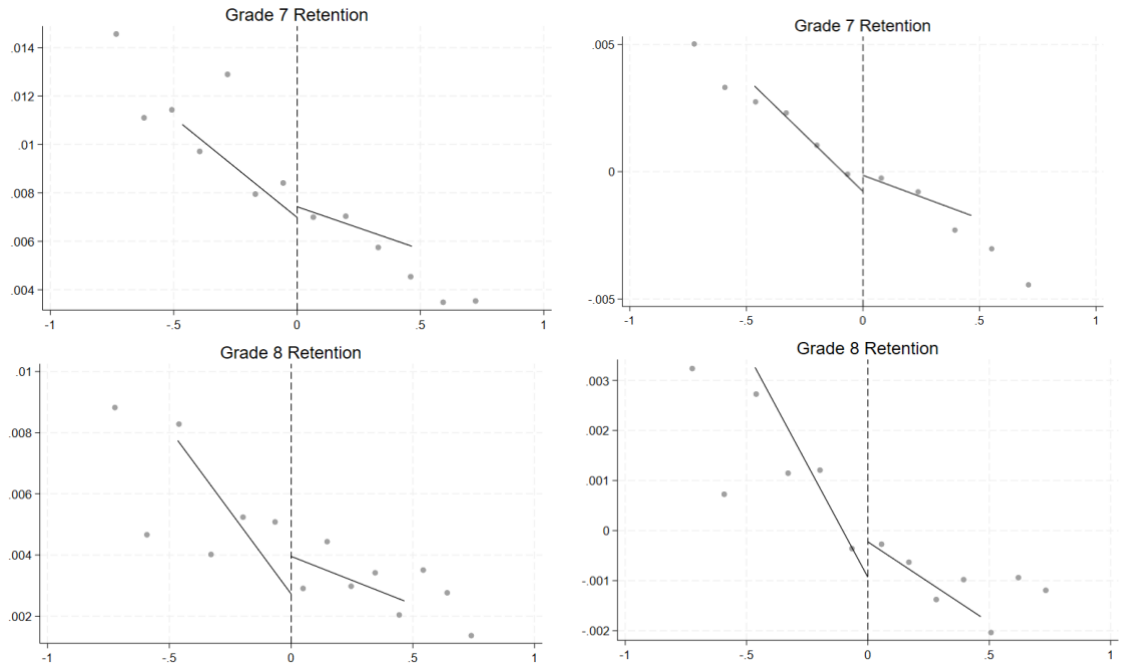
Notes: In Panel A, gray dots represent the mean outcome within evenly spaced bins of the running variable. In Panel B, gray dots represent the mean outcome within 10 bins on each side of the retention threat threshold. Fitted lines correspond to regressions using the variable listed in the figure heading as the dependent variable and include controls for year and students' Grade 3 On-Track status. Effects are estimated within the main discipline bandwidth of 0.617 SD. Bandwidth selected using Calonico et al. (2014) rdbwselect. The x-axis in each graph represents the running variable centered on the retention threat threshold in each year for each test.

Figure A9. Robustness of later-grade retention effects: estimated effect of retention threat on later-grade retention using alternative bins.

Panel A. Evenly Spaced Bins.

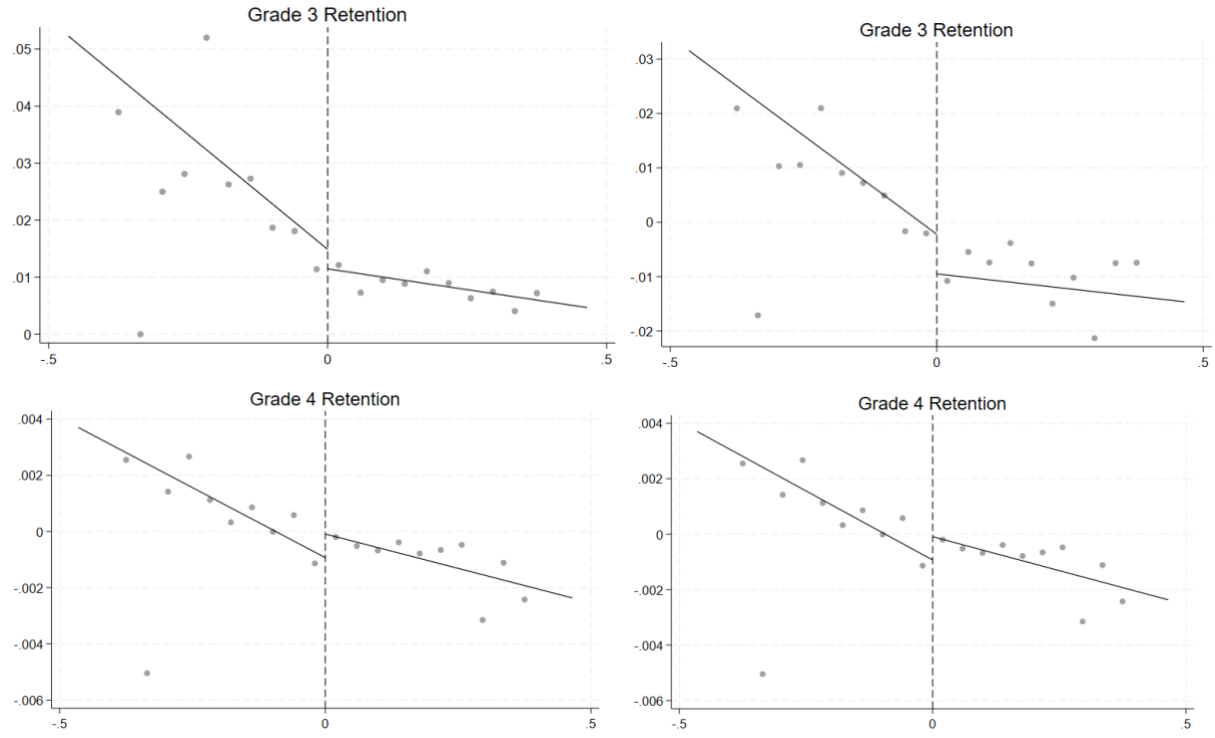
RAW RESIDUALIZED

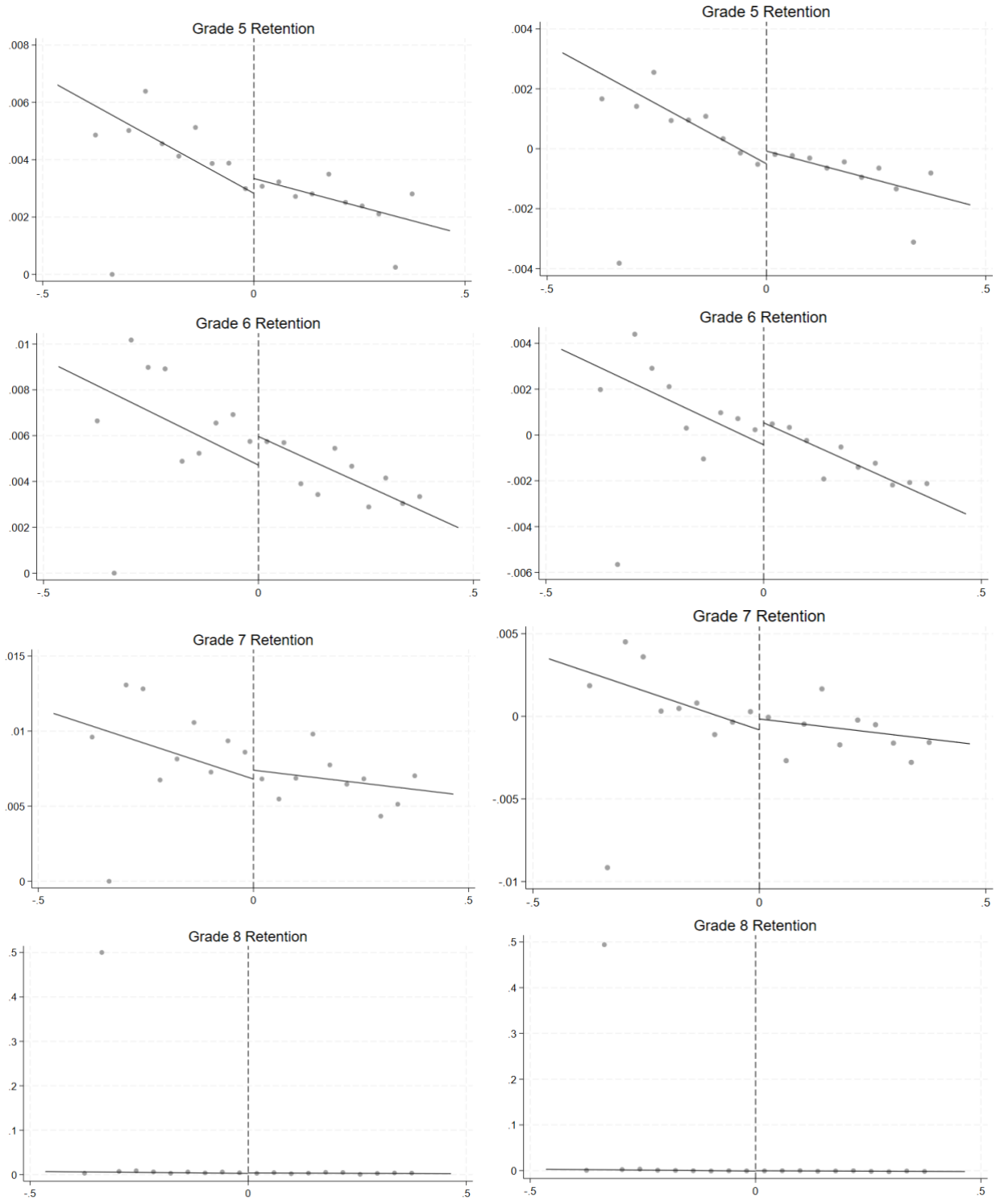




Panel B. 10 Bins per Side of the Threshold.

RAW RESIDUALIZED





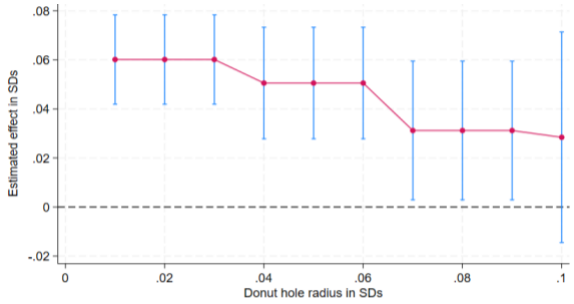
Notes: In Panel A, gray dots represent the mean outcome within evenly spaced bins of the running variable. In Panel B, gray dots represent the mean outcome within 10 bins on each side of the retention threat threshold. Fitted lines correspond to regressions using the variable listed in the figure heading as the dependent variable and include controls for year and students' Grade 3 On-Track status. Effects are estimated within the main discipline bandwidth of 0.465 SD. Bandwidth selected using Calonico et al.

(2014) rdbwselect. The x-axis in each graph represents the running variable centered on the retention threat threshold in each year for each test.

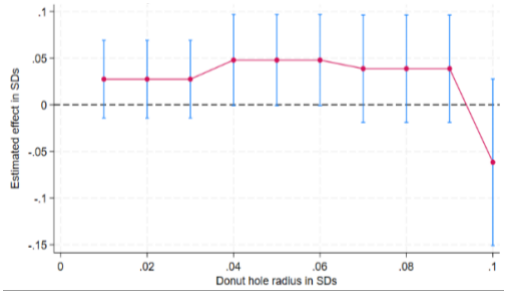
Figure A10. Robustness of effects: donut regression discontinuity estimates.

