Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 25-1188

The Impacts of Grade Retention Policy With Minimal Retention

Jordan S. Berne
University of Michigan

Brian A. Jacob
University of Michigan

Christina Weiland
University of Michigan

Katharine O. Strunk
University of Pennsylvania

State laws that mandate in-grade retention for struggling readers are widespread in the U.S., covering 34% of public-school third graders in 2023-24. This study investigates the impacts of Michigan's third-grade reading law on subsequent test scores and school progress outcomes for the 2020-21 and 2021-22 third-grade cohorts. Using a regression discontinuity (RD) design, we find that being flagged for retention raises students' reading scores in the next school year by 0.045 standard deviations (SD)—a modest but meaningful impact. Because being flagged increases the likelihood of actually being retained by only 3.4 percentage points, the implied effect of retention itself under standard "fuzzy" RD assumptions would be an implausibly large 1.3 SD. This result suggests flagging may affect outcomes via mechanisms other than just retention, a violation of the exclusion restriction. Indeed, we estimate similar effects even in districts that retain no students. Survey evidence suggests flagged students receive more intensive reading support even if they are not retained. Our findings suggest retention may be a much less important component of literacy reforms than previously understood. Finally, given the similarity between Michigan's reading law and those in other states, our findings raise concerns about potential bias in previously estimated retention effects.

VERSION: May 2025

Suggested citation: Berne, Jordan S., Brian A. Jacob, Christina Weiland, and Katharine O. Strunk. (2025). The Impacts of Grade Retention Policy With Minimal Retention. (EdWorkingPaper: 25 -1188). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/tqrb-yh12

The Impacts of Grade Retention Policy With Minimal Retention*

Jordan S. Berne[1], Brian A. Jacob[1,2], Christina Weiland[1], and Katharine O. Strunk[3]

[1]University of Michigan
[2]National Bureau of Economic Research
[3]University of Pennsylvania

April 28, 2025

Abstract

State laws that mandate in-grade retention for struggling readers are widespread in the U.S., covering 34% of public-school third graders in 2023-24. This study investigates the impacts of Michigan's third-grade reading law on subsequent test scores and school progress outcomes for the 2020-21 and 2021-22 third-grade cohorts. Using a regression discontinuity (RD) design, we find that being flagged for retention raises students' reading scores in the next school year by 0.045 standard deviations (SD)—a modest but meaningful impact. Because being flagged increases the likelihood of actually being retained by only 3.4 percentage points, the implied effect of retention itself under standard "fuzzy" RD assumptions would be an implausibly large 1.3 SD. This result suggests flagging may affect outcomes via mechanisms other than just retention, a violation of the exclusion restriction. Indeed, we estimate similar effects even in districts that retain no students. Survey evidence suggests flagged students receive more intensive reading support even if they are not retained. Our findings suggest retention may be a much less important component of literacy reforms than previously understood. Finally, given the similarity between Michigan's reading law and those in other states, our findings raise concerns about potential bias in previously estimated retention effects.

Keywords: Early literacy, grade retention, Read by Grade Three, Michigan

*We thank MDE and CEPI for their partnership, especially Thomas Howell, Richard Lower, and Emily Laidlaw. We also thank members of the broader Read by Grade Three research team, especially Tara Kilbride, Emily Mohr, Josep Nadal-Fernandez, and Tanya Wright. Finally, we thank Katie Laursen and Thomas Staines for excellent research assistance; Nicole Wagner Lam and Jasmina Camo-Biogradlija for assistance with data agreements and project management; Kyle Kwaiser and Thomas Staines for data management; Ozkan Eren and Nikhil Rao for useful comments; and seminar/conference participants at the University of Michigan, AEFP, and APPAM. This research used data structured and maintained by the MERI-Michigan Education Data Center (MEDC). MEDC data are modified for analysis purposes using rules governed by MEDC and are not identical to those data collected and maintained by the Michigan Department of Education and/or Michigan's Center for Educational Performance and Information. This research was funded by the U.S. Department of Education's Institute of Education Sciences grants R305H190004 and R305B200011. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not reflect the views of any other entity.

I Introduction

Strong early literacy skills help set children on a path to school and life success. However, the most recent NAEP results indicate that only 31 percent of fourth graders are proficient in reading, with large disparities between children from historically marginalized groups and their peers (National Assessment of Educational Progress, 2025). Educators have struggled to find effective ways to improve literacy at scale for decades (Gamse et al., 2008; Jacob, 2017; Snow et al., 1998), including via state-level comprehensive literacy reforms. One common component of these reforms—grade retention for those scoring below a state-determined threshold—has been hotly debated, with mixed empirical evidence on short- and long-term student outcomes (e.g., Jacob and Lefgren, 2009; Martorell and Mariano, 2018; Özek, 2015; Schwerdt et al., 2017).

As "science of reading" reforms gain momentum nationally (Schwartz, 2022) and districts struggle to serve students who lost ground during the COVID-19 pandemic, grade retention policies have become more common. As of 2025, 14 states (and formerly Michigan) have a law in effect that ties third-grade promotion to reading proficiency as measured by state standardized tests (ExcelinEd, 2024). Four other states, including Maryland and West Virginia, recently passed retention laws that will take effect in the next few years. As we describe in greater detail below, recent retention laws differ in important ways from those common in earlier decades. Specifically, they include more liberal use of "good cause exemptions" that reduce the proportion of students who are retained and they require schools to provide struggling readers with more support services than past policies.

Rigorous studies of these new literacy reforms are just emerging. While they promise to expand our understanding of grade retention policies, they raise difficult interpretation questions. Not only do the new policies differ from the old policies, but disruptions in educational systems during and after the COVID-19 pandemic mean that the new policies are being implemented in a notably different context. Our study contributes to this new wave of evidence.

We use statewide administrative data from Michigan to study two recent cohorts that faced a test-based retention policy. Children who scored below a predetermined threshold on the state literacy exam were flagged as eligible for retention and their families were informed by letter from the state. We leverage a regression discontinuity (RD) approach to study the policy's impact. In essence, this approach compares children who scored just below the test score cutoff to otherwise identical children who scored just above it.

Our primary focus is on the impact of being *flagged* for retention, commonly called an "Intent-to-Treat" or ITT effect. This impact captures all interventions and behavioral responses that follow flagging. We find that

being flagged increased the likelihood of actually being retained by only 3.4 percentage points. Nevertheless, being flagged raised students' reading test scores in the following year by 0.045 standard deviations (SD), a modest but educationally meaningful impact. Results are qualitatively similar for most subgroups, but there is some indication that historically advantaged subgroups benefited more from the policy, especially by becoming more likely to have an individualized education program (IEP) in the following year.

Policymakers and educators might also be interested in the impact of *actually being retained*, commonly referred to as a "Local Average Treatment Effect" or LATE. Indeed, previous RD papers on grade retention have primarily focused on LATE estimates. However, under standard assumptions, our estimates would imply implausibly large (1.3 SD) effects of grade retention on reading scores in Michigan. (For reference, normative growth in reading between third and fourth grade is only around 0.36 SD; Hill et al. 2008). This result led us to uncover a previously overlooked element of most modern state reading laws that could substantially bias LATE estimates—in Michigan and elsewhere.

In particular, most grade retention laws require that students flagged for retention receive intensive reading support in the following year *regardless of whether they are retained*. These additional supports—as well as behavioral responses from parents and teachers—would undermine the standard LATE assumption that being flagged only affects outcomes via its effect on retention (i.e., the "exclusion restriction"). Using novel survey data we collected from principals around Michigan, we show that schools may have provided reading support (e.g., summer reading programs, high-dosage tutoring, etc.) differentially at the RD cutoff. Moreover, we find positive and statistically significant impacts on reading scores even in districts where *no children were retained*, further underscoring a violation of the exclusion restriction.

These findings make several valuable contributions. First, the nature of the implementation of Michigan's law, combined with our principal survey data, allows us to explore a previously overlooked issue in research on grade retention. We document an important limitation of the most widely used method for studying test-based grade retention. Given the similar structure of Michigan's reading reform and those in other states, our results suggest that retention effect estimates in many previous studies may be misleading.

Second, for policymakers, our paper provides causal estimates of an important reform that affected thousands of children in Michigan. Around the country, 34 percent of public-school third graders live in states with a similar retention law.[1] We also provide suggestive evidence on underappreciated mechanisms through which these retention policies may operate. This contribution is especially relevant as Michigan recently repealed the retention component of its law and as other states (e.g., Maryland) design laws with

[1] Authors' calculation based on the 18 states with third-grade retention policies and state-level 2023-24 enrollment data from the U.S. Department of Education's Common Core of Data (ExcelinEd, 2024).

less emphasis on retention.

II Grade Retention Laws in Michigan and Other States

In an effort to address concerns about reading proficiency among young students, Michigan implemented Public Act 306, informally known as the Read by Grade Three (RBG3) Law, in 2016. The law requires schools to provide early assessment and intervention for struggling readers from kindergarten through third grade using a multi-tiered system of support (MTSS). Specifically, teachers are required to develop individualized reading improvement plans for students who exhibit reading deficiencies, which include strategies such as evidence-based reading instruction, frequent monitoring, and small group or one-on-one tutoring. A central but highly controversial component of the law was a provision that mandated the retention of third-grade students who scored more than one grade level behind on the state's standardized English Language Arts (ELA) test, which corresponds to a score of 1252 or below on the state assessment.

The state is required to notify parents of students identified for retention. Districts must then determine whether each student will be retained or is eligible for a "good cause" exemption and promotion to fourth grade. In certain circumstances, students are eligible for a good cause exemption if they are English learners (EL), have an individualized education program (IEP), have a 504 plan, have been previously retained, or have been enrolled in public school for less than two years. Exemptions can also be granted on the basis of parent requests and/or a portfolio of student work that demonstrates mastery of the necessary skills (Strunk et al., 2022).

Importantly, districts are required to provide additional literacy support to students eligible for retention *even if they are promoted* to fourth grade (Kilbride and Walker, 2023).[2] Michigan's law is not unique in this respect. Most states mandate intensive reading instruction in the next grade for retention-eligible students who are promoted. For example, Mississippi's legislation states, "A student who is promoted to Fourth Grade with a good cause exemption shall be provided an individual reading plan as described in Section 37-177-1(2), which outlines intensive reading instruction and intervention..." (Mississippi State Senate, 2016). Indiana requires schools to complete a Good Cause Exemption Plan for students who score below the mandated cutoff, outlining the support they will receive to facilitate progress toward reading proficiency (Indiana Department of Education, 2021; Indiana State Board of Education, 2021). Maryland and West Virginia have newly passed similar provisions in their statewide literacy reform laws (Maryland

[2]The state recommends that districts provide these same supports to students who score between 1253 and 1271, a group that is substantially below grade level but is not eligible for retention. Previous research suggests that most districts provide these students with at least some services (Kilbride and Walker, 2023).

State Department of Education, 2024; West Virginia House of Delegates, 2023). Retained or not, students scoring below the threshold must receive intensive support in the following school year.

Interestingly, very few students were retained under Michigan's policy. Local educators had negative views about Michigan's retention policy, and most district superintendents indicated that they planned to utilize good cause exemptions heavily (Strunk et al., 2021). See Appendix Figure A1 for one example. As a result, only 6.8% of third graders in the 2020-21 and 2021-22 cohorts who scored below a 1253 on the state assessment were actually retained.[3] Ultimately, the Michigan legislature amended the original RBG3 law, removing the retention component for cohorts beginning in 2023-24.

Such opposition to grade retention is common outside of Michigan too. For example, in 2023-24, the first year of Tennessee's third-grade retention policy, 60 percent of third graders were flagged for retention but only 1.2 percent were ultimately retained (Wegner, 2024). A study of Mississippi's retention policy for third graders in 2014-15 found that only about 15 percent of students scoring below the cutoff were retained. Among students *just* below the cutoff, this figure drops below 10 percent (Mumma and Winters, 2023). In Indiana, frustrated by what was perceived as unreasonably high use of exemptions, legislators recently passed a law tightening up the retention policy (Appleton, 2024).

III Prior Literature

Starting with Jacob and Lefgren (2004), many researchers have used regression discontinuity (RD) approaches to estimate the impact of grade retention. In general, these studies have shown large, positive short-run effects on achievement (see Schwerdt et al. 2017 for Florida, Jacob and Lefgren 2004 for Chicago, Eren et al. 2017 for Louisiana, Mariano et al. 2024 for New York, Mumma and Winters 2023 for Mississippi, and Zhong 2024 for Texas). The evidence on longer-term effects is mixed, including in many of these same contexts (see, for example, Jacob and Lefgren 2009; Schwerdt et al. 2017; Zhong 2024). Evidence of retention's impact on students' social-emotional and behavioral outcomes is also mixed. For example, there is evidence of positive effects on students' sense of belonging in a school in early elementary school (McCombs et al., 2009; Wu et al., 2010); null effects on school suspension in third through eighth grade (Mariano and Martorell, 2013); increases in school discipline incidents in fifth grade (Özek, 2015); and null effects on student absences (Mumma and Winters, 2023; Schwerdt et al., 2017).

Several interesting patterns emerge from prior research. First, there is some indication that retention in grades three and four is more beneficial than retention in later grades. For example, several studies find that

[3]Including students who scored above the threshold, only 0.8% of all test-takers were retained.

retention in these grades has large positive impacts on student achievement through middle school (Mumma and Winters, 2023; Schwerdt et al., 2017). In contrast, there is some evidence that retention in middle school reduces the likelihood of high school graduation (Eren et al., 2017; Jacob and Lefgren, 2009; Mariano et al., 2024). This may be because of motivation and stigma, particularly as peer groups and peer comparisons become more salient in the middle school years (Ryan, 2001; Ryan and Patrick, 2001). Second, most studies find larger achievement effects when comparing students' outcomes in the same grade rather than at the same age (Hwang and Koedel, 2023; Schwerdt et al., 2017).[4]

Finally, there may be larger impacts of retention for some historically marginalized groups than their peers. For example, effects of retention in Florida on disciplinary incidents and suspensions were concentrated among economically disadvantaged students, Black students, and male students (Özek, 2015). In Mississippi, the state retention policy led to large positive effects for Black and Hispanic students and not White students on ELA test scores (Mumma and Winters, 2023). A study in Florida, however, found little evidence of heterogeneity in impacts on students' ELA and math scores in grades 4-6 by student race/ethnicity or gender (Winters and Greene, 2012). Similarly, a study of Indiana's policy found little evidence of heterogeneity in impacts on students' ELA scores, math scores, discipline, or absences in grades 4-7 by student race/ethnicity, gender, free-reduced lunch status, or prior discipline (Hwang and Koedel, 2023).

An important limitation of virtually all quantitative grade retention analyses is that researchers have lacked data on the mechanisms underlying impacts. For example, students who are retained often receive additional school-based services the following year. It also seems likely that parents respond to retention in ways that would influence children's learning (e.g., monitoring performance more carefully or arranging for additional support outside of school). Virtually all studies have been unable to explore these types of mechanisms. As a result, though researchers commonly mention that retention was part of a package of supports, impacts have been attributed primarily to retention. We return to this issue in our Mechanisms and Discussion sections.

IV Data

IV.A Michigan Administrative Datasets

Our analysis relies on several individual-level administrative datasets from the Michigan Department of Education (MDE). The first is a longitudinal dataset on the universe of students in Michigan public

[4]This is not particularly surprising given that retained students in same-grade comparisons not only have an additional year to learn grade-level skills, but are also older than promoted students when they take the exam.

schools. In each year, we observe students' grade level, the school(s) they attend, and a host of student- and school-level demographic and socio-economic characteristics. To measure retention, we construct a binary variable that equals one if a student enrolls in third grade the year after their initial third grade enrollment.[5] We also create measures of school and district mobility by comparing the school (district) in which a third grader took the state assessment to the school (district) they were enrolled in the next fall.[6]

The second dataset contains scores on the Michigan state standardized assessment, called M-STEP, for every student who takes the exam. Importantly, the M-STEP is designed so that scores are comparable across time (horizontally scaled) *and* across grade level (vertically scaled).[7] These scaling properties allow us to compare scores between retained and non-retained students when they are in the same grade *and* when they are the same age.

Lastly, we use data from MDE that contains RBG3-specific information. For all children who take the third-grade M-STEP, we observe which students are deemed "eligible" for retention, which students are granted a good cause exemption, and the reason(s) for each exemption. Districts sent formal letters to the guardians of all students who were eligible for retention, which we also observe in the data.

IV.B Sample

Our sample consists of first-time third graders in the 2020-21 and 2021-22 school years, which we will henceforth refer to as the 2021 and 2022 cohorts. These were the first two cohorts subject to retention under Michigan's RBG3 law.[8]

[5] A very small number of students are coded in the data as being in grades that are unlikely to follow third grade. We drop 4 students from the 2021 cohort who are coded as being in second grade and 12 students from the 2022 cohort who are coded as being in second, sixth, or seventh grade.

[6] When the school (district) in which a student took the third-grade M-STEP is missing, we use the school (district) the student was enrolled in during the spring semester of third grade. When a student is enrolled in more than one school (district), whether in the spring of third grade or the fall of the following year, we use the school (district) in which they were enrolled for the longest period.

[7] In practice, the scores reported to students and researchers are transformed to remove the vertical scaling. We manually undo this transformation for our analysis, putting scores back on a vertical scale. Then, we standardize the scores using the mean and standard deviation from third graders in 2018-19 as reported by the exam's creator (Smarter Balanced).

[8] The original law stipulated retention to begin for third graders in 2019-20, but this was canceled due to the COVID-19 pandemic. We are not able to estimate impacts for the 2023 cohort because MDE did not keep RBG3-specific information for these students, in part because the state legislature had already decided to end retention starting with the 2024 cohort. It is possible that the repeal had anticipatory effects on the 2023 cohort, but this remains unclear.

Table 1: Summary Statistics for Third Graders in 2021 and 2022

	All First-Time Third-Graders (1)	Analysis Sample (2)	Below Cutoff Within 15 Pts (3)	Above Cutoff Within 15 Pts (4)
2022 cohort	0.496	0.571	0.599	0.576
Female	0.487	0.491	0.443	0.447
White	0.631	0.672	0.426	0.497
Black	0.224	0.189	0.439	0.358
Hispanic	0.086	0.081	0.104	0.109
Asian	0.047	0.045	0.019	0.020
NA, NH, or PI	0.013	0.013	0.012	0.015
Economically disadvantaged	0.564	0.541	0.847	0.787
Has an IEP	0.157	0.145	0.282	0.269
English learner (EL)	0.091	0.086	0.122	0.123
Previously retained	0.031	0.027	0.059	0.052
Same district 2+ years	0.100	0.096	0.130	0.120
Math M-STEP score	-	1,292.7	1254.7	1263.5
ELA M-STEP score	-	1,292.7	1247.3	1260.6
ELA M-STEP score below 1253	-	0.054	1.000	0.000
Retained in third grade	0.011	0.008	0.065	0.014
Third-graders in district-year	530	464	690	541
City	0.252	0.214	0.385	0.317
Rural	0.172	0.189	0.142	0.167
Suburb	0.456	0.467	0.378	0.401
Town	0.119	0.130	0.096	0.115
Charter school	0.123	0.116	0.202	0.189
Neighborhood BA+ share	0.169	0.172	0.116	0.129
Neighborhood median HH income	$60,009	$61,318	$44,302	$48,260
Districts	822	803	702	772
District×years	1,630	1,571	1,208	1,468
Students	208,279	168,443	7,612	23,673

Notes: The sample in column 1 includes all first-time third graders in the 2020-21 and 2021-22 school years. Column 2 limits the sample to those with valid ELA M-STEP scores. Columns 3 and 4 limit the sample to those with scores near the retention cutoff. "NA, NH, or PI" stands for Native American, Native Hawaiian, or Pacific Islander. An IEP (individualized education program) is a document that outlines the programs and services a student with disabilities will receive to help them realize their educational goals, sometimes referred to as "special education." Neighborhood characteristics are calculated at the Census block group level. Income is converted to real 2015 dollars. For a wider set of summary statistics, see Appendix Table A1.

Table 1 shows summary statistics for these cohorts, beginning in column 1 with the full set of students. Column 2 narrows the sample to students with valid third-grade M-STEP scores, which comprises our analytical sample. Only 69% of students in the 2021 cohort have valid third-grade M-STEP scores due to COVID-related school disruptions. In contrast, over 93% of students in the 2022 cohort have valid scores.[9] Other differences between columns 1 and 2 stem from some students from certain demographic groups, such as English learners and those with an IEP, being excused from taking the regular M-STEP assessment.

Among all students in column 2, ELA M-STEP scores range from 1203 to 1357, with a mean of 1293 and a standard deviation of 26.[10] The 1253 retention cutoff is located around the 5th percentile.

Columns 3 and 4 focus on students who scored close to the retention cutoff. As we explain next, this is the group that drives our analytical results. These students are substantially more likely to be Black, economically disadvantaged, English learners, and to have an IEP. Consistent with these demographics, they are also more likely to live in lower-income and urban areas and attend charter schools. Importantly, students in columns 3 and 4 appear quite similar in many characteristics, capturing the intuition underlying our methodological approach.

V Empirical Strategy

Students flagged as eligible for retention differ from their peers on a variety of observable (and likely unobservable) characteristics, casting doubt on analytical methods that rely on a simple comparison of flagged versus non-flagged students (or retained versus promoted students). Instead, we utilize a regression discontinuity (RD) design that focuses on students who score very close to the retention cutoff. Intuitively, students who score *just* above and *just* below the cutoff should be equivalent in terms of pre-intervention characteristics, on average, in all ways other than their eligibility for retention.

V.A Estimation Approach

For student i in cohort c, we estimate the reduced form, or Intent-to-Treat (ITT), effect of the policy using the following equation:

$$y_i = \beta_0 + \beta_1^{ITT}\text{Below-Cutoff}_i + f(\text{MSTEP}_i, \text{Below-Cutoff}_i) + \mathbf{B}\mathbf{X}_{ic} + \varepsilon_{ic}, \quad (1)$$

[9] Appendix Table B1 shows summary statistics separately by cohort. Appendix Table B2 shows that our findings are robust to dropping the 2021 cohort.

[10] Throughout this paper, we use preliminary ELA M-STEP scores—which do not account for hand grading of short answers and essays—as our measure of ELA scores for first-time third graders. We do so because preliminary scores are what determine retention eligibility under the RBG3 law. Conversely, we use final ELA M-STEP scores as the outcome variable when estimating impacts on achievement.

where Below-Cutoff is a binary variable that equals one if student i scored below 1253 on the third-grade ELA M-STEP. The parameter of interest is β_1^{ITT}, which captures the causal impact of being flagged for retention for students with scores near the 1253 cutoff. The f function denotes a flexible relationship between the continuous M-STEP score and the outcome y_{it}, which may vary based on the value of Below-Cutoff. To increase statistical precision, we control for the following baseline characteristics ($\mathbf{X}_{ic}$): third-grade cohort, gender, race, eligibility for subsidized school meals, having an IEP, and being an English learner.

Estimating this model requires the selection of a "bandwidth," which in our case refers to the range of baseline M-STEP scores included in the estimation sample. A smaller bandwidth reduces bias by only including students with scores near the cutoff. A larger bandwidth improves precision by including additional students (further away from the cutoff). For our primary specification, we select bandwidths using the optimization procedure developed by Calonico et al. (2018, 2020, 2014, 2019). We use a local linear approach with a triangular kernel, and we conduct inference using the bias-corrected method also developed by Calonico et al., with clustering on the running variable. To test the sensitivity of our results, we re-estimate the model using a variety of alternative specifications. Our results are highly robust (see Section VI.B).

In addition to ITT effects, we are also interested in the effects of retention itself. In Michigan as in other states, retention outcomes fit into a "fuzzy" RD framework, meaning some students below the cutoff may not be retained and some students above it may be retained. In such cases, under certain assumptions, one can estimate the impact of retention itself using the binary indicator Below-Cutoff as an instrumental variable (IV) for the endogenous treatment (i.e., retention). The resulting IV estimand, referred to as a Local Average Treatment Effect (LATE), captures the impact of being retained for "compliers"—students who would be retained if they scored below the cutoff but not if they scored above it.

In Section VI.C, we present evidence that the assumptions necessary for identifying LATEs are likely violated in the Michigan setting. We therefore estimate LATEs only for illustrative purposes. For student i in cohort c, we estimate LATEs using the following two-stage IV model:

$$y_i = \beta_0 + \beta_1^{LATE}\text{Retained}_i + g(\text{MSTEP}_i, \text{Below-Cutoff}_i) + \text{B}\mathbf{X}_{ic} + \varepsilon_{ic} \tag{2}$$

$$\text{Retained}_i = \delta_0 + \delta_1^{FS}\text{Below-Cutoff}_i + h(\text{MSTEP}_i, \text{Below-Cutoff}_i) + \Pi\mathbf{X}_{ic} + \epsilon_{ic}, \tag{3}$$

where δ_1^{FS} is the "first stage" effect of scoring below the cutoff on being retained, and β_1^{LATE} is the effect of being retained on a student's outcomes. As before, our primary specification includes covariates ($\mathbf{X}_{ic}$) for precision, uses optimally selected bandwidths, and uses g and h functions that are linear with potentially different slopes above and below the cutoff.

V.B Identifying Assumptions

Several assumptions are necessary to identify ITT and LATE effects in an RD design. The first key assumption is that unobservable determinants of the outcome do not change discontinuously through the cutoff. One reason this could occur in our context would be if families or schools could manipulate M-STEP scores. Given that the exams are proctored and graded by computer, this seems highly unlikely. To test this, Appendix Figure A2 plots the distribution of ELA M-STEP scores, which appear smooth through the cutoff. Formal tests do not reveal any discontinuities (Appendix Figure A2).

Another reason unobservable factors might change discontinuously through the cutoff is differential attrition from the sample. If students below the cutoff are more (or less) likely to take standardized exams or leave Michigan public schools in the following year, our analysis sample may not be balanced at the cutoff. Fortunately, Appendix Figure A3 shows that differential attrition is not a concern.[11]

Another approach for testing potential violations of this continuity assumption is to examine whether *observable* characteristics change discontinuously through the cutoff. Appendix Figures A4, A5, and A6 show that various student and school covariates are smooth through the cutoff. To test this formally, we re-estimate the ITT model above using a host of student-, school-, and neighborhood-level characteristics as the outcome variable (and with no baseline characteristics included as covariates). Appendix Table A2 presents these estimates.

The estimates are all small in magnitude and, with a few exceptions, not statistically significant. One exception involves our measure of economic disadvantage. Students who score just below the cutoff are 2 percentage points (2.5%) *more* likely to be economically disadvantaged (ED) than their peers scoring just above the cutoff. While this difference is statistically significant, it is small relative to the control mean of 82%. Below, we demonstrate that our results are robust to excluding all covariates (including the economic disadvantage indicator) as well as including additional covariates. To the extent that one is concerned about remaining unobservable factors, the ED discontinuity suggests that our estimates might *understate* the true ITT effects.[12]

If this continuity assumption holds, ITT estimates will be unbiased. To obtain unbiased estimates of the LATE, however, two more assumptions must be met. The first is a standard monotonicity assumption. The second assumption, commonly referred to as an "exclusion restriction," is that falling just below the cutoff only affects outcomes through its impact on treatment. In our context, this means that being eligible

[11] We lose only 0.7% of students due to attrition from the Michigan public school system after third grade. This may happen because a student leaves the state, is homeschooled, or enrolls in a private school.