Panel A. ELA

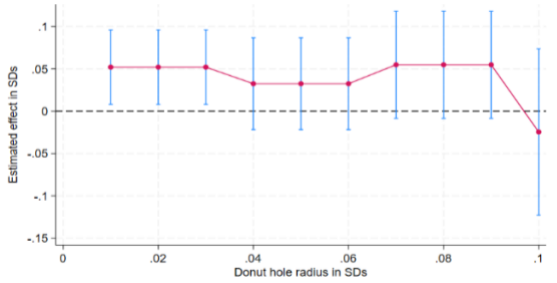
Grade 3



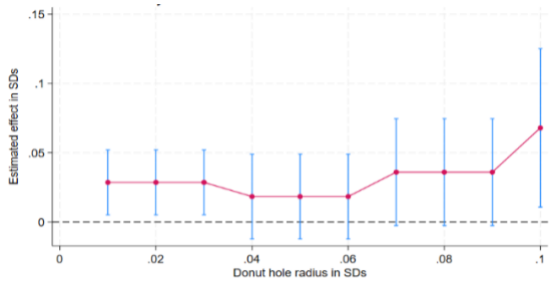
Grade 4



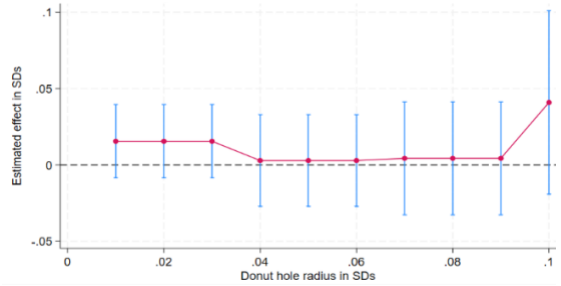
Grade 5



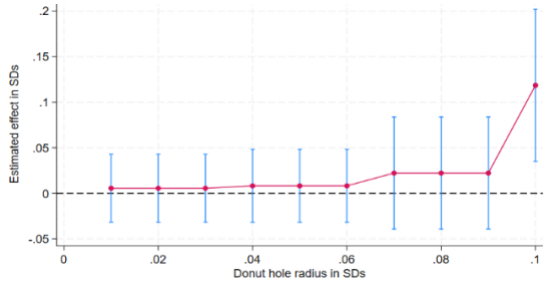
Grade 6



Grade 7

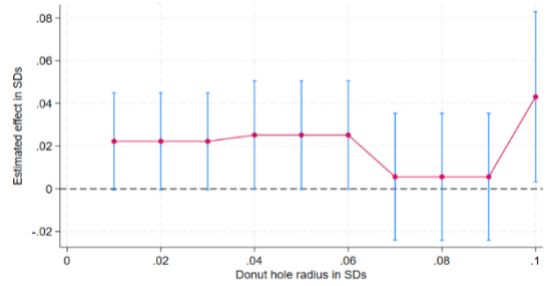


Grade 8

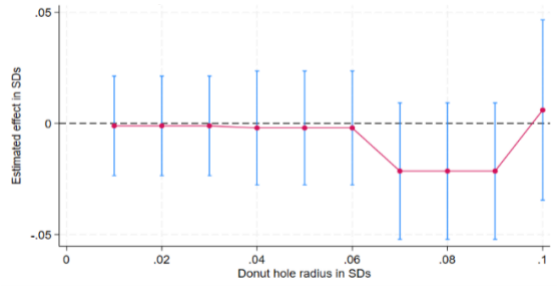


Panel B. Math

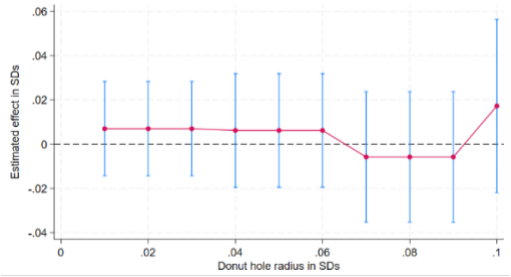
Grade 3



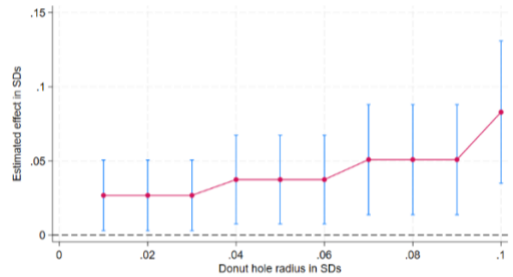
Grade 4



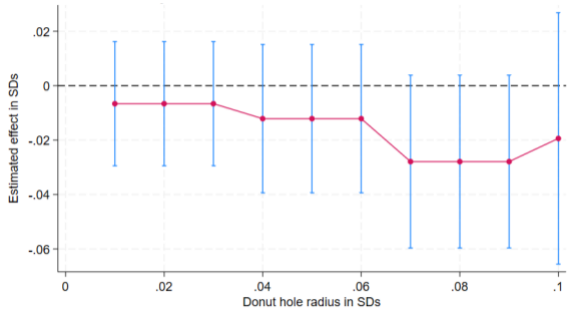
Grade 5



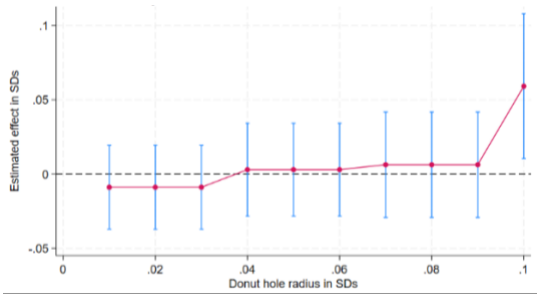
Grade 6



Grade 7

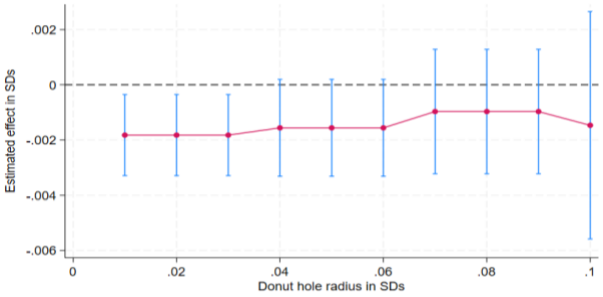


Grade 8

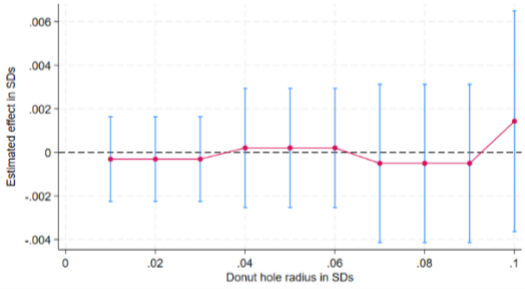


Panel C. Absence Ratio

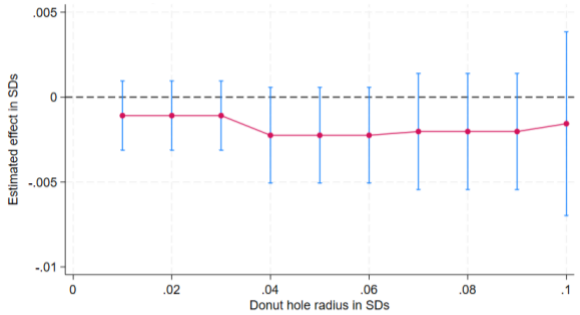
Grade 3



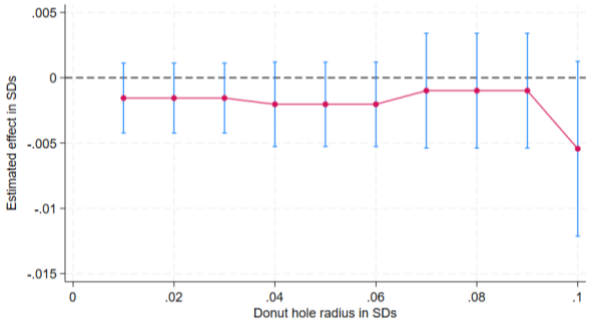
Grade 4



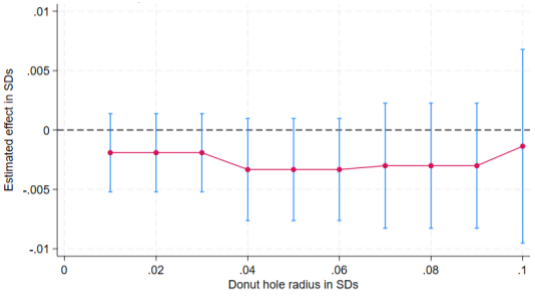
Grade 5



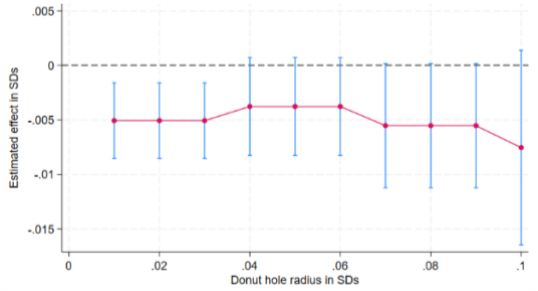
Grade 6



Grade 7



Grade 8

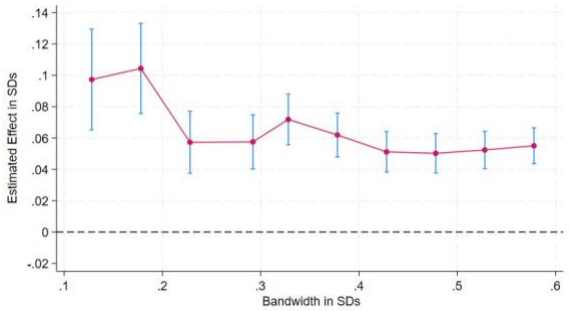


Notes: The figures present robustness of the estimated treatment effect to donut regression discontinuity specifications that exclude an increasing number of observations nearest the threshold. Donut hole radii are measured in standard deviations of the running variable and range from 0.01 SD to 0.10 SD. Points indicate coefficient estimates and vertical bars represent 95% confidence intervals. All specifications are otherwise identical to the main analysis.

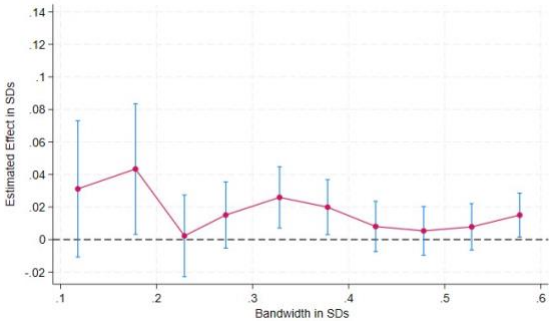
Figure A11. Robustness of effects: alternative bandwidths.