[12] See Appendix B for the same set of internal validity tests conducted separately for the 2021 and 2022 cohorts. The results for each individual cohort closely resemble the overall results.

for retention can only impact a student's future outcomes by influencing whether the student is retained. Given the multifaceted nature of Michigan's retention policy, this is a strong assumption that we do not find credible. We present detailed evidence supporting this belief in Section VI.C.

VI Results

VI.A Impact Estimates

Figure 1 presents a visual representation of our key impact estimates. In each subfigure, the blue dots reflect the average outcome of all students with a particular third-grade ELA score. The vertical dashed lines indicate the retention cutoff. The red lines are linear best-fit lines plotted over the optimally selected bandwidths. The impact estimate in each subfigure is β_1^{ITT} in equation 1 above.

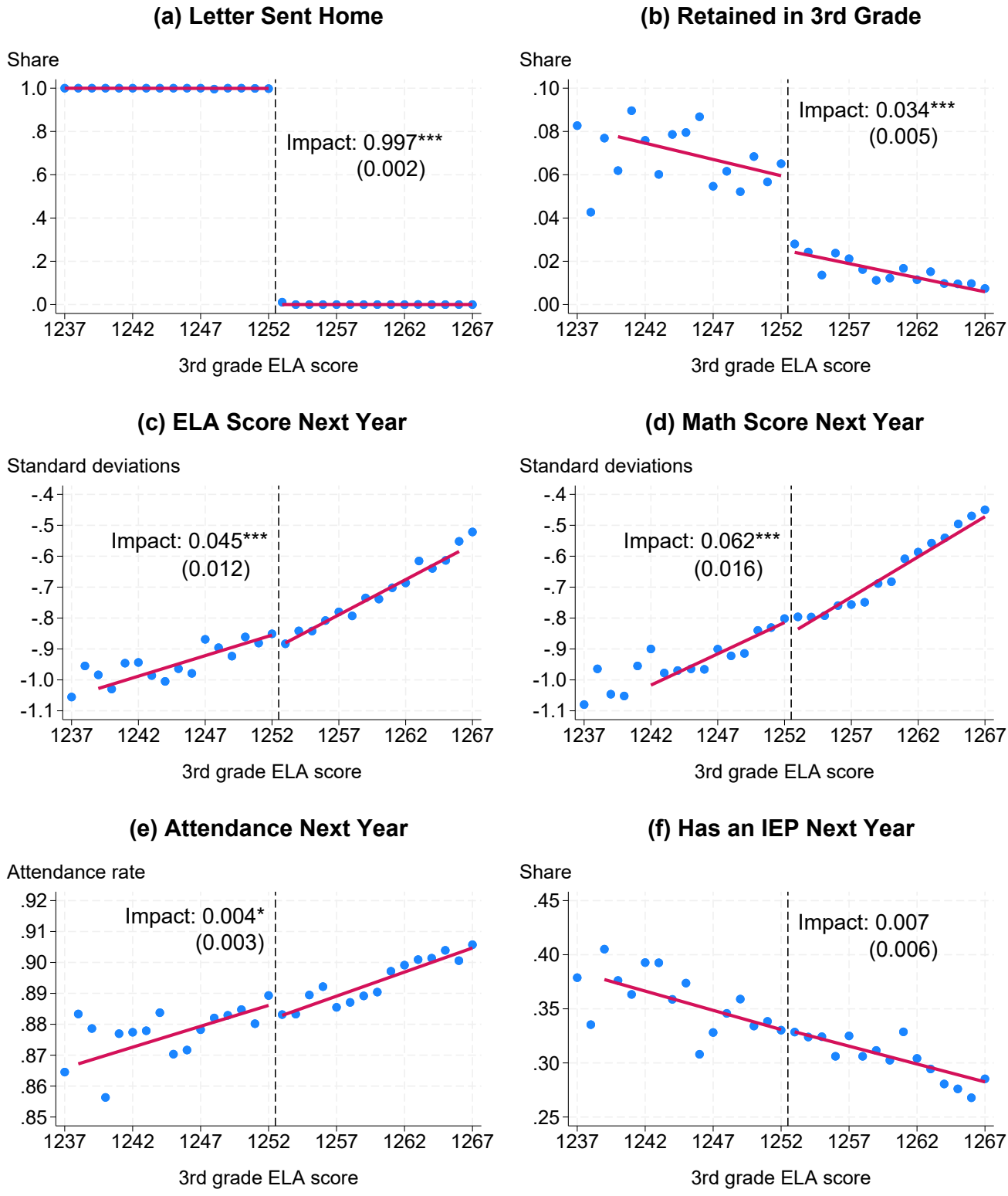
Figure 1(b) shows that the likelihood of retention jumps 3.4 percentage points (from 2.5% to 5.9%) for students who just miss the state-defined ELA cutoff score of 1253. As in Mississippi and Tennessee, very few students were actually retained under Michigan's policy. Students received good cause exemptions for a variety of reasons, the most common being parent request and having an IEP (54.3% and 22.6% of flagged students, respectively).

The next two subfigures (c and d) show average ELA and math scores in the year following the retention decision (fourth grade for most students) separately by third-grade ELA scores. These figures reveal small positive discontinuities just left of the cutoff, suggesting that being flagged for retention increased standardized scores in the following year.[13] These discontinuities reflect "same-age" comparisons. We obtain similar results when we use students' fourth grade scores ("same-grade" comparisons) instead (see Appendix Figure A7 and Appendix Table A3).[14] Both the ELA and math impact estimates are small in magnitude but highly statistically significant.

[13] It is theoretically possible that we estimate positive effects on test scores because schools reallocate resources from students above the RD cutoff to those below it. While we do not believe resource shifting is a key part of the story, testing the theory is unfortunately not possible because there is only modest variation in the share of students on either side of the cutoff.

[14] There is not a consensus in the research or practice community regarding which comparison is most valuable. We present same-age comparisons in the main text, but because so few students were retained, same-age and same-grade estimates are very similar.

Figure 1: ITT Impact Estimates



Notes: The impact estimates printed on each plot are estimated as described in Section V.A. The numbers in parentheses below the impact estimates are standard errors (clustered at the level of the running variable). We denote statistical significance by *** for p<0.01, ** for p<0.05, and * for p<0.1. An IEP (individualized education program) is a document that outlines the programs and services a student with disabilities will receive to help them realize their educational goals, sometimes referred to as "special education."

While visual inspection of the ELA figure suggests the linear form matches the data quite well, the same is not true for math. The three dots immediately to the right of the cutoff in the math figure appear flat, after which there is a notably jump, and then the next three points appear flat as well. Appendix Figure A8 confirms that the impact estimate is close to zero and statistically insignificant when we focus on students closest to the cutoff. For this reason, we are not confident concluding that the policy had an impact on subsequent math achievement, and we focus exclusively on ELA results for the remainder of the paper.

The final two subfigures show results for attendance and IEP status in the following year. In both cases, the estimates are extremely small in magnitude, suggesting that being flagged for retention at the end of third grade did not have an appreciable impact on a student's attendance or their receipt of services related to disability status. Appendix Table A3 presents estimates from our RD analysis for a wider set of outcomes, showing no impacts on switching schools or districts in the next school year.

Table 2 presents ITT estimates separately for several student subgroups.[15] As above, these estimates reflect the impact of being flagged for retention as opposed to actually being retained. Impacts on ELA scores are broadly comparable across gender and economic disadvantage, but differ by race and disability. Looking first at race, we see that the effects are positive and significant for White students but much smaller (and not statistically significant) for Black and Asian students. Point estimates for Hispanic students are comparable to those for White students but less precise and thus not statistically different than zero at conventional levels. The difference between White and Black students is significant at the 10% level (p=0.094). Turning to IEP status, we find that students without an IEP experience a small gain in achievement (0.064 SD) as a result of being flagged for retention. In contrast, students with an IEP see no benefit. The difference between the two subgroups is nearly significant at the 10% level (p=0.110).

[15]The subgroup impact estimates are highly similar when we estimate models with district fixed effects. See Appendix Table A4 for internal validity tests by subgroup. Across all subgroups, we estimate very few discontinuities in enrolling in a Michigan public school and in test-taking in the next school year. We do, however, find a statistically significant decrease in the probability of taking the M-STEP next year for non-economically disadvantaged students flagged for retention. This is likely because these students became more likely to have an IEP in the following year (see Table 2). Although we estimate positive effects on subsequent ELA test scores for this group, we cannot rule out that this finding is at least partially due to attrition from test-taking. We also test for discontinuities in a composite measure of observable characteristics for each subgroup. Mirroring the results for the full sample, we find some small but statistically significant discontinuities going in the opposite direction of our ELA impact estimates.

Table 2: Heterogeneity in ITT Impacts

	Retained		ELA Score		Attendance		Has an IEP	
	Control Mean	Impact Estimate (SE)	Control Mean	Impact Estimate (SE)	Control Mean	Impact Estimate (SE)	Control Mean	Impact Estimate (SE)
Girls	0.032	0.038***	-0.857	0.055***	0.883	0.009***	0.275	0.015***
N=14,680		(0.006)		(0.018)		(0.003)		(0.005)
Boys	0.021	0.034***	-0.866	0.051***	0.883	0.001	0.366	0.002
N=18,317		(0.009)		(0.013)		(0.005)		(0.007)
Econ. disadvantaged	0.029	0.036***	-0.902	0.043***	0.876	0.004	0.312	0.001
N=26,490		(0.006)		(0.016)		(0.003)		(0.005)
Not econ. disadvantaged	0.011	0.025***	-0.678	0.065*	0.917	0.006	0.391	0.033***
N=6,507		(0.006)		(0.039)		(0.005)		(0.012)
White	0.018	0.036***	-0.769	0.056***	0.906	0.002	0.403	0.006
N=15,675		(0.010)		(0.022)		(0.003)		(0.014)
Black	0.038	0.034***	-0.963	0.019	0.855	0.005	0.248	0.004
N=12,625		(0.007)		(0.011)		(0.005)		(0.009)
Hispanic	0.011	0.023	-0.864	0.066	0.888	0.011**	0.310	0.018
N=3,557		(0.026)		(0.059)		(0.004)		(0.015)
Asian	0.038	0.094***	-0.788	-0.016	0.937	-0.006	0.212	0.059*
N=677		(0.036)		(0.081)		(0.006)		(0.039)
English learner (EL)	0.011	0.049***	-0.914	0.077**	0.905	0.012**	0.246	-0.022***
N=4,052		(0.014)		(0.035)		(0.005)		(0.011)
Not an EL	0.028	0.032***	-0.855	0.043***	0.880	0.003	0.337	0.012***
N=28,945		(0.006)		(0.015)		(0.004)		(0.006)
Has an IEP	0.013	0.008	-0.911	0.002	0.891	-0.001	0.965	0.007
N=9,107		(0.006)		(0.028)		(0.005)		(0.010)
No IEP	0.031	0.045***	-0.843	0.064***	0.880	0.006**	0.075	0.008
N=23,890		(0.006)		(0.017)		(0.004)		(0.008)

*** p<0.01, ** p<0.05, * p<0.1

Notes: We obtain these estimates via RD models as described in Section V.A, allowing bandwidths to vary for each outcome and on each side of the cutoff. Standard errors are clustered at the running variable level. Control group means are calculated using students with ELA M-STEP scores within 2 points above the cutoff. Within subgroup, sample sizes may differ across outcomes depending on bandwidth sizes; as a reference point, the first column shows the number of students scoring within 15 points of the retention cutoff for each subgroup. An IEP (individualized education program) is a document that outlines the programs and services a student with disabilities will receive to help them realize their educational goals, sometimes referred to as "special education." Formal tests of equality between ELA impact estimates obtain the following p-values: 0.110 for students with and without an IEP; 0.094 for White and Black students; 0.921 for White and Hispanic students; 0.397 for White and Asian students; and 0.798 for students who are and aren't economically disadvantaged. Formal tests of equality between attendance impact estimates obtain the following p-values: 0.307 for girls and boys; 0.012 for students who are and aren't English learners; and 0.150 for students with and without an IEP. Formal tests of equality between IEP impact estimates obtain the following p-values: 0.003 for girls and boys; 0.001 for students who are and aren't economically disadvantaged; and 0.024 for students who are and aren't English learners.

Impacts on attendance and having an IEP in the following school year also differ across subgroups. Girls experience a positive impact on both of these outcomes, whereas boys experience no impact on either. The impact on having an IEP in the following year is large for children from non-economically disadvantaged families, but close to zero for children from economically disadvantaged families. Interestingly, English learners flagged for retention become *less* likely to have an IEP in the following year, while native English speakers become *more* likely. Though it is difficult to investigate the reasons for this heterogeneity, differences in obtaining IEPs by family income and EL status could reflect differences in family knowledge of school resources or differences in the schools students attend. It is also possible that when EL students are flagged for retention, schools adjust by providing them additional EL services, which might mitigate the need for services provided through an IEP.[16]

VI.B Robustness Checks

To probe the validity of our RD estimates, we conduct a battery of robustness checks. We present estimates using alternative bandwidths in Appendix Figure A8; placebo estimates using false cutoff points in Appendix Figure A9; estimates from quadratic models in Appendix Table A5; estimates from models with no covariates and with additional covariates in Appendix Table A5; and, following Kolesár and Rothe (2018), standard errors computed without clustering on the running variable in Appendix Table A5.

Across the board, we find consistent evidence that being flagged for retention has a small (but detectable) impact on being retained and on ELA scores. The estimates are quantitatively and qualitatively similar across the many specifications we run. The estimates for attendance and having an IEP are also highly similar across specifications, but because they are less statistically significant in the first place, they sometimes waver between significance and insignificance.