Panel A. ELA

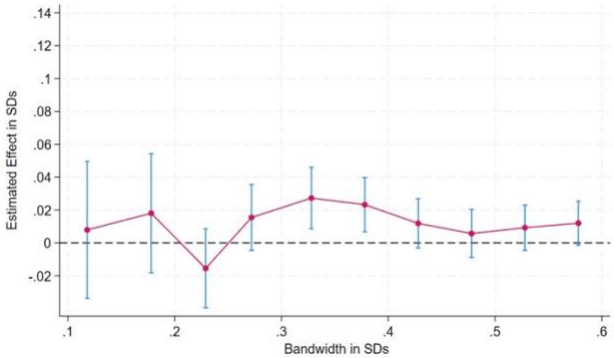
Grade 3



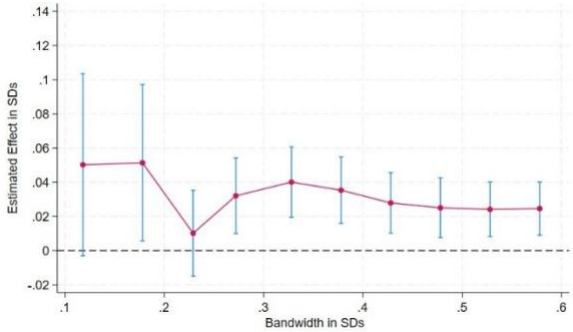
Grade 4



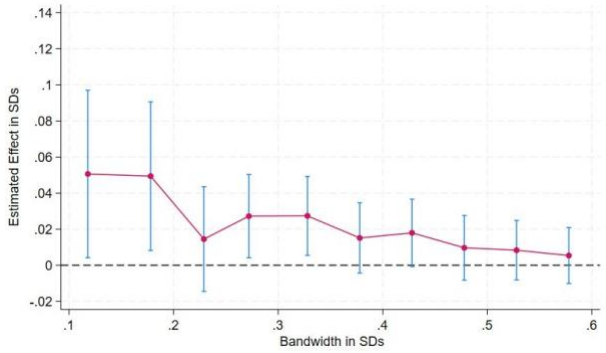
Grade 5



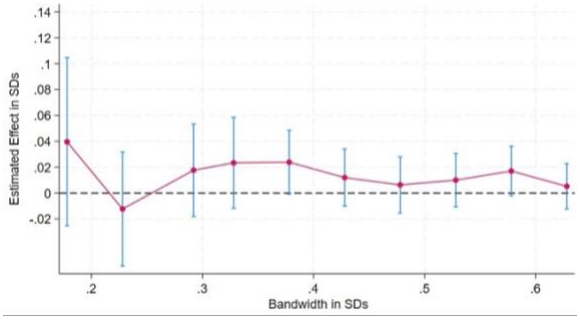
Grade 6



Grade 7

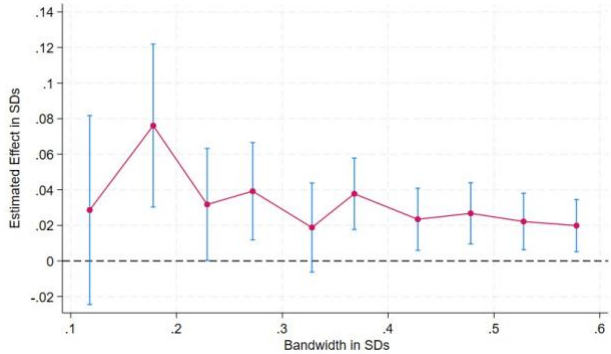


Grade 8

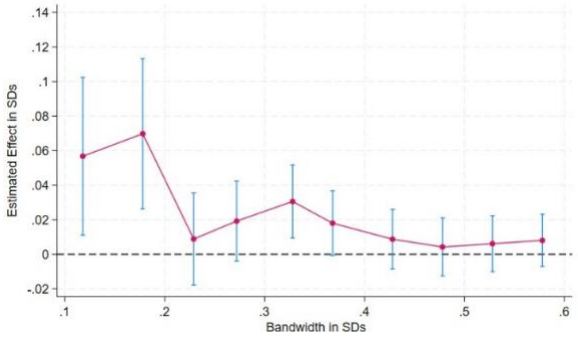


Panel B. Math

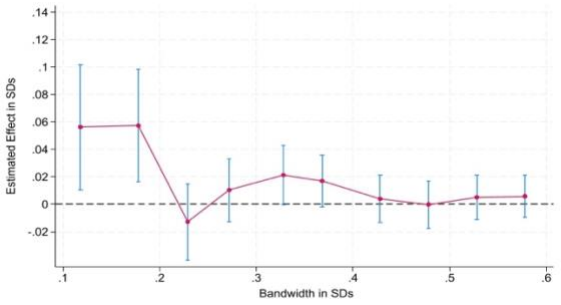
Grade 3



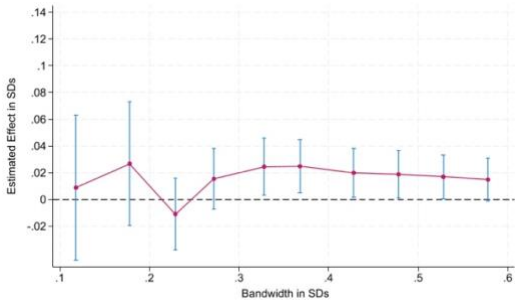
Grade 4



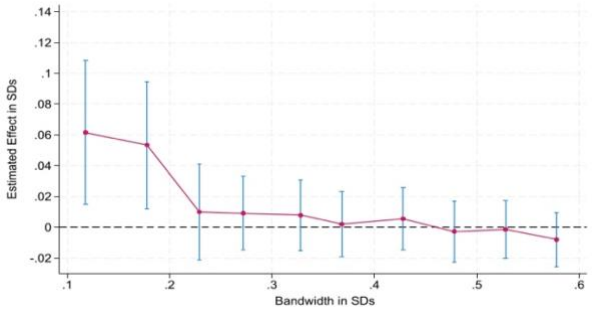
Grade 5



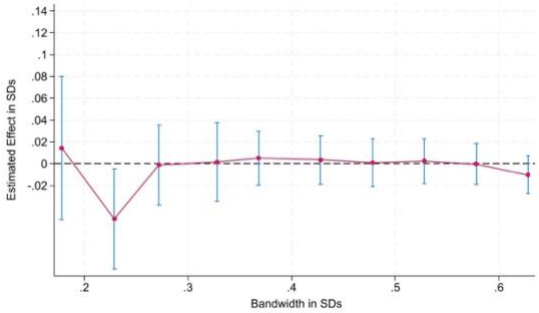
Grade 6



Grade 7

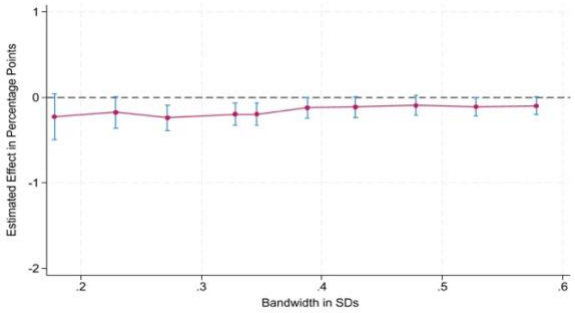


Grade 8

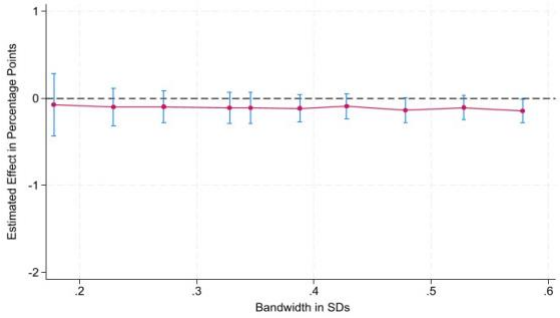


Panel C. Absence Ratio

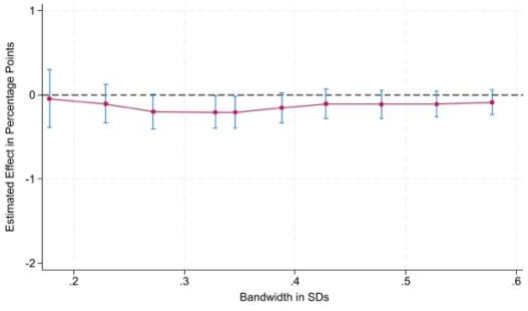
Grade 3



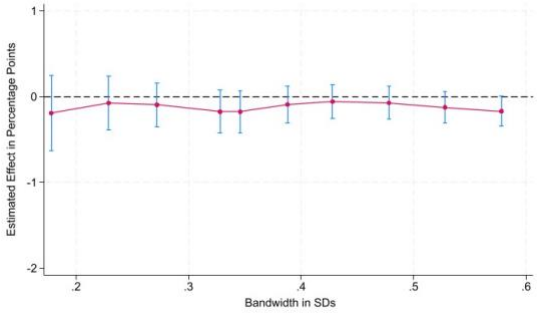
Grade 4



Grade 5

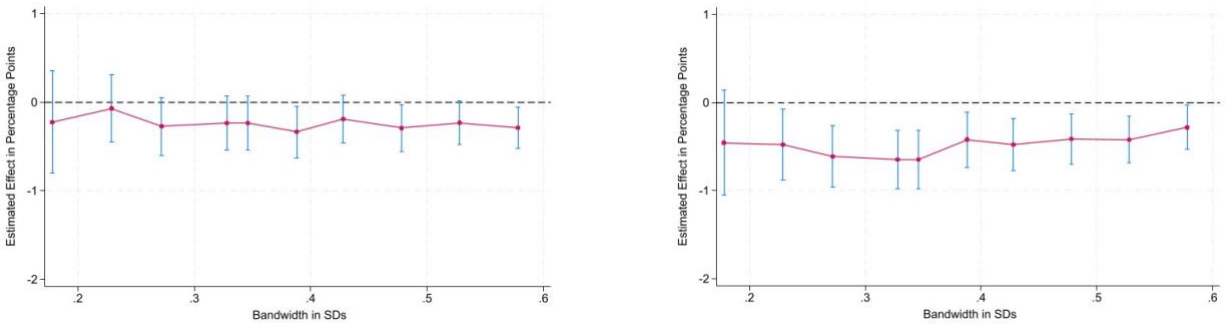


Grade 6



Grade 7

Grade 8

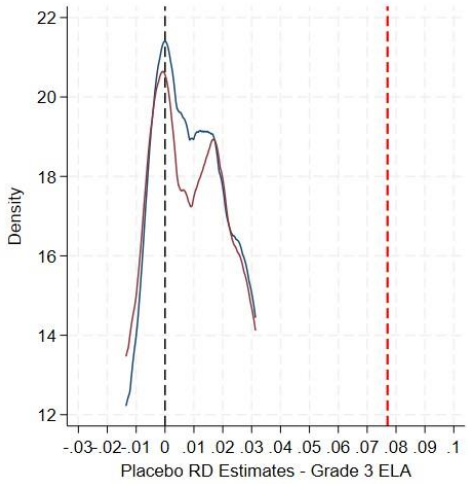


Notes: The figure displays robustness of the estimated treatment effect to alternative bandwidth choices around the regression discontinuity threshold. Points indicate coefficient estimates and vertical bars represent 95% confidence intervals. Bandwidths are expressed in standard deviations of the running variable. All specifications are otherwise identical to the main model.

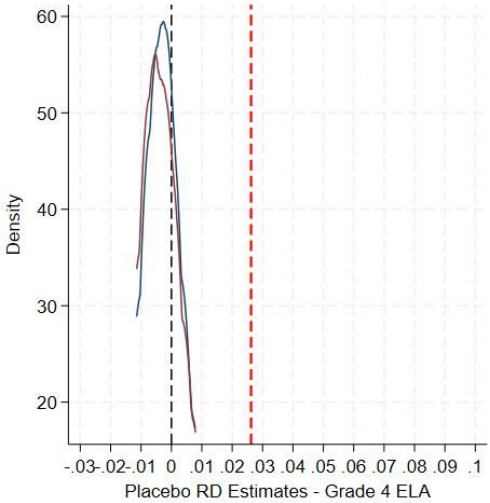
Figure A12. Distribution of effects at placebo cutoffs.

Panel A. ELA

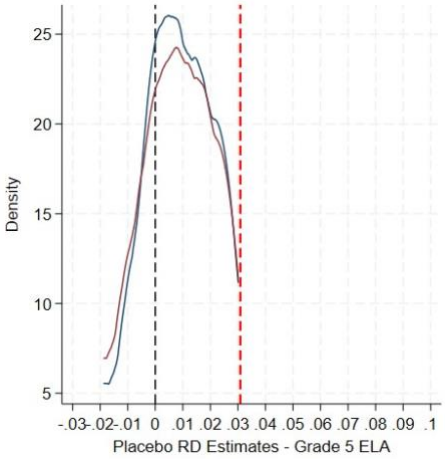
Grade 3



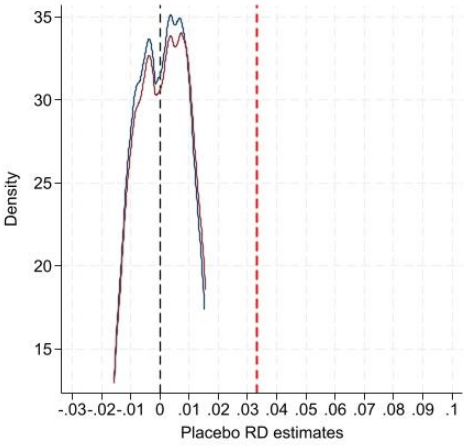
Grade 4



Grade 5

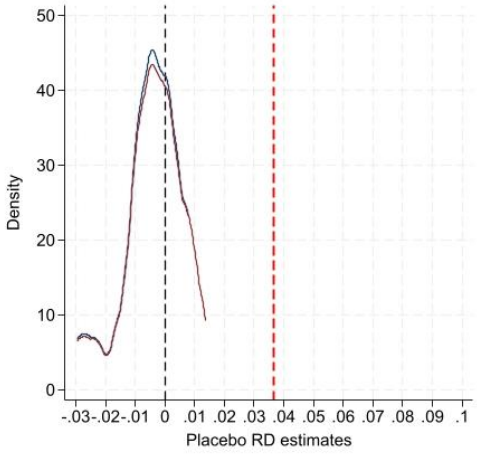


Grade 6

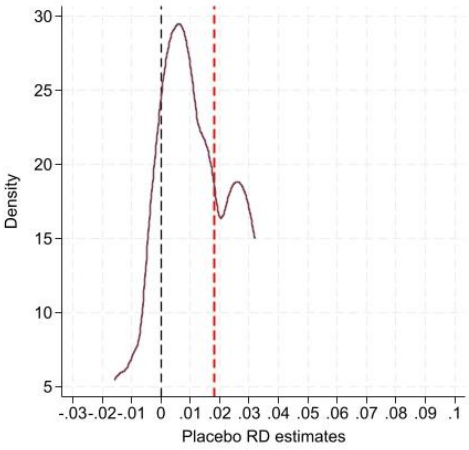


Panel B. Math

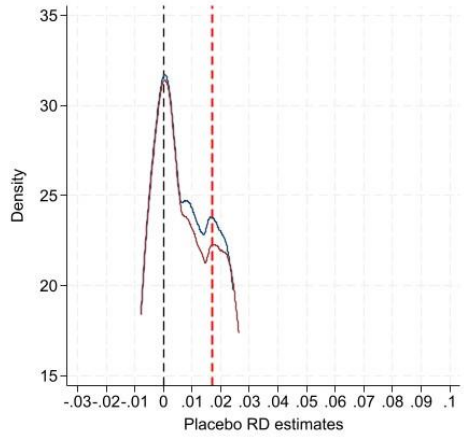
Grade 3



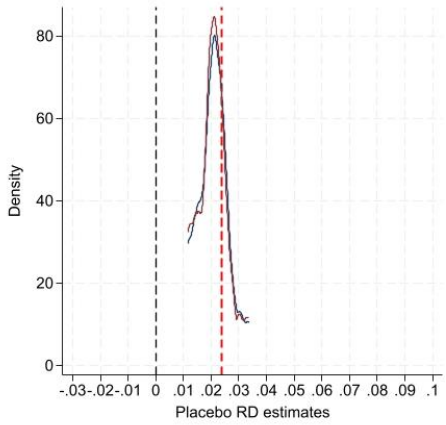
Grade 4



Grade 5

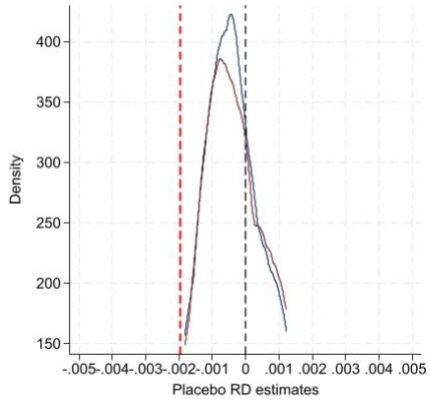


Grade 6

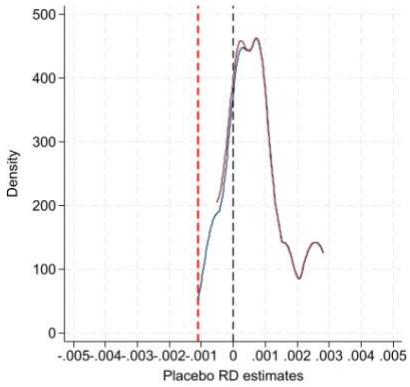


Panel C. Absence

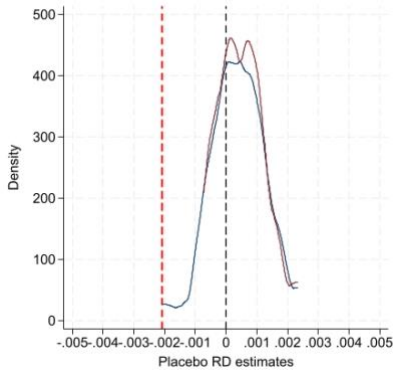
Grade 3



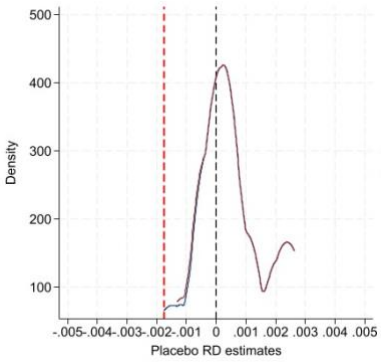
Grade 4



Grade 5



Grade 6



Notes: Each plot shows the distribution of placebo RD estimates generated using false promotion thresholds. Blue lines denote placebo estimates using cutoffs above the true threshold and red lines denote placebo estimates using cutoffs below the true threshold; the red dashed vertical line indicates the estimated effect at the true promotion threshold, while the black dashed vertical line indicates zero. Placebo cutoffs above the true threshold are estimated using only control observations, while placebo cutoffs below the true threshold are estimated using only treated observations, ensuring each placebo estimate is based solely on students with the same treatment status. Effects are estimated within outcome-specific bandwidths, calculated for models including the On-Track variable and cohort fixed effects. Bandwidths are selected using rdbwselect (Calonico et al., 2014). Analyses are presented through Grade 6 due to reduced statistical power in later grades.

Appendix Table A1. RIMP reading intervention program codes.

Summer Reading Programs
Extended Learning Time (Each Week)
Guided Reading (Small Group Instruction)
Increased Reading Time
Interactive Writing
Leveled Literacy Intervention
Literacy Collaborative
One-on-One Tutoring or Mentoring
Orton-Gillingham
Peer-Assisted Learning Strategies
Phonemic Awareness and Phonemic Decoding
Proactive Intervention
Reading Recovery
Responsive Intervention
Shared Reading
Sight Word Instruction
Success for All
Supplemental Instruction in Decoding Skills
Other Explicit Instruction of Comprehension Interventions
Other Explicit Instruction of Fluency Interventions
Other Explicit Instruction of Vocabulary Interventions
Other Language Experience Approach
Other Intervention

Notes: This table lists the reading intervention program codes recorded in Ohio Reading Improvement and Monitoring Plans (RIMPs). These interventions are assigned to improve literacy achievement and reflect the reading supports and instructional services provided to students by schools and districts.

Table A2. Robustness of ELA and Math Effects to Sequentially Saturated Specifications

	(1) Grade 3	(2) Grade 4	(3) Grade 5	(4) Grade 6	(5) Grade 7	(6) Grade 8
Covariates: Cohort						
ELA						
Threat (Fall)	0.073***	0.021**	0.027***	0.035***	0.024**	0.022
	(0.008)	(0.010)	(0.010)	(0.011)	(0.011)	(0.018)
Bandwidth	0.322	0.322	0.322	0.322	0.322	0.322
N	116,634	107,987	106,289	77,007	71,905	76,401
Math						
Threat (Fall)	0.031***	0.012	0.012	0.019*	-0.003	0.001
	(0.010)	(0.010)	(0.009)	(0.011)	(0.011)	(0.013)
Bandwidth	0.373	0.373	0.373	0.373	0.373	0.373
N	129,206	124,105	121,851	85,893	86,918	87,519
Covariates: Cohort, On-Track						
ELA						
Threat (Fall)	0.077***	0.026***	0.031***	0.039***	0.026**	0.023
	(0.008)	(0.010)	(0.010)	(0.011)	(0.011)	(0.017)
Bandwidth	0.328	0.328	0.328	0.328	0.328	0.328
N	120,368	111,848	110,090	80,733	75,580	76,441
Math						
Threat (Fall)	0.037***	0.018*	0.017*	0.024**	0.002	0.005
	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.013)
Bandwidth	0.368	0.368	0.368	0.368	0.368	0.368
N	128,915	124,105	121,851	85,893	86,918	87,519
Covariates: Cohort, On-Track, Observables at Baseline						
ELA						
Threat (Fall)	0.072***	0.021**	0.024***	0.029***	0.022**	0.018
	(0.008)	(0.010)	(0.010)	(0.011)	(0.011)	(0.017)
Bandwidth	0.328	0.328	0.328	0.328	0.328	0.328
N	120,368	111,848	110,090	80,733	75,580	76,441
Math						
Threat (Fall)	0.036***	0.021**	0.027***	0.035***	0.024**	0.022
	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.018)
Bandwidth	0.368	0.368	0.368	0.368	0.368	0.368
N	128,915	124,105	121,851	85,893	86,918	87,519
Covariates: Cohort, On-Track, Observables at Baseline with updated bandwidths.						
ELA						
Threat (Fall)	0.072***	0.021**	0.024***	0.029***	0.022**	0.018
	(0.008)	(0.010)	(0.010)	(0.011)	(0.011)	(0.017)
Bandwidth	0.331	0.331	0.331	0.331	0.331	0.331
N	120,368	111,848	110,090	80,733	75,580	76,441
Math						
Threat (Fall)	0.013	0.032***	0.022***	0.025**	0.009	0.001
	(0.013)	(0.012)	(0.011)	(0.011)	(0.012)	(0.018)
Bandwidth	0.317	0.317	0.317	0.317	0.317	0.317
N	111,396	107,041	105,236	76,015	71,129	75,643

Notes: Coefficient estimates in each column are from separate regressions using the variable in the row heading as the dependent variable and estimated separately for each grade. All estimates are from models which include the running score, On-Track, and cohort fixed effects. Effects are estimated within each outcome-specific bandwidth, calculated for a model that includes cohort fixed effects and the On-Track

variable. Bandwidth selected using Calonico et al. (2014) rdbwselect. Robust standard errors clustered on the running variable are shown in parentheses. *p< 0.10; **p< 0.05; ***p< 0.01.