VI.C Mechanisms

We find that being flagged as retention-eligible increased the likelihood of actually being retained by only 3.4 percentage points, yet increased standardized test scores by 0.045 SD. These two facts imply implausibly large ($.045/.034 \approx 1.3$ SD) effects of grade retention in the fuzzy RD framework described above (see Appendix Table A3 for formal LATE estimates). For reference, normative growth in reading between third and fourth grade is only around 0.36 SD (Hill et al., 2008).[17] Moreover, while prior research has found

[16] Kilbride et al. (2024) report similar patterns and discuss potential explanations.

[17] Jacob and Lefgren (2004) provide another useful reference point because they estimate retention effects in a setting with high retention compliance and no explicit violation of the exclusion restriction from school-based services. They find that being retained in third grade increased reading scores by 41% of the average annual gain, compared to our LATE estimate of 361%.

positive and relatively large impacts of grade retention, virtually no estimates come close to exceeding 1 SD. In this section, we investigate what might explain such seemingly incongruous results.

We begin by examining the impact of the policy in districts that did not retain *any* students. In these districts, we can rule out grade retention as a mechanism driving impacts. 67.4% of district-year observations with at least one retention-eligible student did not retain any students. Indeed, 55.8% of districts with a retention-eligible student did not retain any students in any year of the policy.[18]

Appendix Table A6 compares districts with zero versus non-zero retention rates. As expected, district-years with zero retention rates are somewhat smaller than other district-years, with only 143 third graders and 5 retention-eligible third graders compared with 184 and 12 in other district-years. However, even if we restrict the sample to district-years with at least 10 retention-eligible students, we find that 45.8% did not retain any students.[19] Controlling for the number of retention-eligible third graders in a district-year, districts with no retention have substantially higher shares of White students (16 percentage points) and lower shares of Black (16pp) and economically disadvantaged (13pp) students. They are also much less likely to be charter schools (23pp) (Appendix Table A7).

Table 3 shows the impact of being flagged for retention separately for districts that did and did not retain students.[20] The estimates suggest that being flagged for retention has a positive impact on student achievement even in districts where *no* students were actually retained (0.062 SD, SE=0.021). When we limit the analysis to districts with at least 10 retention-eligible students, we find even larger estimates (0.099 SD, SE=0.020). Impacts tend to be *smaller* in places that retained students.[21]

The existence of positive effects even in districts that did not retain students suggests that other factors were improving student literacy skills. One potential mechanism involves parents. If a student scores below the cutoff, the state sends a letter to parents informing them of their child's reading deficiency and urging them to contact their school to discuss the "best path forward" (see template letter in Appendix Figure A12).

[18]For the purpose of these calculations, we focus only on students scoring within a bandwidth of ±18 points of the cutoff. We do this because these are the students who provide the identification for our treatment effect estimates. Further, we focus on district-years with at least one student on each side of the cutoff (within 18 points). This restriction excludes roughly 22.1% of district-year observations. Lastly, we define a district-year as having no retention if no students within 18 points on either side of the cutoff were retained. Given these restrictions and definitions, 10.4% of the district-years we classify as having no retention have at least one retained student outside of our bandwidth.

[19]Retention compliance was fairly uniformly low even in these district-years. Focusing on students within 18 points of the retention cutoff, the share of students retained below the cutoff minus the share of students retained above the cutoff was less than 15 percentage points in 82 percent of these district-years.

[20]Appendix Figure A10 shows that there is no discontinuity in observable characteristics (as measured by a composite predicted ELA score) at the retention cutoff in any of the four district retention samples.

[21]It is unclear why the ELA impacts are smaller in district-years with retention. One explanation could be that district-years with retention had smaller discontinuities at the retention cutoff in other reading interventions. Appendix Figures C4 and C5 provide some evidence for this possibility, but the differences between district-years with and without retention are imprecise. A smaller discontinuity in other school-based interventions would improve our ability to estimate LATEs for these district-years, but we would still be unable to rule out bias from parent and teacher responses.

Table 3: ITT Impact Estimates, by Presence of Retention in District-Year

	1+ Student on Each Side of Cutoff		10+ Students on Each Side of Cutoff	
	No Retention (1)	Has Retention (2)	No Retention (3)	Has Retention (4)
Retained	-	0.068***	-	0.041***
	-	(0.009)	-	(0.012)
	-	[0.048]	-	[0.038]
ELA score next year	0.062***	0.044**	0.099***	0.049
	(0.021)	(0.018)	(0.020)	(0.029)
	[-0.795]	[-0.936]	[-0.858]	[-0.974]
Attendance rate next year	0.001	0.008***	0.003	0.011***
	(0.006)	(0.003)	(0.011)	(0.004)
	[0.897]	[0.869]	[0.890]	[0.860]
Has an IEP next year	0.011	0.008*	-0.001	0.009*
	(0.010)	(0.006)	(0.011)	(0.007)
	[0.355]	[0.289]	[0.318]	[0.261]
Districts	568	311	77	90
District×years	825	399	92	109
Students	19,095	18,409	6,760	11,947

*** p<0.01, ** p<0.05, * p<0.1

Notes: We obtain these estimates via RD models as described in Section V.A, allowing bandwidths to vary for each outcome and on each side of the cutoff. Standard errors, shown in parentheses under the point estimates, are clustered at the running variable level. Control group means, shown in brackets under the point estimates, are calculated using students with ELA M-STEP scores within 2 points above the cutoff. Columns 1 and 2 (3 and 4) restrict the sample to district-years with at least 1 student (10 students) on each side of the cutoff within 18 points. A district-year is classified as "no retention" if no student within 18 points on either side of the cutoff is retained in third grade. Note that the same district may appear in multiple columns if its retention status or the number of students it has within the bandwidth differs across years. We use students scoring more than 18 points away from the cutoff for estimation, but the "students" row at the bottom of the table is limited to those within 18 points for illustrative purposes. An IEP (individualized education program) is a document that outlines the programs and services a student with disabilities will receive to help them realize their educational goals, sometimes referred to as "special education." See Appendix Figure A11 for ITT plots for ELA scores.

This letter can provide parents with new information and/or make a child's already-known challenges more salient. Parents might respond by seeking help for their child outside of school and/or by advocating for more in-school assistance. Figlio et al. (2023) find that third-grade retention of a child has a positive spillover effect on their siblings, which the authors speculate is partially due to parental responses. Importantly, parent responses may occur whether or not a child is actually retained.

A second, complementary mechanism involves support provided by a child's school. As noted above, the law in Michigan (and many other states) requires schools to provide additional literacy support to students eligible for retention *even if they are promoted* to fourth grade (Kilbride and Walker, 2023). Finally, being flagged for retention could influence student learning via its impact on informal teacher attitudes and

Table 4: Share of Schools Offering Additional Supports to Students (Next Year)

	Student Group				
	Below Cutoff Retained (1)	Below Cutoff Promoted (2)	Above Cutoff Promoted (3)	P-value of (1)-(2)	P-value of (2)-(3)
Increased dosage (%)					
Summer reading programs	72.1	71.0	64.5	0.572	0.001
Assign to a high-quality teacher	67.1	69.7	64.8	0.132	0.001
Extra instructional time in literacy	71.3	73.6	68.4	0.160	0.002
High-dosage tutoring	39.7	35.9	28.8	0.083	0.007
Work with families on home reading	68.1	62.9	58.5	0.002	0.008
Before- or after-school literacy interventions	32.4	32.1	29.2	0.870	0.055
Small group instruction during school	82.8	84.6	82.8	0.250	0.297
Supplemental virtual learning	18.8	19.3	18.5	0.618	0.406
Other support (%)					
Literacy intervention curriculum	48.6	48.8	46.2	0.848	0.059
Focus on essential skills	73.9	74.2	71.0	0.873	0.077
Data-driven instruction	82.0	84.6	83.8	0.059	0.602
Total number of supports (#)	6.3	6.4	6.0	0.804	0.000

Notes: These results are from 383 principal surveys that have complete responses for each student group in the table. Among these survey responses, 227 correspond to the 2021 cohort and 156 correspond to the 2022 cohort. "High-dosage tutoring" was only asked in the survey for the 2022 cohort. Items are listed in ascending order based on the last column. See Appendix C for details on the survey.

behaviors. The flag is a clear signal to a child's future teacher that the student is in need of additional support. While parent involvement and formal school supports would likely have positive impacts on children, the signal provided to teachers could, in theory, negatively impact a student by lowering a teacher's expectations for the child (Weinstein, 2004).[22]

While we do not have any direct evidence on parent involvement or teacher beliefs, we do have information relating to school supports. As part of a larger research project studying the RBG3 law, we administered surveys to school principals in the 2021-22 and 2022-23 school years. Table 4 presents evidence from the survey on additional supports provided to students who scored near the retention cutoff in third grade. To the best of our knowledge, this is the only survey to ask school administrators about services separately for low-achieving students who were promoted as well as those who were retained. Specifically, principals were asked to indicate whether their school provided each of 11 different supports to three different groups: students who (i) scored below the cutoff and were retained, (ii) scored below the cutoff and were promoted, and (iii) scored just above the cutoff.[23] These results are from 383 principals that gave complete responses to this survey question. For more details on the survey, see Appendix C. As shown in Appendix Table C1,

[22] In theory, being flagged for retention could influence a child's performance by changing their expectations for themselves. This is certainly plausible for students who are actually retained. However, we believe this channel is less relevant for students who receive a good cause exemption and advance to fourth grade because they likely interpret promotion as a positive signal of their ability despite a weak test performance.

[23] The survey asks about students who scored within 18 points above the cutoff, 1253-1271, because state guidance recommends districts provide additional support to these students.

the schools with survey responses are very similar to other schools in the state.

Table 4 shows that principals provided extra support to students scoring below *and* just above the retention cutoff. Some of the most common supports were summer reading programs, extra instructional time in literacy, and working with families on home reading. A comparison of columns 1 and 2 reveals that students flagged for retention received similar levels of additional support *regardless* of whether they were retained or promoted. A comparison of columns 1 and 2 versus 3 reveals that flagged students received more support in the following year than non-flagged students.[24] This difference helps rationalize the positive ITT estimates alongside the small first-stage (retention) estimates.

These findings have two key implications. First, our ITT estimates come from a true policy bundle, not any single treatment. Second, the discontinuity in supports at the cutoff is a clear violation of the exclusion restriction—the assumption one would need to isolate the effect of retention itself (the LATE in the framework above). Put differently, we cannot estimate the effect of being retained even in districts with a strong first stage on retention.

VII Discussion

The effects of grade retention laws continue to be a hotly debated topic, even as additional states move to implement such policies. Our study of Michigan's third-grade retention law makes several novel methodological and substantive contributions to this evidence base. We find that children who scored just below the retention threshold showed modest but educationally meaningful improvements in their reading scores one year later compared to children who just barely passed. Most interestingly, reading benefits accrued even in districts that did not retain *any* students. Survey evidence suggests that these literacy improvements may be attributable to supplemental support provided to low-achieving students who were promoted via good cause exemptions. These results have important implications for research and policy.

First, our findings shine light on two neglected issues in the literature: the potential bias in previous estimates of retention effects and the critical importance of implementation data in evaluations of multifaceted literacy laws. Prior studies recognized that *retained* students likely received a package of various interventions, ranging from additional support during the retained year to the influence of parent, teacher, and student expectations (Mariano et al., 2024; Mumma and Winters, 2023; Özek, 2015; Winters and Greene, 2012). However, the prior literature did not seriously contemplate that being flagged for retention might influence

[24] In Table C6 we confirm that our results hold when we link principal responses to the student administrative data and limit the survey sample to school-years with at least one flagged, promoted student and at least one non-flagged, promoted student in the data.

student learning even in the absence of actual retention.[25]

The existence of such influences violates a key assumption underlying the estimation of retention effects (LATEs) within an RD framework and suggests that LATE estimates (but not ITT estimates) in the prior literature might be biased upward, as we believe is the case for our Michigan results. The task ahead for researchers is to collect implementation data that contain information on the services individual students receive, both prior to and following the retention decision. This will likely require administrative data that are typically housed at the school and/or district (rather than state) level. It may also necessitate broader teams with expertise in survey and qualitative research.

From a policy perspective, our findings raise questions about the utility of test-based grade retention, which is part of state literacy law for more than a third of all public-school third graders in the U.S. Retention is not only unpopular among students, parents, and educators, it is also costly and introduces practical challenges for schools trying to balance class sizes and teaching loads between grades. In states that recently implemented literacy laws, the use of good cause exemptions has been extremely common, resulting in very small impacts on retention—for example, roughly 3.4pp in Michigan and 6pp in Mississippi (Mumma and Winters, 2023). In Tennessee, 60% of third graders scored below the retention threshold in 2023-24, exceeding capacity for retention; only 1.2% were ultimately retained (Wegner, 2024).

To the extent that supplemental services improve academic performance for low-achieving students, it may be possible to amend state policies in a way that promotes learning with less political backlash. To be clear, our analysis was not designed to evaluate the causal impact of such services. There is evidence that some supports such as tutoring (Jacob et al., 2016; Nickow et al., 2024), teacher coaching (Kraft et al., 2018), additional instruction time (Figlio et al., 2018), and tailored interventions (Baker et al., 2015; Coyne et al., 2018; Cruz et al., 2023) improve student learning, but more research is needed to determine which supports are most effective, for which students. Maryland's approach offers one path; there, parents must opt into grade retention and they are given a menu of supplemental reading support programs to select from (Maryland State Department of Education, 2024).

Finally, our subgroup findings provide suggestive evidence that Michigan's third-grade reading law benefited historically more advantaged subgroups more so than their peers. The prior literature on test-based retention is mixed in terms of which subgroups benefit most. Several studies find that historically marginalized groups benefit disproportionately; others find little heterogeneity. More evidence on which subgroups benefit from reading laws and whether mechanisms differ by group should be a priority in future studies.