Table A3. Regressions of Treatment Status on Baseline Student Covariates

	Threat
Female	-0.000
	(0.001)
Asian	0.005
	(0.007)
Black	-0.002
	(0.003)
Hispanic	-0.003
	(0.004)
White	0.001
	(0.003)
FRPL	0.000
	(0.002)
ELL	0.002
	(0.004)
SWD	-0.002
	(0.002)
Attrition	-0.006
	(0.005)
On-Track	-0.006***
	(0.002)
P-Value of Tests	
Female-White	0.493
Asian-White	0.334
FRPL – SWD	0.516
Female-SWD	0.720
Female-Attrition	0.692
All: Female-On-Track	0.082
N	123,082

Notes: Coefficient estimates are from regressions of treatment status on baseline Grade 3 student covariates estimated within the ELA bandwidth. Robust standard errors are shown in parentheses. Reported p-values correspond to tests of equality between selected coefficients. The final row reports the p-values from joint F-tests of covariate significance. *p< 0.10; **p< 0.05; ***p< 0.01.

Table A4. Test of Differences in Attrition Across Grades

	(1) Grade 3	(2) Grade 4	(3) Grade 5	(4) Grade 6	(5) Grade 7	(6) Grade 8
Missing ELA Test Score	-0.002	0.002	-0.004	-0.001	-0.009	-0.010
	(0.002)	(0.004)	(0.005)	(0.005)	(0.005)	(0.009)
Missing Math Test Score	-0.003	-0.000	-0.004	0.006	-0.004	-0.005
	(0.003)	(0.004)	(0.004)	(0.005)	(0.005)	(0.006)
Missing Attendance Records	-0.003	-0.000	-0.004	0.000	0.001	0.001
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Missing Disciplinary Records	-	0.003**	0.004*	0.003	0.004*	0.004
		(0.002)	(0.002)	(0.002)	(0.002)	(0.003)

Notes: Coefficient estimates in each column are from separate regressions replacing the dependent variable in the primary RD specification with indicators for missing outcome data. All models include the running variable, cohort fixed effects, and the Grade 3 On-Track variable and are estimated within the main outcome-specific bandwidths in Grade 3 selected using Calonico et al. (2014) rdbwselect including cohort fixed-effects and On-Track. Robust standard errors are shown in parentheses. Grade 3 estimates for missing disciplinary records are omitted because no disciplinary records are missing. ***p< 0.01; **p< 0.05; *p< 0.10**.**

Table A5. Robustness of ELA, Math, and Absence Effects to Model Specification

	(1) Grade 3	(2) Grade 4	(3) Grade 5	(4) Grade 6	(5) Grade 7	(6) Grade 8
Quadratic Specification						
ELA						
Threat (Fall)	0.134***	0.046*	0.044*	0.083***	0.074**	0.185**
	(0.022)	(0.027)	(0.027)	(0.029)	(0.031)	(0.092)
Math						
Threat (Fall)	0.140***	0.039	0.037	-0.010	0.049*	-0.050
	(0.030)	(0.027)	(0.026)	(0.027)	(0.028)	(0.060)
Absence						
Threat (Fall)	-0.001	-0.003	-0.000	-0.002	-0.001	-0.009**
	(0.002)	(0.003)	(0.002)	(0.004)	(0.004)	(0.004)
No Clustering						
ELA						
Threat (Fall)	0.072***	0.026**	0.031***	0.039***	0.027**	0.023
	(0.008)	(0.010)	(0.011)	(0.011)	(0.011)	(0.018)
Math						
Threat (Fall)	0.037***	0.018*	0.017*	0.024**	0.002	0.005
	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.012)
Absence						
Threat (Fall)	-0.002***	-0.001	-0.002**	-0.002	-0.002	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
School-Level Clustering						
ELA						
Threat (Fall)	0.056**	0.040	0.042**	0.063***	0.043**	0.013
	(0.020)	(0.051)	(0.019)	(0.024)	(0.017)	(0.017)
Math						
Threat (Fall)	0.110***	0.041***	0.035*	0.033	0.035	-0.012
	(0.034)	(0.011)	(0.018)	(0.022)	(0.026)	(0.020)
Absence						
Threat (Fall)	-0.003***	-0.001	-0.005***	-0.002	0.003	-0.020
	(0.001)	(0.001)	(0.002)	(0.005)	(0.004)	(0.011)
rdrobust						
ELA	0.148***	0.040***	0.073***	0.063***	0.049***	0.081***
Threat (Fall)	(0.009)	(0.011)	(0.010)	(0.012)	(0.012)	(0.019)
Math	0.120***	0.051***	0.041***	0.033***	0.012*	0.022
Threat (Fall)	(0.012)	(0.012)	(0.011)	(0.012)	(0.012)	(0.017)
Absence	-0.004***	-0.001	-0.005***	-0.002*	0.003**	-0.021***
Threat (Fall)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
RDHonest						
ELA						
Threat (Fall)	0.095***	-0.019	0.068**	0.180***	0.025**	-0.464**
	(0.021)	(0.012)	(0.027)	(0.028)	(0.030)	(0.177)
Math						
Threat (Fall)	0.238***	0.005	0.022***	0.113**	-0.033	-0.521
	(0.036)	(0.013)	(0.011)	(0.056)	(0.032)	(0.148)
Absence						
Threat (Fall)	-0.003***	-0.009***	-0.013***	0.010***	-0.003	-0.066***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.004)	(0.004)

Notes: Coefficient estimates in each cell are from separate regressions using the variable in the row heading as the dependent variable and estimated for the grade indicated by the column header. All estimates are from models which include the running score, On-Track, and cohort fixed effects. Effects are estimated within each outcome-specific bandwidth, calculated for a model that includes cohort fixed effects and the On-Track variable. Bandwidth

selected using Calonico et al. (2014) rdbwselect including controls for cohort and On-Track.. Robust standard errors clustered on the running variable are shown in parentheses unless indicated otherwise. *p< 0.10; **p< 0.05; ***p< 0.01.

Table A6. Heterogeneity in effect of retention threat on Grade 3 and Grade 6 nonacademic outcomes by student characteristics.

	(1) On Track	(2) Not On Track	(3) Female	(4) Male	(5) Asian	(6) Black	(7) Hispanic	(8) White	(9) FRPL	(10) Non-FRPL	(11) ELL	(12) Non-ELL	(13) SWD	(14) Non-SWD
Grade 3	-0.003***	-0.001	-0.002*	-0.002**	-0.005	0.002	-0.006**	-0.001	-0.001	-0.002*	-0.000	-0.002	-0.005**	-0.002**
Absence	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.002)	(0.001)
P-Value of Test	0.221		0.964			0.102			0.734		0.541		0.124	
N	80,867	35,029	55,516	60,380	1,782	20,870	7,464	79,169	69,500	46,396	5,642	110,254	14,790	101,106
Discipline	-0.003	0.001	-0.003	-0.002	0.017	0.002	0.008	-0.001	-0.001	-0.004	-0.001	-0.003	0.005	-0.004
	(0.002)	(0.004)	(0.002)	(0.003)	(0.011)	(0.008)	(0.009)	(0.002)	(0.004)	(0.003)	(0.007)	(0.002)	(0.006)	(0.002)
P-Value of Test	0.408		0.810			0.330			0.548		0.835		0.148	
N	154,912	65,287	105,691	114,508	3,665	41,418	14,585	147,894	134,231	85,968	11,466	208,733	29,988	190,211
Retention	-0.000	0.018***	0.005**	0.009***	0.018**	0.055***	0.043***	0.016***	0.028***	0.003	0.013*	0.007***	0.011**	0.006***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.010)	(0.005)	(0.007)	(0.002)	(0.003)	(0.002)	(0.007)	(0.002)	(0.004)	(0.002)
P-Value of Test	0.000		0.215			0.000			0.000		0.350		0.268	
N	118,583	48,989	80,422	87,150	2,751	31,262	11,119	112,841	102,192	65,380	8,689	158,883	22,238	145,334
Grade 6	-0.002	-0.001	-0.002	-0.002	-0.008	-0.007	-0.001	-0.002	-0.005**	-0.002	0.002	-0.002	-0.000	-0.002
Absence	(0.002)	(0.002)	(0.002)	(0.002)	(0.006)	(0.004)	(0.005)	(0.001)	(0.002)	(0.001)	(0.005)	(0.001)	(0.004)	(0.001)
P-Value of Test	0.588		0.896			0.553			0.238		0.442		0.616	
N	80,521	33,286	54,582	59,225	1,652	20,384	7,225	78,075	68,667	45,140	5,433	108,374	14,546	99,261
Discipline	0.007*	-0.012**	-0.005	0.004	0.009	-0.015	-0.021	-0.004	-0.013***	-0.003	-0.011	0.001	-0.005	0.001
	(0.003)	(0.005)	(0.004)	(0.004)	(0.022)	(0.011)	(0.013)	(0.004)	(0.005)	(0.004)	(0.011)	(0.003)	(0.008)	(0.003)
P-Value of Test	0.003		0.128			0.439			0.138		0.835		0.510	
N	144,430	61,652	98,934	107,148	3,066	37,649	13,197	140,464	124,933	81,149	9,976	196,106	28,064	178,018
Retention	0.000	-0.001	-0.001	0.001	-0.003*	-0.004	0.006	0.001	0.000	-0.000	0.000	-0.000	-0.001	-0.000
	(0.001)	(0.002)	(0.000)	(0.001)	(0.002)	(0.003)	(0.004)	(0.001)	(0.002)	(0.001)	(0.003)	(0.001)	(0.002)	(0.001)
P-Value of Test	0.660		0.944			0.055			0.713		0.901		0.640	
N	110,972	46,421	75,499	81,894	2,298	28,716	10,091	107,362	95,599	61,794	7,583	149,810	20,906	136,487

Notes: Coefficient estimates for each group: sex, race, FRPL, ELL, and SWD come from separate regressions where the variable in the row heading is the dependent variable. The coefficient on the dummy variable Threat or Threat plus student type is displayed to show the estimated effect of retention threat for students of all types within a group. All estimates displayed are from models which control for and interacts student type indicator variables with the running variable. The model also controls for year and a student's On-Track status in Grade 3. Effects are estimated within outcome-specific bandwidths determined in Grade 3 that include controls for cohort and On-Track. Robust standard errors clustered on the running variable are shown in parentheses. The p-value of test

indicates the statistical significance of a test that student types within the same group are equivalent. The reported Ns correspond to each analytic sample; however, each pair of estimates within a group comes from the same regression. *p< 0.10; **p< 0.05; ***p< 0.01.

Table A7. Heterogeneity in effect of retention threat on Grade 4 outcomes by student characteristics.

	(1) On Track	(2) Not On Track	(3) Female	(4) Male	(5) Asian	(6) Black	(7) Hispanic	(8) White	(9) FRPL	(10) Non-FRPL	(11) ELL	(12) Non-ELL	(13) SWD	(14) Non-SWD
Academic Outcomes														
	0.007	0.062***	0.058***	-0.002	0.186**	0.127***	0.129***	0.096***	0.104***	0.085***	0.002	0.027***	-0.036	0.033**
ELA	(0.013)	(0.017)	(0.014)	(0.015)	(0.086)	(0.023)	(0.039)	(0.012)	(0.015)	(0.020)	(0.039)	(0.011)	(0.033)	(0.011)
P-Value of Test	0.011		0.003			0.443			0.448		0.540		0.052	
N	79,262	32,586	53,703	58,145	1,737	19,660	7,199	76,940	66,492	45,356	5,509	106,339	14,023	97,825
Math	0.005	0.036	0.035**	0.003	0.016	0.140***	0.122***	0.056***	0.073***	0.104***	0.026	0.018*	-0.000	0.018*
	(0.013)	(0.018)	(0.014)	(0.015)	(0.107)	(0.022)	(0.042)	(0.014)	(0.015)	(0.021)	(0.042)	(0.011)	(0.028)	(0.011)
P-Value of Test	0.161		0.118			0.004			0.231		0.832		0.583	
N	87,325	36,780	59,511	64,594	1,907	22,017	8,116	79,709	74,078	50,027	6,195	117,910	15,876	108,229
Non-Academic Outcomes														
	-0.001	-0.002	-0.001	-0.001	0.014	-0.000	-0.006	-0.000	-0.002	-0.001	0.008	-0.002	0.001	-0.001
Absence	(0.001)	(0.002)	(0.001)	(0.001)	(0.011)	(0.003)	(0.004)	(0.001)	(0.002)	(0.001)	(0.005)	(0.001)	(0.003)	(0.001)
P-Value of Test	0.511		0.937			0.343			0.776		0.085		0.416	
N	80,992	33,288	54,841	59,439	1,788	20,207	7,369	78,422	68,269	46,011	5,653	108,627	14,542	99,738
Discipline	-0.003	-0.003	-0.004	-0.003	0.014	0.015	-0.028***	-0.005*	-0.007*	-0.003	-0.016**	-0.003	-0.008	-0.003
	(0.003)	(0.004)	(0.003)	(0.004)	(0.010)	(0.009)	(0.009)	(0.003)	(0.004)	(0.003)	(0.008)	(0.002)	(0.007)	(0.002)
P-Value of Test	0.974		0.026			0.067			0.351		0.091		0.452	
N	149,789	63,889	102,559	111,119	3,413	39,499	13,970	144,565	129,845	83,833	10,820	202,858	29,143	184,535
Retention	-0.000	-0.002	-0.002	-0.000	0.002**	0.001	0.002	-0.000	-0.002	-0.002	-0.006***	-0.001	0.000	-0.001
	(0.009)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
P-Value of Test	0.273		0.429			0.763			0.998		0.020		0.520	
N	115,092	48,109	78,303	84,898	2,572	30,091	10,700	110,505	99,347	63,854	8,237	154,964	21,716	141,485

Notes: Coefficient estimates for each group: sex, race, FRPL, ELL, and SWD come from separate regressions where the variable in the row heading is the dependent variable. The coefficient on the dummy variable Threat or Threat plus student type is displayed to show the estimated effect of retention threat for students of all types within a group. All estimates displayed are from models which control for and interacts student type indicator variables with the running variable. The model also controls for year and a student's On-Track status in Grade 3. Effects are estimated within outcome-specific bandwidths determined in Grade 3 that include controls for cohort and On-Track. Robust standard errors clustered on the running variable are shown in parentheses. The p-value of test indicates the statistical significance of a test that student types within the same group are equivalent. The reported Ns correspond to each analytic sample; however, each pair of estimates within a group comes from the same regression. *p< 0.10; **p< 0.05; ***p< 0.01.

Table A8. Heterogeneity in effect of retention threat on Grade 5 outcomes by student characteristics.

	(1) On Track	(2) Not On Track	(3) Female	(4) Male	(5) Asian	(6) Black	(7) Hispanic	(8) White	(9) FRPL	(10) Non-FRPL	(11) ELL	(12) Non-ELL	(13) SWD	(14) Non-SWD
Academic Outcomes														
ELA	0.003	0.089***	0.053***	0.011***	0.186*	0.106***	0.135***	0.113***	0.120***	0.100***	0.013	0.032***	0.037	0.025**
	(0.012)	(0.018)	(0.014)	(0.014)	(0.098)	(0.025)	(0.040)	(0.040)	(0.016)	(0.019)	(0.036)	(0.010)	(0.031)	(0.010)
P-Value of Test	0.000		0.035			0.823			0.428		0.674		0.730	
N	78,098	31,992	52,845	57,245	1,672	19,216	7,031	75,972	65,391	44,699	5,397	104,693	13,762	96,328
Math	0.009	0.028	0.027*	0.009	0.123	0.072***	0.054	0.077***	0.085***	0.066***	0.050	0.015	0.015	0.013
	(0.012)	(0.018)	(0.013)	(0.014)	(0.120)	(0.021)	(0.043)	(0.014)	(0.015)	(0.019)	(0.044)	(0.010)	(0.028)	(0.011)
P-Value of Test	0.375		0.377			0.928			0.440		0.437		0.952	
N	85,931	35,920	58,414	63,437	1,835	21,407	7,914	78,709	72,622	49,229	6,037	115,814	15,494	106,357
Non-Academic Outcomes														
Absence	-0.002*	-0.002	-0.003*	-0.001*	-0.013	-0.004	-0.006	-0.001	-0.005***	-0.001	0.001	-0.002**	-0.001	-0.002**
	(0.001)	(0.002)	(0.002)	(0.001)	(0.010)	(0.004)	(0.005)	(0.001)	(0.002)	(0.001)	(0.003)	(0.001)	(0.003)	(0.001)
P-Value of Test	0.948		0.395			0.338			0.033		0.434		0.779	
N	80,798	33,425	54,790	59,433	1,730	20,351	7,355	78,287	68,700	45,523	5,620	108,603	14,598	99,625
Discipline	0.005*	0.002	-0.003	0.008**	-0.005	0.021	-0.018	0.002	0.005	-0.003	-0.005	0.004	0.004	-0.003
	(0.003)	(0.005)	(0.002)	(0.004)	(0.014)	(0.010)	(0.011)	(0.003)	(0.005)	(0.004)	(0.009)	(0.002)	(0.007)	(0.002)
P-Value of Test	0.568		0.026			0.067			0.185		0.835		0.942	
N	146,811	62,680	100,585	108,906	3,237	38,470	13,603	142,251	127,211	82,280	10,430	199,061	28,529	180,962
Retention	0.001	-0.001	0.000	0.000	0.002**	0.001	0.002	-0.005	-0.001	0.000	-0.001	0.000	-0.001	0.000
	(0.009)	(0.001)	(0.000)	(0.001)	(0.002)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)	(0.007)	(0.001)	(0.001)	(0.001)
P-Value of Test	0.174		0.944			0.763			0.222		0.607		0.556	
N	112,838	47,187	76,792	83,233	2,431	29,352	10,419	108,729	97,370	62,655	7,931	152,094	21,262	138,763

Notes: Coefficient estimates for each group: sex, race, FRPL, ELL, and SWD come from separate regressions where the variable in the row heading is the dependent variable. The coefficient on the dummy variable Threat or Threat plus student type is displayed to show the estimated effect of retention threat for students of all types within a group. All estimates displayed are from models which control for and interacts student type indicator variables with the running variable. The model also controls for year and a student's On-Track status in Grade 3. Effects are estimated within outcome-specific bandwidths determined in Grade 3 that include controls for cohort and On-Track. Robust standard errors clustered on the running variable are shown in parentheses. The p-value of test indicates the statistical significance of a test that student types within the same group are equivalent. The reported Ns correspond to each analytic sample; however, each pair of estimates within a group comes from the same regression. *p< 0.10; **p< 0.05; ***p< 0.01.

Table A9. Heterogeneity in effect of retention threat on Grade 7 outcomes by student characteristics.

	(1) On Track	(2) Not On Track	(3) Female	(4) Male	(5) Asian	(6) Black	(7) Hispanic	(8) White	(9) FRPL	(10) Non-FRPL	(11) ELL	(12) Non-ELL	(13) SWD	(14) Non-SWD
Academic Outcomes														
	0.016	0.056***	0.055***	-0.001	0.041	0.083***	0.062	0.102***	0.110***	0.093***	-0.057	0.030***	-0.002	0.029**
ELA	(0.014)	(0.021)	(0.016)	(0.016)	(0.119)	(0.032)	(0.051)	(0.016)	(0.019)	(0.022)	(0.051)	(0.012)	(0.036)	(0.012)
P-Value of Test	0.114		0.013			0.806			0.556		0.091		0.432	
N	55,025	20,555	36,196	39,384	1,094	12,631	4,686	53,012	44,043	31,537	3,403	72,177	8,986	66,594
	-0.011	0.021	0.016	-0.011	0.209	0.078***	0.081*	0.036**	0.068***	0.030	-0.082	0.005	-0.034	0.005
Math	(0.013)	(0.020)	(0.016)	(0.016)	(0.130)	(0.024)	(0.050)	(0.015)	(0.016)	(0.022)	(0.051)	(0.011)	(0.032)	(0.012)
P-Value of Test	0.187		0.220			0.199			0.171		0.098		0.259	
N	62,542	24,376	41,563	45,355	1,230	15,379	5,397	58,280	50,952	35,966	4,026	82,892	10,654	76,264
Non-Academic Outcomes														
	-0.003	-0.002	-0.004*	-0.001	-0.000	-0.001	-0.009	-0.004**	-0.006**	-0.002	0.000	-0.002	-0.000	-0.003
Absence	(0.002)	(0.003)	(0.002)	(0.002)	(0.008)	(0.005)	(0.007)	(0.002)	(0.003)	(0.002)	(0.006)	(0.002)	(0.004)	(0.002)
P-Value of Test	0.956		0.382			0.784			0.328		0.636		0.578	
N	80,412	33,117	54,477	59,052	1,620	20,269	7,193	77,999	68,711	44,818	5,356	108,173	14,521	99,008
	0.001	-0.006	-0.001	-0.003	0.017	0.004	-0.015	-0.008**	-0.003	-0.013	-0.018	-0.001	0.005	-0.003
Discipline	(0.004)	(0.005)	(0.004)	(0.004)	(0.020)	(0.010)	(0.013)	(0.004)	(0.005)	(0.005)	(0.011)	(0.003)	(0.008)	(0.003)
P-Value of Test	0.222		0.753			0.399			0.160		0.132		0.347	
N	142,424	60,806	97,598	105,632	2,984	36,865	12,947	138,921	123,100	80,130	9,745	193,485	27,632	175,598
	-0.002	-0.001	-0.000	-0.002	-0.004	0.001	0.004	-0.003**	-0.004**	-0.002	0.002	-0.001	0.001	-0.001
Retention	(0.001)	(0.002)	(0.001)	(0.001)	(0.005)	(0.003)	(0.004)	(0.001)	(0.002)	(0.001)	(0.004)	(0.001)	(0.002)	(0.001)
P-Value of Test	0.639		0.485			0.751			0.360		0.359		0.454	
N	109,395	45,776	74,470	80,701	2,237	28,123	9,893	106,140	94,178	60,993	7,392	147,779	20,582	134,589

Notes: Coefficient estimates for each group: sex, race, FRPL, ELL, and SWD come from separate regressions where the variable in the row heading is the dependent variable. The coefficient on the dummy variable Threat or Threat plus student type is displayed to show the estimated effect of retention threat for students of all types within a group. All estimates displayed are from models which control for and interacts student type indicator variables with the running variable. The model also controls for year and a student's On-Track status in Grade 3. Effects are estimated within outcome-specific bandwidths determined in Grade 3 that include controls for cohort and On-Track. Robust standard errors clustered on the running variable are shown in parentheses. The p-value of test indicates the statistical significance of a test that student types within the same group are equivalent. The reported Ns correspond to each analytic sample; however, each pair of estimates within a group comes from the same regression. *p< 0.10; **p< 0.05; ***p< 0.01.