[25] One exception involves studies that examine the effects of summer school tied to promotion cutoffs (Eren et al., 2017; Jacob and Lefgren, 2004).

Our study has several important limitations. Due to the pandemic, outcome data were available at lower rates in our first cohort. We showed in Appendix B that our findings are qualitatively similar across cohorts, but nonetheless, the generalizability of our findings is not as complete as we would prefer. Our study also took place in Michigan only; it is possible our findings may not generalize to other states that differ in demographics or in the specifics of their state reading laws. We also lacked measures of other important outcomes such as students' social-emotional skills and behavior that prior literature shows is important to understanding the effects of these laws (Mariano and Martorell, 2013; McCombs et al., 2009; Özek, 2015; Wu et al., 2010). Finally, although our study is unusual in having some data on the implementation of non-retention components of a reading law, these data were more limited than we would have liked, especially for understanding subgroup findings.

Taken together, our findings show that comprehensive third-grade reading retention laws like Michigan's can move the needle modestly in improving children's reading skills. However, retention may be a much less important driver of these effects than previously understood.

References

Appleton, Aleksandra, "Indiana's revamped reading law could have big consequences for students from low-income families," *Chalkbeat Indiana*, July 2024.

Baker, Scott K., Keith Smolkowski, Erin A. Chaparro, Jill L. Smith, and Hank Fien, "Using Regression Discontinuity to Test the Impact of a Tier 2 Reading Intervention in First Grade," *Journal of Research on Educational Effectiveness*, 2015, *8* (2), 218–244.

Calonico, Sebastian, Matias D. Cattaneo, and Max H. Farrell, "On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference," *Journal of the American Statistical Association*, 2018, *113* (522), 767–779.

_ , _ , and _ , "Optimal Bandwidth Choice for Robust Bias-Corrected Inference in Regression Discontinuity Designs," *The Econometrics Journal*, 2020, *23* (2), 192–210.

_ , Matias D Cattaneo, and Rocío Titiunik, "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 2014, *82* (6), 2295–2326.

_ , Matias D. Cattaneo, Max H. Farrell, and Rocío Titiunik, "Regression Discontinuity Designs using Covariates," *Review of Economics and Statistics*, 2019, *101* (3), 442–451.

Coyne, Michael D., Alexandra Oldham, Sara M. Dougherty, Kelly Leonard, Taylor Koriakin, Nicholas A. Gage, and Martha Gillis, "Evaluating the Effects of Supplemental Reading Intervention within an MTSS or RTI Reading Reform Initiative Using a Regression Discontinuity Design," *Exceptional Children*, 2018, *84* (4), 350–367.

Cruz, Joana, Sara Mendes, Daniela Alves, Susana Marques, and Irene Cadime, "Fluency Training for Struggling Readers: Examining the Effects of a Tier-2 Intervention in Third Graders," *Education Sciences*, 2023, *13* (9), 908.

Eren, Ozkan, Briggs Depew, and Stephen Barnes, "Test-based Promotion Policies, Dropping Out, and Juvenile Crime," *Journal of Public Economics*, 2017, *153*, 9–31.

ExcelinEd, "Comprehensive Early Literacy Policy: A State-by-State Analysis of Fundamental Principles," https://excelined.org/wp-content/uploads/2021/10/ExcelinEd_PolicyToolkit_EarlyLiteracy_StatebyStateAnalysis_2021.pdf January 2024. Accessed: 2025-02-17.

Figlio, David, Kristian L. Holden, and Umut Özek, "Do students benefit from longer school days? Regression discontinuity evidence from Florida's additional hour of literacy instruction," *Economics of Education Review*, 2018, *67*, 171–183.

_ , Krzysztof Karbownik, and Umut Özek, "Sibling Spillovers May Enhance the Efficacy of Targeted School Policies," Technical Report Working Paper 31406, National Bureau of Economic Research, Cambridge, MA 2023.

Gamse, Beth C., Robin T. Jacob, Megan Horst, Beth Boulay, and Fatih Unlu, "Reading First Impact Study: Final Report," Technical Report NCEE 2009-4038, National Center for Education Evaluation and Regional Assistance November 2008.

Hill, Carolyn J., Howard S. Bloom, Rebeck Black, and Mark W. Lipsey, "Empirical Benchmarks for Interpreting Effect Sizes in Research," *Child Development*, 2008, *2*, 172–177.

Hwang, NaYong and Cory Koedel, "Helping or Hurting: The Effects of Retention in the Third Grade on Student Outcomes," *Educational Evaluation and Policy Analysis*, 2023.

Indiana Department of Education, "Process & Considerations for Issuing a Good Cause Exemption," https://www.in.gov/doe/files/Process-and-Considerations-for-Issuing-a-Good-Cause-Exemption.pdf 2021. Accessed: 2025-02-17.

Indiana State Board of Education, "SBOE 10-13-21 IREAD-3 Guidance Memo," https://www.in.gov/doe/files/IREAD-3-memo-updated-guidance-October-2021.pdf 2021. Accessed: 2025-02-17.

Jacob, Brian A., "When evidence is not enough: Findings from a randomized evaluation of Evidence-Based Literacy Instruction (EBLI)," *Labour Economics*, 2017, *45*, 5–16.

_ and Lars Lefgren, "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis," *The Review of Economics and Statistics*, 2004, *86* (1), 226–244.

_ and _ , "The effect of grade retention on high school completion," *American Economic Journal: Applied Economics*, 2009, *1* (3), 33–58.

Jacob, Robin, Catherine Armstrong, Alice B. Bowden, and Yanzhi Pan, "Leveraging Volunteers: An Experimental Evaluation of a Tutoring Program for Struggling Readers," *Journal of Research on Educational Effectiveness*, 2016, *9* (sup1), 67–92.

Kilbride, Tara and Seth Walker, "Read by Grade Three Retention Eligiblity: 2023 Update," Technical Report, Education Policy Innovation Collaborative November 2023.

__ , Josep Nadal-Fernandez, and Seth Walker, "Retention and Exemption Decisions Under the Read by Grade Three Law," Technical Report, Education Policy Innovation Collaborative June 2024.

Kolesár, Michal and Christoph Rothe, "Inference in Regression Discontinuity Designs with a Discrete Running Variable," *American Economic Review*, 2018, *108* (8), 2277–2304.

Kraft, Matthew A., David Blazar, and Dylan Hogan, "The Effect of Teacher Coaching on Instruction and Achievement: A Meta-analysis of the Causal Evidence," *Review of Educational Research*, 2018, *88* (4), 547–588.

Mariano, Louis T. and Paco Martorell, "The Academic Effects of Summer Instruction and Retention in New York City," *Educational Evaluation and Policy Analysis*, 2013, *35* (1), 96–117.

Mariano, Louis T, Paco Martorell, and Tiffany Berglund, "The effects of grade retention on high school outcomes: Evidence from New York City schools," *Journal of Research on Educational Effectiveness*, 2024, pp. 1–31.

Martorell, Paco and Louis T. Mariano, "The Causal Effects of Grade Retention on Behavioral Outcomes," *Journal of Research on Educational Effectiveness*, 2018, *11* (2), 192–216.

Maryland State Department of Education, "Comprehensive Literacy Policy Update," https://marylandpublicschools.org/stateboard/Documents/EdPolicyCommittee/09122024/Proposed-Literacy-Policy-A.pdf 2024. Accessed: 2025-02-17.

McCombs, Jennifer Sloan, Sheila Nataraj Kirby, and Louis T. Mariano, "Ending social promotion without leaving children behind: The case of New York City," Technical Report, RAND Corporation 2009.

McCrary, Justin, "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, 2008, *142* (2), 698–714.

Mississippi State Senate, "Senate Bill 2157: Literacy-Based Promotion Act; provide standards for individual reading plans and clarify assessments," https://legiscan.com/MS/text/SB2157/id/1390375 2016. Accessed: 2025-02-17.

Mumma, Kirsten Slungaard and Marcus A. Winters, "The Effect of Retention Under Mississippi's Test-Based Promotion Policy," Technical Report Working Paper 2023-1, Wheelock Educaitonal Policy Center, Boston, MA 2023.

National Assessment of Educational Progress, "NAEP Report Card: Reading," https://www.nationsreportcard.gov/reading/nation/achievement/?grade=4 2025. Accessed: 2025-02-17.

Nickow, Andrew, Philip Oreopoulos, and Vincent Quan, "The Promise of Tutoring for PreK-12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence," *American Educational Research Journal*, 2024, *61*, 74–107.

Özek, Umut, "Hold Back To Move Forward? Early Grade Retention And Student Misbehavior," *Education Finance and Policy*, 2015, *10* (3), 350–377.

Ryan, Allison M., "The peer group as a context for the development of young adolescent motivation and achievement," *Child Development*, 2001, *72* (4), 1135–1150.

__ and Helen Patrick, "The classroom social environment and changes in adolescents' motivation and engagement during middle school," *American Educational Research Journal*, 2001, *38* (2), 437–460.

Schwartz, Sarah, "Which States Have Passed 'Science of Reading' Laws? What's in Them?," *Education Week*, July 2022.

Schwerdt, Guido, Martin R. West, and Marcus A. Winters, "The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida," *Journal of Public Economics*, 2017, *152*, 154–169.

Snow, Catherine E., Susan M. Burns, and Peg Griffin, *Preventing Reading Difficulties in Young Children*, Washington, DC: National Academies Press, 1998.

Strunk, Katharine O., Tanya S. Wright, John Westall, Tara Kilbride, Qiong Zhu, Amy Cummings, Andrew Utter, and Madeline Mavrogordato, "Michigan's Read by Grade Three Law: Year Two Report," Technical Report, Education Policy Innovation Collaborative February 2022.

_ , _ , Tara Kilbride, Qiong Zhu, Amy Cummings, Joanne West, Meg Turner, and Craig De Voteo, "Michigan's Read by Grade Three Law: Year One Report," Technical Report, Education Policy Innovation Collaborative March 2021.

Wegner, Rachel, "What to know as TCAPs start and Tennessee third, fourth graders face being held backs," *The Tennessean*, March 2024.

Weinstein, Rhona S., *Reaching higher: The power of expectations in schooling*, Harvard University Press, 2004.

West Virginia House of Delegates, "House Bill 3035: Third Grade Success Act," https://legiscan.com/WV/bill/HB3035/2023 2023. Accessed: 2025-02-17.

Winters, Marcus A. and Jay P. Greene, "The Medium-Run Effects of Florida's Test-Based Promotion Policy," *Education Finance and Policy*, 2012, *7* (3), 305–330.

Wu, Wei, Stephen G. West, and Jan N. Hughes, "Effect of grade retention in first grade on psychosocial outcomes," *Educational Psychology*, 2010, *102*, 135–152.

Zhong, Jiee, "Early Grade Retention Harms Adult Earnings," 2024. Working Paper.

A Additional Tables and Figures

Table A1: Summary Statistics for Third Graders in 2021 and 2022

	All First-Time Third-Graders	Analysis Sample	Analysis Sample Within Bandwidth
Student Characteristics			
2022 cohort	0.496	0.571	0.582
Female	0.487	0.491	0.446
White	0.631	0.672	0.480
Black	0.224	0.189	0.378
Hispanic	0.086	0.081	0.108
Asian	0.047	0.045	0.020
NA, NH, or PI	0.013	0.013	0.014
Economically disadvantaged	0.564	0.541	0.802
Has an IEP	0.157	0.145	0.272
English learner (EL)	0.091	0.086	0.123
Previously retained	0.031	0.027	0.054
Same district 2+ years	0.100	0.096	0.123
Math M-STEP score	-	1,292.7	1,261.4
ELA M-STEP score	-	1,292.7	1,257.4
ELA M-STEP score below 1253	-	0.054	0.243
Retained in third grade	0.011	0.008	0.027
Other Characteristics			
Third-graders in district-year	530	464	577
City	0.252	0.214	0.333
Rural	0.172	0.189	0.161
Suburb	0.456	0.467	0.396
Town	0.119	0.130	0.110
Charter school	0.123	0.116	0.192
School ED share	0.561	0.543	0.700
School IEP share	0.160	0.157	0.163
School EL share	0.084	0.080	0.098
School White share	0.633	0.667	0.518
School Black share	0.184	0.154	0.297
School Hispanic share	0.087	0.083	0.100
School Asian share	0.037	0.036	0.024
School NA/NH/PI share	0.005	0.006	0.006
Neighborhood BA+ share	0.169	0.172	0.126
Neighborhood median HH income	$60,009	$61,318	$47,307
Districts	822	803	780
District×years	1,630	1,571	1,496
Students	208,279	168,443	31,285

Notes: The final column restricts the sample to students scoring within ±15 points of the retention cutoff. "NA, NH, or PI" stands for Native American, Native Hawaiian, or Pacific Islander. An IEP (individualized education program) is a document that outlines the programs and services a student with disabilities will receive to help them realize their educational goals, sometimes referred to as "special education." Neighborhood characteristics are calculated at the Census block group level. Income is converted to real 2015 dollars.

Table A2: Tests for Covariate Continuity Through the Cutoff

	Control Mean	Point Estimate	P-value
Composite			
Predicted ELA M-STEP score	1,281.0	-0.5	0.000
Student-Level Characteristics			
2022 cohort	0.563	0.015	0.525
Female	0.436	0.011	0.167
White	0.445	0.007	0.602
Black	0.413	-0.011	0.237
Hispanic	0.109	-0.002	0.981
Asian	0.021	0.004	0.061
NA, NH, or PI	0.012	0.001	0.342
Economically disadvantaged	0.821	0.020	0.000
Has an IEP	0.276	-0.007	0.809
English learner (EL)	0.114	0.007	0.355
Previously retained	0.058	0.003	0.604
Same district 2+ years	0.126	-0.002	0.505
Other Characteristics			
City	0.365	0.003	0.887
Rural	0.158	-0.007	0.525
Suburb	0.372	0.009	0.458
Town	0.105	-0.003	0.626
Charter school	0.198	0.000	0.843
School ED share	0.715	0.016	0.000
School IEP share	0.162	-0.001	0.993
School EL share	0.095	0.004	0.143
School White share	0.496	-0.014	0.071
School Black share	0.319	0.007	0.789
School Hispanic share	0.099	0.003	0.216
School Asian share	0.024	-0.001	0.378
School NA/NH/PI share	0.005	0.000	0.368
Neighborhood BA+ share	0.123	-0.004	0.147
Neighborhood median HH income	45,811.7	-829.1	0.038

Notes: The sample includes first-time third-grade students with valid M-STEP scores the following year. We estimate models using the procedure developed by Calonico et al. (2018, 2020, 2014, 2019), allowing bandwidths to vary for each outcome and on each side of the cutoff. Standard errors are clustered at the running variable level. Control group means are calculated using students with ELA M-STEP scores within 2 points above the cutoff. "NA, NH, or PI" stands for Native American, Native Hawaiian, or Pacific Islander. An IEP (individualized education program) is a document that outlines the programs and services a student with disabilities will receive to help them realize their educational goals, sometimes referred to as "special education." Neighborhood characteristics are calculated at the Census block group level. Income is converted to real 2015 dollars.

Table A3: ITT and LATE Estimates

	No Exclusion Restriction Needed		Exclusion Restriction Needed			
	Control Mean	ITT Estimate (SE)	Complier Control Mean	LATE Estimate (SE)	Optimal Bandwidth	Observations Within Bandwidth
Analysis Sample (N=168,443)						
Enrolled (next year)	0.993	-0.001	-	-	(-12,35)	71,761
		(0.001)		-		
Took M-STEP (next year)	0.941	-0.001	-	-	(-16,32)	66,020
		(0.005)		-		
Took M-STEP (fourth grade)	0.939	-0.001	-	-	(-14,32)	65,756
		(0.004)		-		
Conditional on Enrollment Next Year (N=167,347)						
Retained (next year)	0.026	0.034***	-	-	(-12,17)	36,057
		(0.005)		-		
Attendance rate (next year)	0.883	0.004*	0.904	0.121*	(-14,16)	34,190
		(0.003)		(0.091)		
Has an IEP (next year)	0.326	0.007	0.053	0.206	(-13,19)	39,895
		(0.006)		(0.156)		
Switched schools (next year)	0.176	0.004	0.302	0.100	(-12,26)	52,861
		(0.007)		(0.188)		
Switched districts (next year)	0.099	0.006	0.124	0.176	(-13,29)	59,058
		(0.007)		(0.177)		
Same-Age Test Score Impacts (N=162,355)						
Math score (next year)	-0.797	0.062***	-2.242	1.584***	(-10,26)	50,288
		(0.016)		(0.385)		
ELA score (next year)	-0.862	0.045***	-2.596	1.256***	(-13,14)	29,319
		(0.012)		(0.323)		
Same-Grade Test Score Impacts (N=162,295)						
Math score (fourth grade)	-1.007	0.075***	-2.656	1.970***	(-10,27)	52,239
		(0.017)		(0.516)		
ELA score (fourth grade)	-1.133	0.055***	-3.266	1.629***	(-14,12)	25,759
		(0.012)		(0.387)		

*** p<0.01, ** p<0.05, * p<0.1

Notes: We estimate models using the procedure developed by Calonico et al. (2018, 2020, 2014, 2019), allowing bandwidths to vary for each outcome and on each side of the cutoff. Standard errors are clustered at the running variable level. Control group means are calculated using students with ELA M-STEP scores within 2 points above the cutoff. Complier control means are obtained by estimating fuzzy RD models, following Section V.A, but where the outcome variable is $y_i \times 1[\text{Retained}_i = 0]$, the endogenous treatment variable is $1[\text{Retained}_i = 0]$, and the instrument is Below-Cutoff$_i$. The "optimal bandwidths" are selected when estimating the ITT models; for consistency, we use the same bandwidths when estimating the LATEs. An IEP (individualized education program) is a document that outlines the programs and services a student with disabilities will receive to help them realize their educational goals, sometimes referred to as "special education."

Table A4: Validity Checks for Student Subgroups

	Enrolled Next Year		Took M-STEP Next Year		Predicted ELA M-STEP Score	
	Control Mean	Impact Estimate (SE)	Control Mean	Impact Estimate (SE)	Control Mean	Impact Estimate (SE)
Girls	0.991	0.000	0.932	0.007	1281.7	-0.4
N=14,680		(0.002)		(0.008)		(0.4)
Boys	0.994	-0.003*	0.948	-0.008	1280.4	-0.8***
N=18,317		(0.003)		(0.006)		(0.3)
Econ. disadvantaged	0.993	-0.001	0.941	0.003	1278.0	-0.1
N=26,490		(0.002)		(0.004)		(0.2)
Not econ. disadvantaged	0.991	-0.002	0.939	-0.027***	1295.0	0.0
N=6,507		(0.005)		(0.009)		(0.4)
White	0.992	0.004	0.947	-0.002	1288.4	-0.7**
N=15,675		(0.003)		(0.005)		(0.4)
Black	0.992	-0.008***	0.938	-0.004	1272.9	-0.7***
N=12,625		(0.003)		(0.007)		(0.2)
Hispanic	0.996	0.000	0.941	-0.005	1279.0	0.6
N=3,557		(0.004)		(0.013)		(0.9)
Asian	1.000	-0.003	0.923	0.028	1294.5	-4.8***
N=677		(0.019)		(0.020)		(1.3)
English learner (EL)	0.989	0.005	0.930	0.006	1278.0	0.0
N=4,052		(0.005)		(0.012)		(0.8)
Not an EL	0.993	-0.002	0.942	0.000	1281.4	-0.6***
N=28,945		(0.001)		(0.005)		(0.2)
Has an IEP	0.990	0.000	0.920	0.004	1273.2	-0.8***
N=9,107		(0.002)		(0.010)		(0.3)
No IEP	0.994	-0.002	0.949	0.000	1284.0	-0.5**
N=23,890		(0.001)		(0.005)		(0.2)

*** p<0.01, ** p<0.05, * p<0.1

Notes: We estimate models using the procedure developed by Calonico et al. (2018, 2020, 2014, 2019), allowing bandwidths to vary for each outcome and on each side of the cutoff. Standard errors are clustered at the running variable level. Control group means are calculated using students with ELA M-STEP scores within 2 points above the cutoff. The observation counts in the first column are for students scoring ±15 points of the retention cutoff. An IEP (individualized education program) is a document that outlines the programs and services a student with disabilities will receive to help them realize their educational goals, sometimes referred to as "special education."

Table A5: ITT Impact Estimates From Alternative Specifications

	Control Mean	Primary Estimates (1)	Quadratic Model (2)	No Covariates (3)	Additional Covariates (4)	Robust SEs (No Clustering) (5)
Analysis Sample (N=168,443)						
Enrolled (next year)	0.993	0.001	-0.003**	-0.001	-0.001	-0.001
		(0.001)	(0.002)	(0.001)	(0.001)	(0.003)
Took M-STEP (next year)	0.941	-0.001	0.000	0.000	0.001	-0.001
		(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Took M-STEP (fourth grade)	0.939	-0.001	0.004	0.000	0.002	-0.001
		(0.004)	(0.005)	(0.004)	(0.004)	(0.006)
Conditional on Enrollment Next Year (N=167,347)						
Retained (next year)	0.026	0.034***	0.038***	0.035***	0.036***	0.034***
		(0.005)	(0.004)	(0.005)	(0.005)	(0.007)
Attendance rate (next year)	0.883	0.004*	0.006**	0.005**	0.005**	0.004*
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Has an IEP (next year)	0.326	0.007	0.012***	0.004	0.009**	0.007
		(0.006)	(0.005)	(0.007)	(0.005)	(0.008)
Switched schools (next year)	0.176	0.004	0.013*	0.003	0.003	0.003
		(0.007)	(0.008)	(0.007)	(0.007)	(0.011)
Switched districts (next year)	0.099	0.006	0.009	0.006	0.007	0.007
		(0.007)	(0.007)	(0.007)	(0.006)	(0.009)
Same-Age Test Score Impacts (N=162,355)						
Math score (next year)	-0.797	0.062***	0.055***	0.063***	0.066***	0.060***
		(0.016)	(0.017)	(0.017)	(0.016)	(0.021)
ELA score (next year)	-0.862	0.045***	0.038**	0.039***	0.042***	0.044***
		(0.012)	(0.016)	(0.012)	(0.012)	(0.017)
Same-Grade Test Score Impacts (N=162,295)						
Math score (fourth grade)	-1.007	0.075***	0.062***	0.071***	0.076***	0.070***
		(0.017)	(0.022)	(0.020)	(0.020)	(0.021)
ELA score (fourth grade)	-1.133	0.055***	0.05***	0.048***	0.051***	0.053***
		(0.012)	(0.014)	(0.012)	(0.011)	(0.018)

*** p<0.01, ** p<0.05, * p<0.1

Notes: We estimate models using the procedure developed by Calonico et al. (2018, 2020, 2014, 2019), allowing bandwidths to vary for each outcome and on each side of the cutoff. Standard errors are clustered at the running variable level. Control group means are calculated using students with ELA M-STEP scores within 2 points above the cutoff. Column 1 reprints the estimates from our primary specification, as discussed in the paper. Column 2 uses a quadratic relationship between the running variable and the outcome, rather than a linear relationship. Column 3 drops all covariates from the baseline model. Column 4 adds two neighborhood-level covariates to the baseline model: (i) the share of the population age 25+ with a bachelor's degree, and (ii) the log of median household income. We exclude these two variables from our baseline specification because they are missing for 2.5% and 3.9% of our analysis sample, respectively. Column 5 uses the primary specification except that standard errors are computed using a heteroskedasticity-robust plug-in residuals variance estimator, without clustering. An IEP (individualized education program) is a document that outlines the programs and services a student with disabilities will receive to help them realize their educational goals, sometimes referred to as "special education." Neighborhood characteristics are calculated at the Census block group level.

Table A6: Summary Statistics for Third Graders in 2021 and 2022, by Presence of Retention in District-Year

	1+ Student on Each Side of Cutoff		10+ Students on Each Side of Cutoff	
	No Retention (1)	Has Retention (2)	No Retention (3)	Has Retention (4)
Student Characteristics				
2022 cohort	0.557	0.604	0.624	0.697
Female	0.490	0.492	0.488	0.489
White	0.736	0.536	0.668	0.447
Black	0.133	0.306	0.183	0.374
Hispanic	0.073	0.102	0.074	0.118
Asian	0.044	0.045	0.065	0.054
NA, NH, or PI	0.014	0.011	0.010	0.007
Economically disadvantaged	0.476	0.675	0.471	0.725
Has an IEP	0.146	0.143	0.144	0.140
English learner (EL)	0.065	0.130	0.096	0.184
Previously retained	0.020	0.037	0.017	0.037
Same district 2+ years	0.091	0.102	0.084	0.097
Math M-STEP score	1,295.4	1,285.8	1,293.8	1,281.9
ELA M-STEP score	1,295.1	1,286.6	1,293.3	1,283.5
ELA M-STEP score below 1253	0.043	0.085	0.056	0.110
Retained in third grade	0.001	0.019	0.001	0.015
Other Characteristics				
Third-graders in district-year	360	710	684	1,105
City	0.115	0.414	0.216	0.596
Rural	0.193	0.121	0.013	0.036
Suburb	0.539	0.373	0.678	0.334
Town	0.153	0.092	0.093	0.034
Charter school	0.081	0.156	0.030	0.082
School ED share	0.480	0.673	0.473	0.719
School IEP share	0.157	0.159	0.157	0.158
School EL share	0.062	0.119	0.090	0.168
School White share	0.731	0.532	0.662	0.444
School Black share	0.101	0.265	0.148	0.329
School Hispanic share	0.074	0.104	0.075	0.119
School Asian share	0.035	0.038	0.055	0.045
School NA/NH/PI share	0.005	0.005	0.003	0.002
Neighborhood BA+ share	0.187	0.144	0.202	0.144
Neighborhood median HH income	$65,716	$52,308	$70,235	$50,436
Districts	568	311	77	90
District×years	825	399	92	109
Students	97,825	58,087	30,296	32,323

Notes: As described further in Section VI.C, columns 1 and 2 (3 and 4) restrict the sample to district-years with at least 1 student (10 students) on each side of the cutoff within 18 points. A district-year is classified as "no retention" if no student within 18 points on either side of the cutoff is retained in third grade. "NA, NH, or PI" stands for Native American, Native Hawaiian, or Pacific Islander. An IEP (individualized education program) is a document that outlines the programs and services a student with disabilities will receive to help them realize their educational goals, sometimes referred to as "special education." Neighborhood characteristics are calculated at the Census block group level. Income is converted to real 2015 dollars.