Table A10. Heterogeneity in effect of retention threat on Grade 8 outcomes by student characteristics.

	(1) On Track	(2) Not On Track	(3) Female	(4) Male	(5) Asian	(6) Black	(7) Hispanic	(8) White	(9) FRPL	(10) Non-FRPL	(11) ELL	(12) Non-ELL	(13) SWD	(14) Non-SWD
Academic Outcomes														
ELA	0.004	0.070**	0.018	0.035	-0.110	0.106***	0.131***	0.115***	0.112***	0.096***	0.079	0.021	-0.018	0.029**
	(0.021)	(0.021)	(0.026)	(0.025)	(0.125)	(0.032)	(0.053)	(0.016)	(0.026)	(0.032)	(0.086)	(0.019)	(0.057)	(0.012)
P-Value of Test	0.105		0.634			0.339			0.686		0.509		0.675	
N	54,748	21,693	36,521	39,920	1,156	12,614	4,830	53,658	44,412	32,029	3,642	72,799	9,475	66,966
Math	-0.009	0.041	0.018	-0.006	0.229*	0.078***	0.029	0.054***	0.071***	0.055**	-0.018	0.006	-0.034	-0.005
	(0.015)	(0.021)	(0.017)	(0.017)	(0.139)	(0.026)	(0.051)	(0.016)	(0.017)	(0.025)	(0.057)	(0.013)	(0.032)	(0.013)
P-Value of Test	0.056		0.328			0.464			0.601		0.675		0.307	
N	62,178	25,341	41,783	45,736	1,298	14,543	5,629	56,268	51,000	36,519	4,230	83,289	11,133	76,386
Non-Academic Outcomes														
Absence	-0.007***	-0.007**	-0.006**	-0.007***	-0.013	-0.020	-0.009	-0.006***	-0.013***	-0.005***	-0.002	-0.007***	-0.001	-0.007***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.013)	(0.006)	(0.008)	(0.002)	(0.003)	(0.002)	(0.006)	(0.002)	(0.004)	(0.002)
P-Value of Test	0.974		0.787			0.149			0.029		0.430		0.199	
N	79,185	32,494	53,599	58,080	1,588	19,791	7,069	76,913	67,339	44,340	5,235	106,444	14,266	97,413
Discipline	0.011***	-0.004	0.007	0.003	-0.011	0.007	-0.021	-0.004	0.001	-0.009*	0.005	0.006	-0.001	0.006**
	(0.004)	(0.005)	(0.004)	(0.004)	(0.022)	(0.010)	(0.014)	(0.004)	(0.005)	(0.005)	(0.012)	(0.003)	(0.008)	(0.003)
P-Value of Test	0.033		0.507			0.415			0.159		0.946		0.408	
N	140,150	59,399	95,824	103,725	2,919	35,807	12,653	136,903	120,285	79,264	9,480	190,069	27,139	172,410
Retention	-0.001	0.001	-0.000	-0.001	-0.000	-0.004*	-0.001	0.000	-0.001	-0.002	0.003	-0.001	-0.001	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.001)	(0.002)	(0.001)	(0.003)	(0.001)	(0.002)	(0.001)
P-Value of Test	0.299		0.783			0.328			0.860		0.359		0.707	
N	107,644	44,842	73,166	79,320	2,193	27,363	9,703	104,632	92,181	60,305	7,214	145,272	20,239	132,247

Notes: Coefficient estimates for each group: sex, race, FRPL, ELL, and SWD come from separate regressions where the variable in the row heading is the dependent variable. The coefficient on the dummy variable Threat or Threat plus student type is displayed to show the estimated effect of retention threat for students of all types within a group. All estimates displayed are from models which control for and interacts student type indicator variables with the running variable. The model also controls for year and a student's On-Track status in Grade 3. Effects are estimated within outcome-specific bandwidths determined in Grade 3 that include controls for cohort and On-Track. Robust standard errors clustered on the running variable are shown in parentheses. The p-value of test indicates the statistical significance of a test that student types within the same group are equivalent. The reported Ns correspond to each analytic sample; however, each pair of estimates within a group comes from the same regression. *p< 0.10; **p< 0.05; ***p< 0.01.

Table A11. Estimated effect of retention threat on Grade 4 ELA by school and district characteristics.

	(1) Level of District Financing	(2) Predicted Share Subject to Retention Threat^	(3) Level of District Financing Controlling for Retention Threat^
Panel A: District Finance			
Quintile 5 # Threat	0.044		0.032
	(0.033)		(0.034)
Quintile 4 # Threat	0.041		0.038
	(0.036)		(0.036)
Quintile 3 # Threat	0.008		0.008
	(0.035)		(0.035)
Quintile 2 # Threat	-0.010		-0.006
	(0.038)		(0.038)
Threat	0.005		-0.055
	(0.025)		(0.034)
Panel B: Share Subject to Threat			
Quintile 5 # Threat		0.087***	0.099***
		(0.030)	(0.031)
Quintile 4 # Threat		0.054*	0.059*
		(0.031)	(0.033)
Quintile 3 # Threat		0.023	0.032
		(0.031)	(0.031)
Quintile 2 # Threat		0.054*	0.030
		(0.033)	(0.032)
Threat		-0.015	-0.055
		(0.022)	(0.034)
P-value of test:			Finance; Risk
5 = 4	0.930	0.297	0.859 ; 0.213
5 = 3	0.217	0.038	0.409 ; 0.037
5 = 2	0.097	0.028	0.233 ; 0.034
5 = 1	0.182	0.003	0.335 ; 0.002
4 = 3	0.305	0.365	0.342 ; 0.431
4 = 2	0.148	0.311	0.198 ; 0.403
4 = 1	0.249	0.104	0.288 ; 0.077
3 = 2	0.595	0.909	0.676 ; 0.959
3 = 1	0.809	0.458	0.826 ; 0.306
2 = 1	0.795	0.532	0.865 ; 0.332
All	0.337	0.040	0.572 ; 0.033
5 = 4 + 3 + 2	0.954	0.887	0.925 ; 0.723
4 = 3 = 2	0.323	0.548	0.407 ; 0.655

Notes: Coefficients in each column are from a separate regression using Grade 4 ELA as the dependent variable and our main RD model fully interacted with each quintile and the running variable in a model that includes On-Track and cohort fixed effects. The coefficients shown are the interaction of Threat and a school's quintile. P-

values indicate t-tests of differences in outcomes across coefficients. *p< 0.10; **p< 0.05; ***p< 0.01. Robust standard errors are clustered at the school level.

Table A12. Estimated effect of retention threat on Grade 5 ELA by school and district characteristics.

	(1) Level of District Financing	(2) Predicted Share Subject to Retention Threat^	(3) Level of District Financing Controlling for Retention Threat^
Panel A: District Finance			
Quintile 5 # Threat	0.049		0.037
	(0.034)		(0.035)
Quintile 4 # Threat	0.075**		0.076**
	(0.034)		(0.034)
Quintile 3 # Threat	0.038		0.039
	(0.034)		(0.035)
Quintile 2 # Threat	0.016		0.022
	(0.034)		(0.034)
Threat	-0.011		-0.072
	(0.027)		(0.035)
Panel B: Share Subject to Threat			
Quintile 5 # Threat		0.046	0.064*
		(0.033)	(0.034)
Quintile 4 # Threat		0.043	0.046
		(0.032)	(0.032)
Quintile 3 # Threat		-0.008	-0.004
		(0.030)	(0.031)
Quintile 2 # Threat		0.009	0.015
		(0.030)	(0.031)
Threat		0.009	-0.072
		(0.023)	(0.035)
P-value of test:			Finance; Risk
5 = 4	0.661	0.390	0.057 ; 0.562
5 = 3	0.196	0.848	0.122 ; 0.030
5 = 2	0.051	0.001	0.043 ; 0.119
5 = 1	0.000	0.008	0.770 ; 0.057
4 = 3	0.105	0.241	0.331 ; 0.093
4 = 2	0.026	0.005	0.013 ; 0.299
4 = 1	0.000	0.044	0.882 ; 0.153
3 = 2	0.484	0.000	0.001 ; 0.505
3 = 1	0.008	0.002	0.350 ; 0.901
2 = 1	0.041	0.396	0.058 ; 0.631
All	0.001	0.000	0.020 ; 0.157
5 = 4 + 3 + 2	0.012	0.533	0.757 ; 0.906
4 = 3 = 2	0.075	0.000	0.003 ; 0.242

Notes: Coefficients in each column are from a separate regression using Grade 5 ELA as the dependent variable and our main RD model fully interacted with each quintile and the running variable in a model that includes On-Track and cohort fixed effects. The coefficients shown are the interaction of Threat and a school's quintile. P-

values indicate t-tests of differences in outcomes across coefficients. *p< 0.10; **p< 0.05; ***p< 0.01. Robust standard errors are clustered at the school level.

Table A13. Estimated effect of retention threat on Grade 7 ELA by school and district characteristics.

	(1) Level of District Financing	(2) Predicted Share Subject to Retention Threat^	(3) Level of District Financing Controlling for Retention Threat^
Panel A: District Finance			
Quintile 5 # Threat	0.068*		0.058
	(0.029)		(0.037)
Quintile 4 # Threat	0.136***		0.123***
	(0.037)		(0.038)
Quintile 3 # Threat	0.084**		0.077**
	(0.037)		(0.037)
Quintile 2 # Threat	0.107***		0.021
	(0.038)		(0.034)
Threat	-0.059		-0.130
	(0.028)		(0.037)
Panel B: Share Subject to Threat			
Quintile 5 # Threat		-0.009	0.016
		(0.036)	(0.037)
Quintile 4 # Threat		0.078**	0.084**
		(0.023)	(0.038)
Quintile 3 # Threat		0.002	-0.002
		(0.201)	(0.034)
Quintile 2 # Threat		0.023	0.021
		(0.034)	(0.034)
Threat		0.006	-0.130
		(0.025)	(0.037)
P-value of test:			Finance; Risk
5 = 4	0.045	0.022	0.057 ; 0.072
5 = 3	0.629	0.740	0.598 ; 0.615
5 = 2	0.246	0.360	0.367 ; 0.894
5 = 1	0.068	0.810	0.176 ; 0.664
4 = 3	0.135	0.037	0.118 ; 0.020
4 = 2	0.410	0.132	0.186 ; 0.092
4 = 1	0.000	0.036	0.613 ; 0.027
3 = 2	0.505	0.535	0.001 ; 0.499
3 = 1	0.025	0.007	0.409 ; 0.959
2 = 1	0.004	0.494	0.041 ; 0.549
All	0.006	0.147	0.005 ; 0.156
5 = 4 + 3 + 2	0.001	0.110	0.012 ; 0.218
4 = 3 = 2	0.326	0.000	0.410 ; 0.063

Notes: Coefficients in each column are from a separate regression using Grade 7 ELA as the dependent variable and our main RD model fully interacted with each quintile and the running variable in a model that includes On-Track and cohort fixed effects. The coefficients shown are the interaction of Threat and a school's quintile. P-

values indicate t-tests of differences in outcomes across coefficients. *p< 0.10; **p< 0.05; ***p< 0.01. Robust standard errors are clustered at the school level.

Table A14. Estimated effect of retention threat on Grade 8 ELA by school and district characteristics.

	(1) Level of District Financing	(2) Predicted Share Subject to Retention Threat^	(3) Level of District Financing Controlling for Retention Threat^
Panel A: District Finance			
Quintile 5 # Threat	0.087		0.031
	(0.061)		(0.061)
Quintile 4 # Threat	0.055		0.026
	(0.063)		(0.061)
Quintile 3 # Threat	0.032		0.017
	(0.064)		(0.061)
Quintile 2 # Threat	0.033		0.032
	(0.065)		(0.063)
Threat	-0.027		-0.054
	(0.082)		(0.066)
Panel B: Share Subject to Threat			
Quintile 5 # Threat		0.032	0.062
		(0.058)	(0.059)
Quintile 4 # Threat		0.145**	0.162***
		(0.061)	(0.062)
Quintile 3 # Threat		-0.063	-0.053
		(0.058)	(0.058)
Quintile 2 # Threat		-0.128**	-0.118**
		(0.057)	(0.058)
Threat		0.025	-0.054
		(0.044)	(0.066)
P-value of test:			Finance; Risk
5 = 4	0.661	0.044	0.915 ; 0.076
5 = 3	0.196	0.066	0.785 ; 0.033
5 = 2	0.051	0.002	0.989 ; 0.001
5 = 1	0.000	0.575	0.608 ; 0.297
4 = 3	0.105	0.000	0.871 ; 0.000
4 = 2	0.026	0.000	0.912 ; 0.000
4 = 1	0.000	0.018	0.674 ; 0.009
3 = 2	0.484	0.209	0.792 ; 0.209
3 = 1	0.008	0.274	0.781 ; 0.365
2 = 1	0.041	0.026	0.615 ; 0.042
All	0.001	0.000	0.986 ; 0.000
5 = 4 + 3 + 2	0.012	0.499	0.722 ; 0.543
4 = 3 = 2	0.075	0.000	0.965 ; 0.000

Notes: Coefficients in each column are from a separate regression using Grade 8 ELA as the dependent variable and our main RD model fully interacted with each quintile and the running variable in a model that includes On-Track and cohort fixed effects. The coefficients shown are the interaction of Threat and a school's quintile. P-

values indicate t-tests of differences in outcomes across coefficients. *p< 0.10; **p< 0.05; ***p< 0.01. Robust standard errors are clustered at the school level.

Appendix Table A15. Estimated effect of retention threat on probability of assignment to reading interventions in Grade 3.

	(1)	(2)	(3)	(4)	(5)
	Any Int.	Extra Time	Small Group	Reading Int.	Writing Int.
Threat (Fall)	0.065***	0.051***	0.036***	0.018***	0.002
	(0.007)	(0.007)	(0.007)	(0.005)	(0.003)
N	123,079	123,079	123,079	123,079	123,079
Avg. Pr(assign) control group	0.485	0.265	0.306	0.105	0.027

Note: Coefficient estimates in each column are from separate regressions. The variable in the column heading is the dependent variable. All estimates control for the running variable, cohort fixed effects, and a student's On-Track status in Grade 3. Effects are estimated within the bandwidth for ELA determined in Grade 3 which was estimating including controls for cohort and On-Track. Robust standard errors clustered on the running variable are shown in parentheses. *p< 0.10; **p< 0.05; ***p< 0.01.

APPENDIX B: DECREASED EFFORT AMONG PROMOTED STUDENTS?

Increased achievement and effort among those facing the threat of retention throughout Grade 3 or decreased achievement and effort among those who know they will be promoted to Grade 4 could drive observed differences in spring ELA achievement. In this appendix we present an interrupted time-series analysis which suggests increased effort and achievement among students subject to retention accounts for our effect estimates.

To implement this analysis, we convert all student third grade ELA test scores from the years 2009 through 2015 into standardized scores with a mean of zero and standard deviation (SD) of one using the mean and standard deviations measured in the original scaled score across all six of these years. The ELA test and scale used in Ohio changes after 2015, and so those subsequent years are excluded from this analysis. Next, we fit a trend line using the years before the adoption of the retention mandate (2009-2013) and the following regression model.

$$ELA_{it} = \beta_0 + \beta_1 T_t + X_{it} B + \varepsilon_{it}$$

Where ELA is the standardized ELA test score, T is a year counter beginning at 0 in 2009 and running through 4 in 2013, and X is a set of student level covariates measured prior to third grade. If each of the X variables are centered on 0, then β_0 and β_1 define the intercept and slope of the pre-treatment trend in mean test scores, adjusted for the covariates.

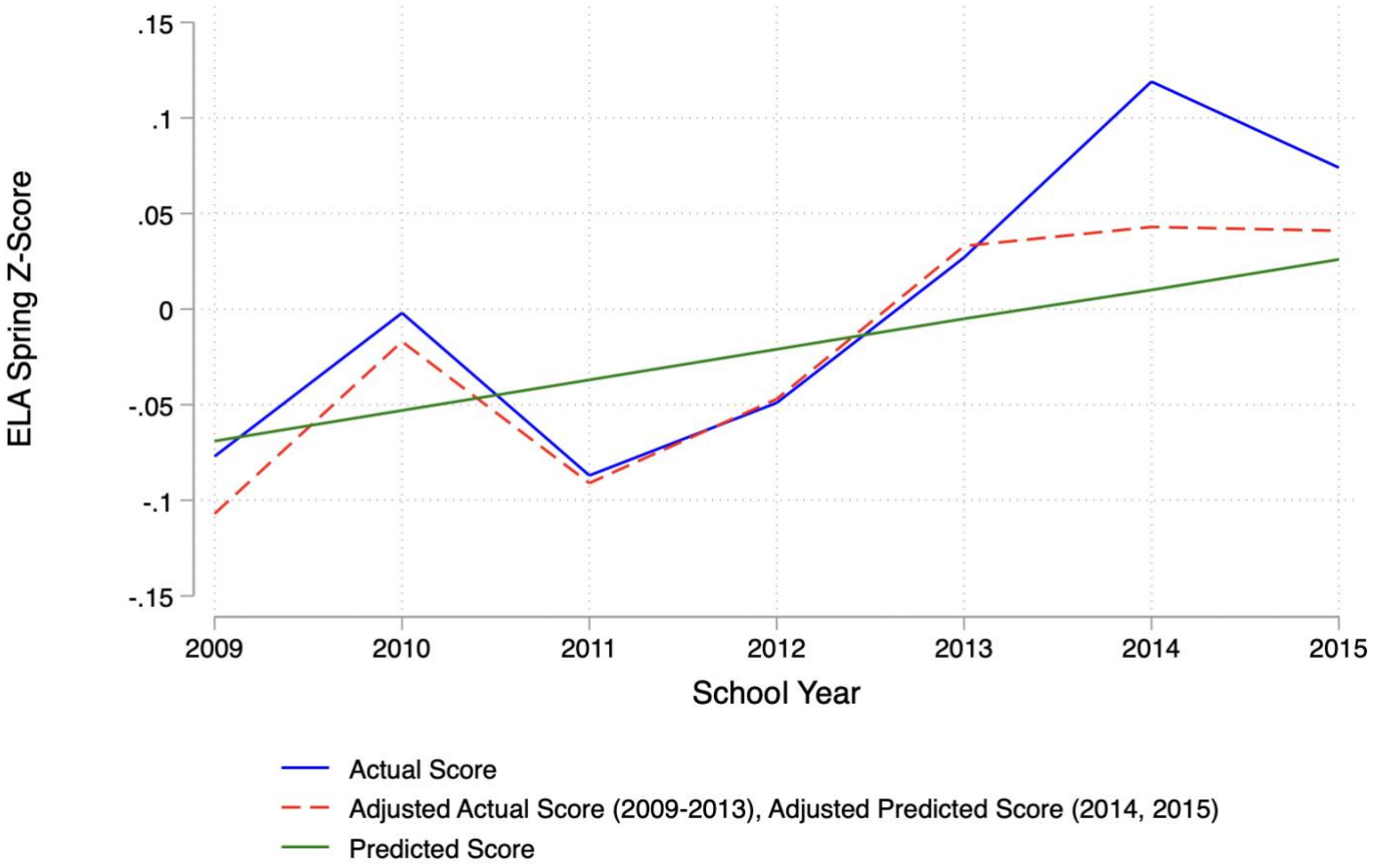
We then project this trend line into the 2014 and 2015 school years to get a predicted mean test score for those years. If the only reason for a change in trends beginning in 2014 was the effect of the retention mandate on the achievement of students subject to retention threat, then we would expect the actual mean test scores to exceed the predicted scores by our estimate

of the effect of retention threat times the number of students subject to retention threat.[31] For 2014, this value is 0.094 x 0.314 = 0.030 and for 2015 it is 0.038 x 0.339 = 0.013. If the actual mean scores for 2014 and 2015 exceed the predicted means by these amounts or more, then that suggests that our estimated impacts reflect positive increases in effort and achievement of students subject to retention threat. If actual mean scores are less than the predicted mean scores plus these estimated impacts, then that suggests that our estimated impacts at least partially reflect decreased effort and achievement among those not subject retention threat. Of course, this exercise assumes that nothing else effects deviations from pre-treatment trends in 2014 and 2015 other than the effects of retention threat, which is a strong assumption. Thus, the results of this analysis are only suggestive.

Figure B1 presents the results. The straight solid green line on the graph represents the estimated pre-treatment trend in third grade ELA test scores and is projected using data from 2009-2013 for 2014 and 2015. The solid blue line represents the actual ELA z-scores. The dashed red line prior to 2014 represents actual scores adjusted by the average residuals in each year. In years 2014 and 2015, the dashed red line shows predicted values plus the estimated impacts of retention threat. The actual, observed mean score is 0.105 SD and 0.046 SD above predicted mean values, and 0.076 and 0.033 SD above predicted mean values plus the estimated effects of retention in 2014 and 2015, respectively. Both of these amounts exceed the estimated impacts of retention threat based on our regression discontinuity analysis, 0.030 and 0.013, and thus, suggests that the RD estimates reflect increased achievement among those subject to retention threat rather than decreased achievement among those not subject to the retention threat.

[31] This calculation assumes that the local average treatment effect we estimate equals the average effect of the treatment on the treated.

Figure B1. Trends in Spring ELA achievement, 2009-2015.



Notes: Figure B1 shows an interrupted time-series analysis of Grade 3 ELA achievement. All scores from 2009–2015 are standardized using the pooled mean and standard deviation across years. The solid green line shows the pre-policy trend estimated using 2009–2013 data and projected forward. The solid blue line represents observed mean ELA z-scores. The dashed red line shows covariate-adjusted mean scores (2009–2013) and projected values plus estimated treatment effects in 2014–2015.