Table A7: Regression-Adjusted Differences in Characteristics of District-Years With and Without Retention

	1+ Student on Each Side of Cutoff			10+ Students on Each Side of Cutoff		
	Difference	Standard Error	P-value	Difference	Standard Error	P-value
Female	-0.013	0.005	0.005	-0.018	0.006	0.003
White	0.161	0.024	0.000	0.181	0.050	0.000
Black	-0.162	0.024	0.000	-0.183	0.054	0.001
Hispanic	-0.011	0.007	0.138	-0.016	0.019	0.402
Asian	0.004	0.006	0.501	0.015	0.009	0.125
NA, NH, or PI	0.007	0.004	0.072	0.003	0.001	0.030
Economically disadvantaged	-0.134	0.015	0.000	-0.180	0.036	0.000
Has an IEP	0.016	0.006	0.009	0.005	0.010	0.628
English leaner (EL)	-0.009	0.011	0.397	-0.009	0.026	0.736
Previously retained	-0.020	0.003	0.000	-0.021	0.005	0.000
Same district 2+ years	-0.025	0.007	0.001	-0.027	0.009	0.003
Math M-STEP score	6.1	1.8	0.001	8.8	2.8	0.002
ELA M-STEP score	4.8	0.8	0.000	7.3	2.0	0.000
ELA M-STEP score below 1253	-0.018	0.004	0.000	-0.038	0.011	0.001
Third-graders in district-year	45	12	0.000	123	48	0.011
City	-0.140	0.029	0.000	-0.176	0.072	0.016
Rural	0.049	0.029	0.095	-0.054	0.031	0.081
Suburb	0.042	0.032	0.193	0.171	0.076	0.027
Town	0.050	0.023	0.030	0.059	0.047	0.211
Charter school	-0.225	0.031	0.000	-0.196	0.061	0.002
Neighborhood BA+ share	0.031	0.004	0.000	0.044	0.011	0.000
Neighborhood median HH income	8,767	1,079	0.000	12,968	2,921	0.000
District×years		1,224			201	

Notes: As described further in Section VI.C, the first (second) three columns restrict the sample to district-years with at least 1 student (10 students) on each side of the cutoff within 18 points. A district-year is classified as "no retention" if no student within 18 points on either side of the cutoff is retained in third grade. To conduct this analysis, we first collapse our student-level third-grader data to be at the district-year level. Each characteristic indicated in the first column of the table is a student-weighted mean, other than the number of third-graders in a district-year, which is a sum. Then, we regress each characteristic in the first column on a binary indicator for whether a district-year has no retention, controlling for year and a cubic polynomial for the number of retention-eligible third graders in the district-year. Standard errors are clustered at the district level. "NA, NH, or PI" stands for Native American, Native Hawaiian, or Pacific Islander. An IEP (individualized education program) is a document that outlines the programs and services a student with disabilities will receive to help them realize their educational goals, sometimes referred to as "special education." Neighborhood characteristics are calculated at the Census block group level. Income is converted to real 2015 dollars.

Figure A1: Letter From a District to Guardians About Grade Retention Intentions

Exceptional
ANN ARBOR PUBLIC SCHOOLS

Ann Arbor Public Schools Instructional Services Division

2555 S. State Street
Ann Arbor, MI

November 20, 2017

Dear Parents,

The Ann Arbor Public Schools is committed to providing a high quality education for all students. Ensuring that students are proficient and skillful readers and writers is one of our highest district priorities. In 2016, the State of Michigan passed legislation that immediately impacts reading expectations for all students in kindergarten through third grade. This legislation became Public Act 306 and is referred to as the **Third Grade Reading Law**. This Third Grade Reading Law sets out standards for student achievement and guidelines to ensure students achieve grade level proficiency by the end of third grade.

AAPS has been in compliance with many elements of this new law for many years--we are taking this opportunity to strengthen even further the district's reading and writing program. The law has several requirements that must be integrated into every school district's Pre-K through 3rd grade reading program over the next three years. We want to assure parents that the Ann Arbor Public Schools has a solid, robust plan to meet the requirements of the law and to provide strong and adaptive instructional support for young readers.

Attached to this cover letter is a general overview of the elements of the Third Grade Reading law and how parents/ families will be a strong partner in this achievement. Every student in grades K-5 will have a personalized learning plan (PLP) and all students who are not now grade level proficient in reading will have an Individual Reading Improvement Plan (IRIP). This plan will address the student's challenge areas and strategies for improving them. Teachers will actively engage parents in generating these plans to include effective "at home" reading activities.

Reading proficiency is a strong indicator of future career and college readiness. The Ann Arbor Public Schools continues to take steps to assure that literacy skills are embedded in all subject areas throughout the school day. Family engagement plays a vital role in a child's success as a reader. Your collaboration is both appreciated and necessary to ensure your student is a skillful reader and writer.

You may have heard the Third Grade Reading legislation includes language about third grade retention beginning in 2019-2020. We want to assure parents that the Ann Arbor Public Schools **does not** intend to implement this element. While the State of Michigan will monitor and report annually on a student's level of proficiency as measured by the MSTEP (Michigan Student Test of Educational Progress), local school districts have the authority to modify this provision if certain exemptions are documented. We firmly believe new enhancements to our core literacy program with additional supports and collaboration with families will ensure our K-3 students develop skills to be reading proficient.

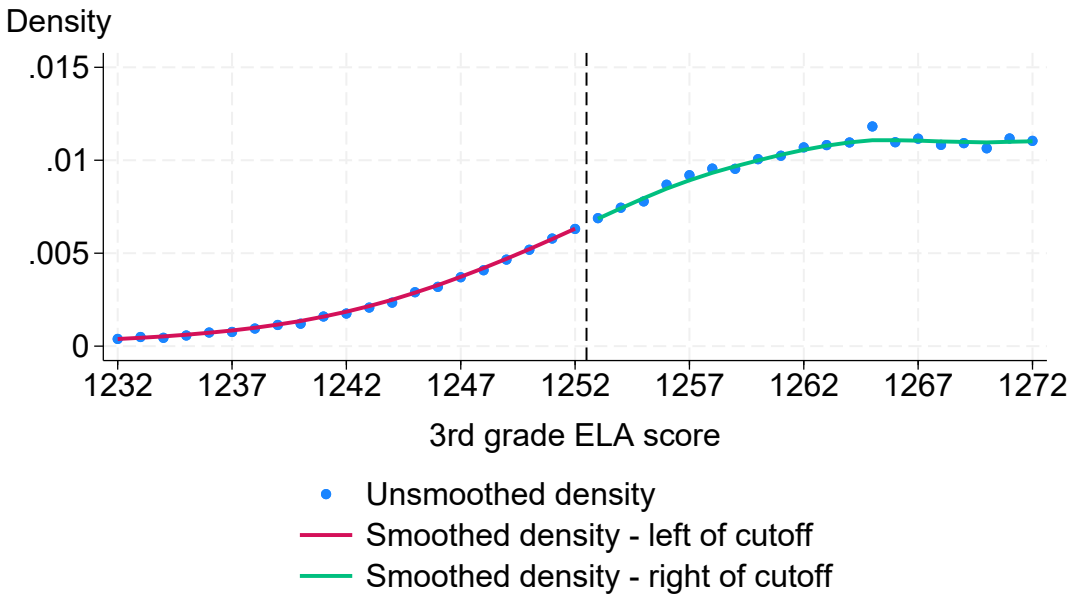
If you have any questions or concerns please do not hesitate to contact your student's principal or classroom teacher. As partners in your child's education we encourage you to communicate frequently with your child's teacher about his/her progress. Together, we will ensure our students will develop critical reading skills necessary to thrive in our diverse and global world.

In partnership,

Lee Ann Dickinson-Kelley
Assistant Superintendent for K-12 Instruction

Notes: The text in the red box indicates the intention of the Ann Arbor Public Schools to not retain children in third grade as part of the Read by Grade Three (RBG3) law. Note that the retention component of RBG3 did not begin with the 2019-20 cohort of third-graders as the letter says it would because of disruptions caused by the COVID-19 pandemic.

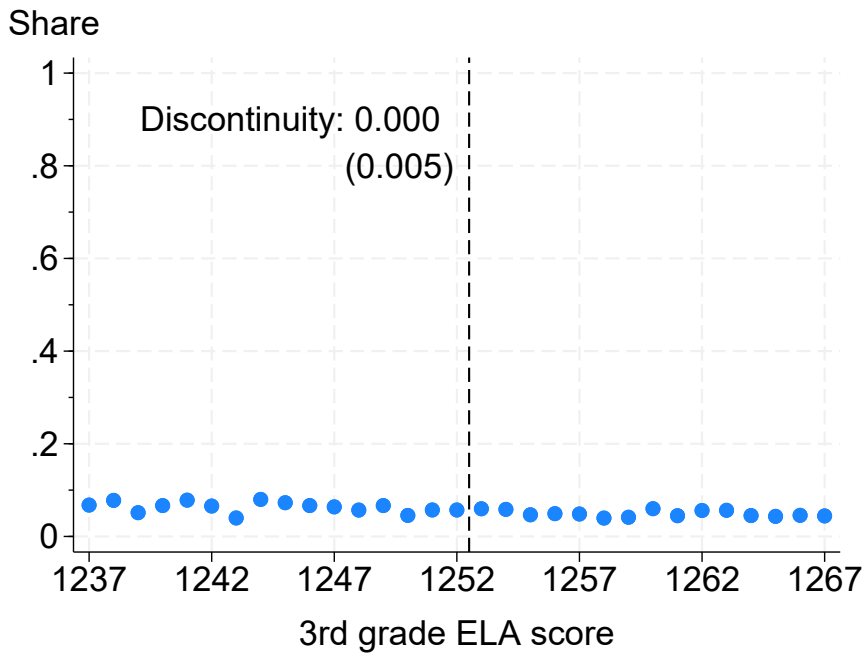
Figure A2: Third-Grade ELA Score Distribution



	T-statistic	P-value
Calonico et al. (2020)	0.429	0.668
McCrary (2008)	0.052	0.960

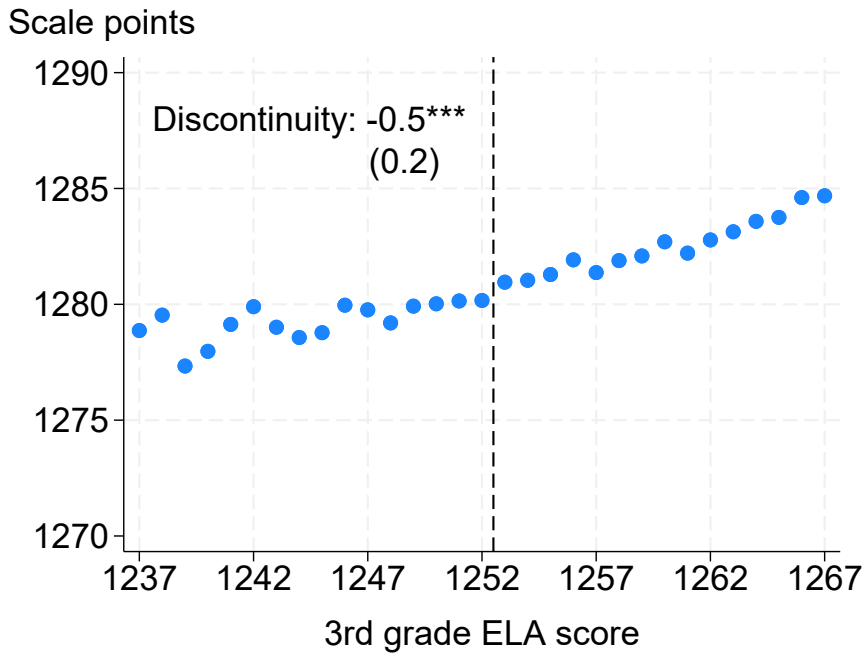
Notes: The sample includes first-time third-grade students with valid M-STEP scores the following year. In the Calonico et al. (2020) analysis, the bandwidth selection procedure selects 22.5 points for the bandwidth on both sides of the cutoff. In the McCrary (2008) analysis, we use a bandwidth of 5 points on both sides of the cutoff. Inference in the McCrary (2008) analysis is conducted using 1,000 bootstrapped samples. In the figure, the smoothed densities are obtained using the McCrary (2008) method with a 5-point bandwidth.

Figure A3: Share of Third Graders Without an M-STEP Score Next Year



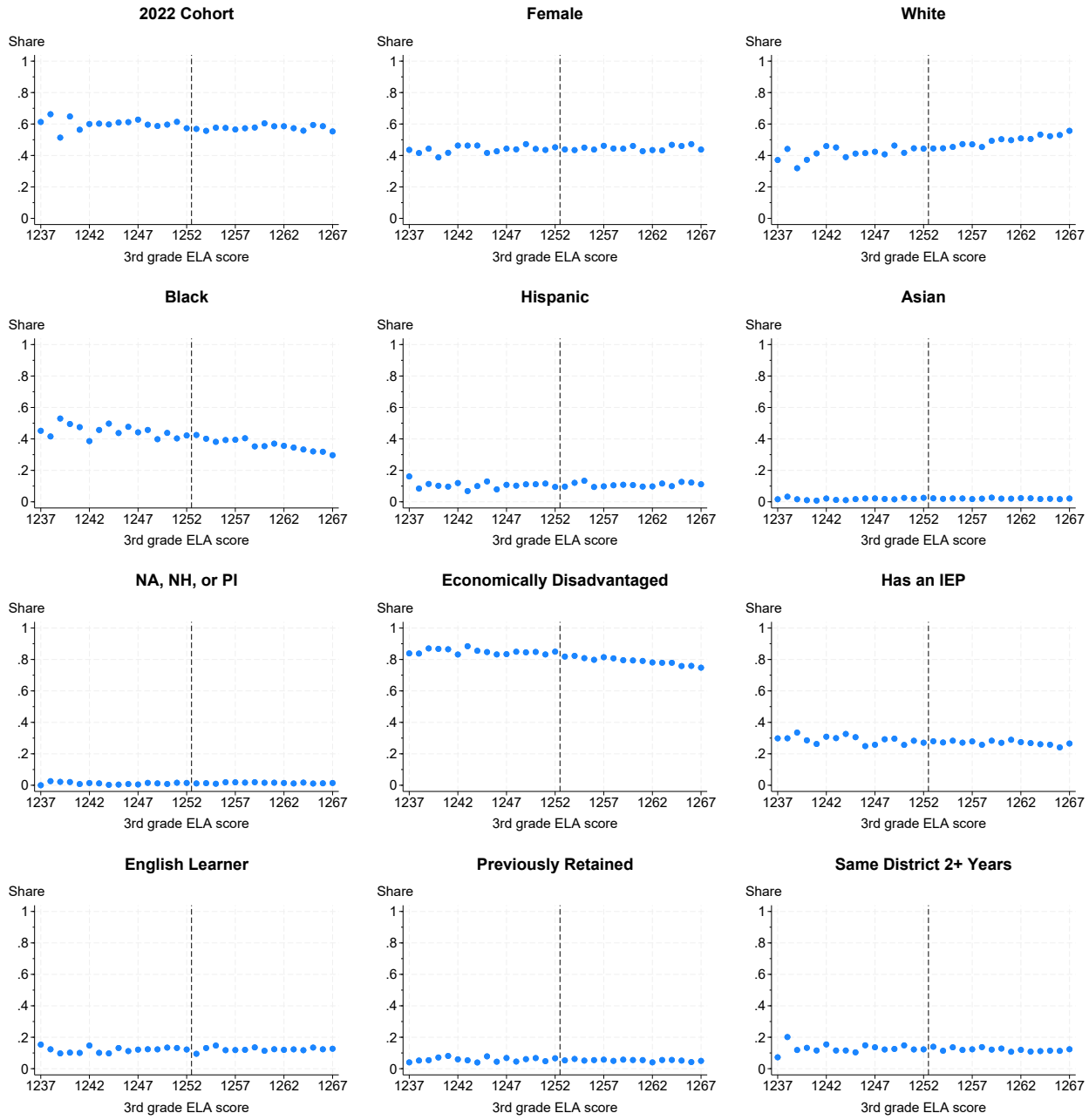
Notes: Each dot gives the share of students, for a given third-grade M-STEP score, that do not have a valid M-STEP score in the following year. This may occur either because a student did not take the test the following year or because they did not enroll in a Michigan public school the following year. The discontinuity estimate printed on the plot is estimated as described in Section V.A, but without including covariates. The number in parentheses is the standard error of the estimate (clustered at the level of the running variable). We denote statistical significance by *** for p<0.01, ** for p<0.05, and * for p<0.1.

Figure A4: Continuity in Predicted ELA Score Through the Cutoff



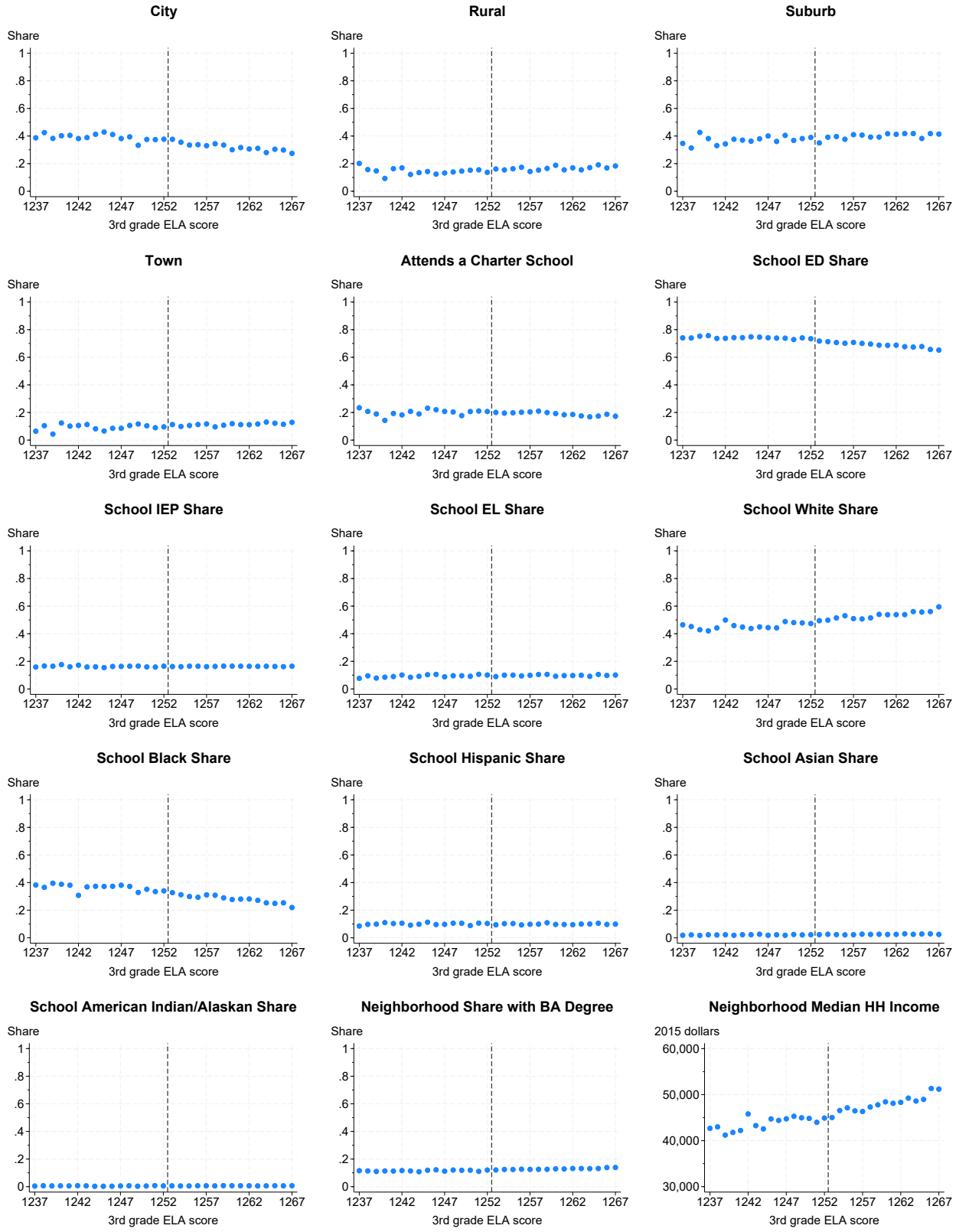
Notes: The sample includes first-time third-grade students with valid M-STEP scores the following year. Each dot gives the average predicted third-grade ELA M-STEP score for students with a given actual third-grade M-STEP ELA score. Predicted scores are computed by regressing third-grade ELA scores on the following covariates: sex, race, economic disadvantage (ED) status, baseline IEP status, baseline English learner (EL) status, being previously retained in-grade, being enrolled in a district for at least two years, attending a charter school, school-level race shares, school-level ED share, school-level IEP share, and school-level EL share. The discontinuity estimate printed on the plot is estimated as described in Section V.A, but without including covariates. The number in parentheses is the standard error of the estimate (clustered at the level of the running variable). We denote statistical significance by *** for $p<0.01$, ** for $p<0.05$, and * for $p<0.1$. The discontinuity estimate is robust to alternative estimation specifications.

Figure A5: Covariate Continuity Through the Cutoff I



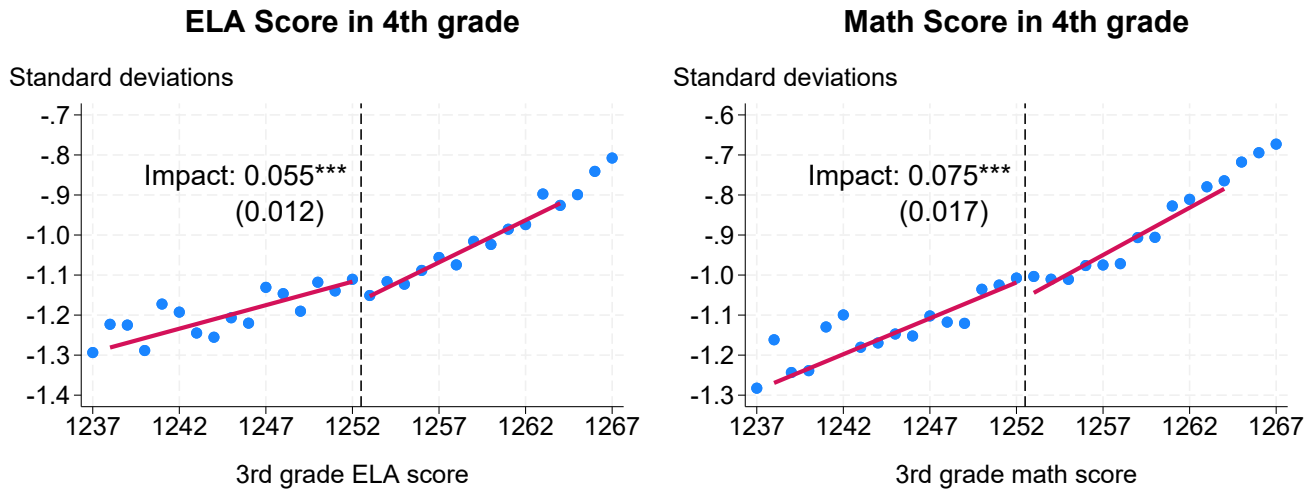
Notes: The sample includes first-time third-grade students with valid M-STEP scores the following year. Each dot gives the share of students with a given characteristics for a given third-grade M-STEP ELA score. "NA, NH, or PI" stands for Native American, Native Hawaiian, or Pacific Islander. An IEP (individualized education program) is a document that outlines the programs and services a student with disabilities will receive to help them realize their educational goals, sometimes referred to as "special education."

Figure A6: Covariate Continuity Through the Cutoff II



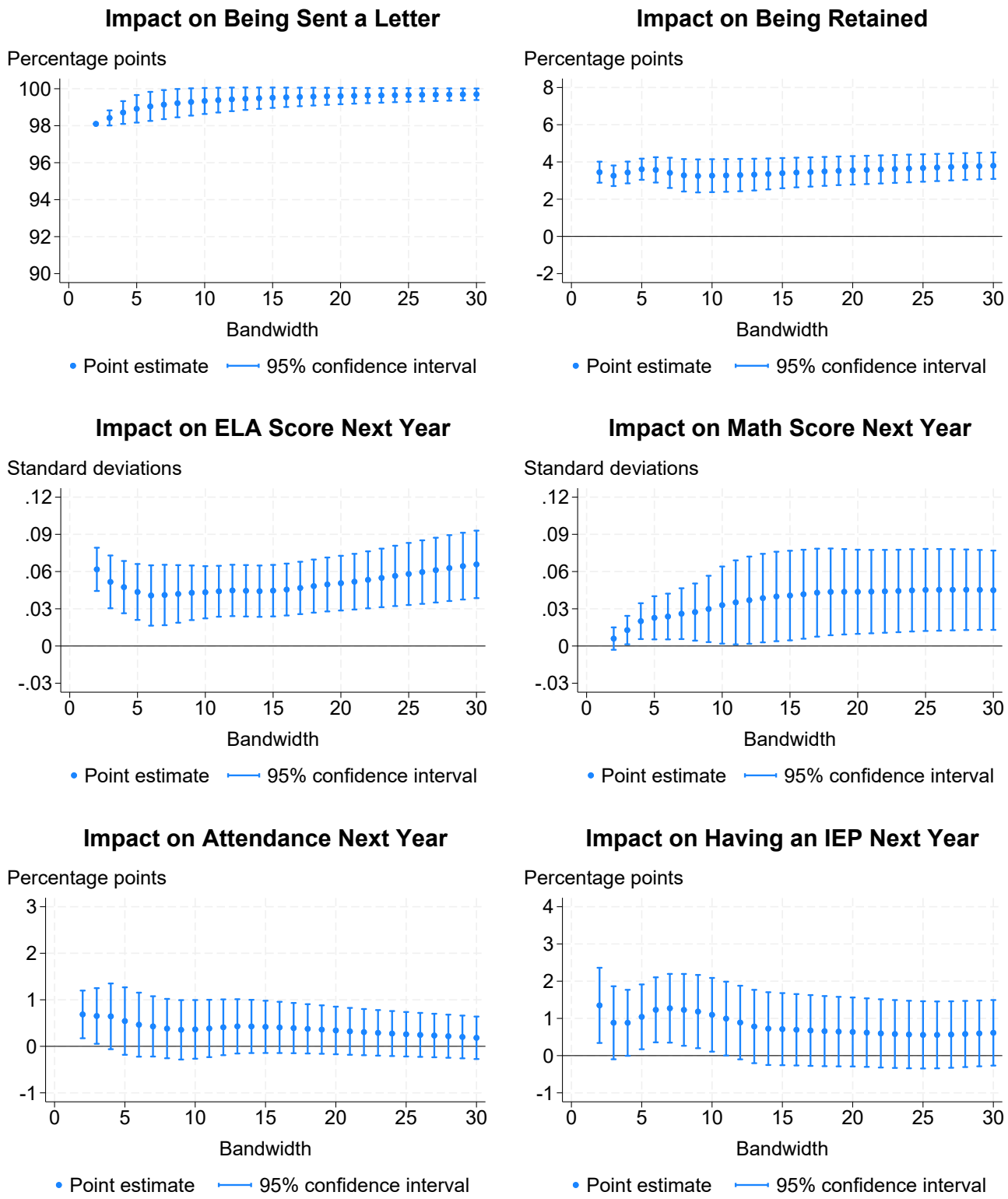
Notes: The sample includes first-time third-grade students with valid M-STEP scores the following year. Each dot gives the share of students with a given characteristics for a given third-grade M-STEP ELA score. An IEP (individualized education program) is a document that outlines the programs and services a student with disabilities will receive to help them realize their educational goals, sometimes referred to as "special education." Neighborhood characteristics are calculated at the Census block group level.

Figure A7: "Same-Grade" ITT Impact Estimates



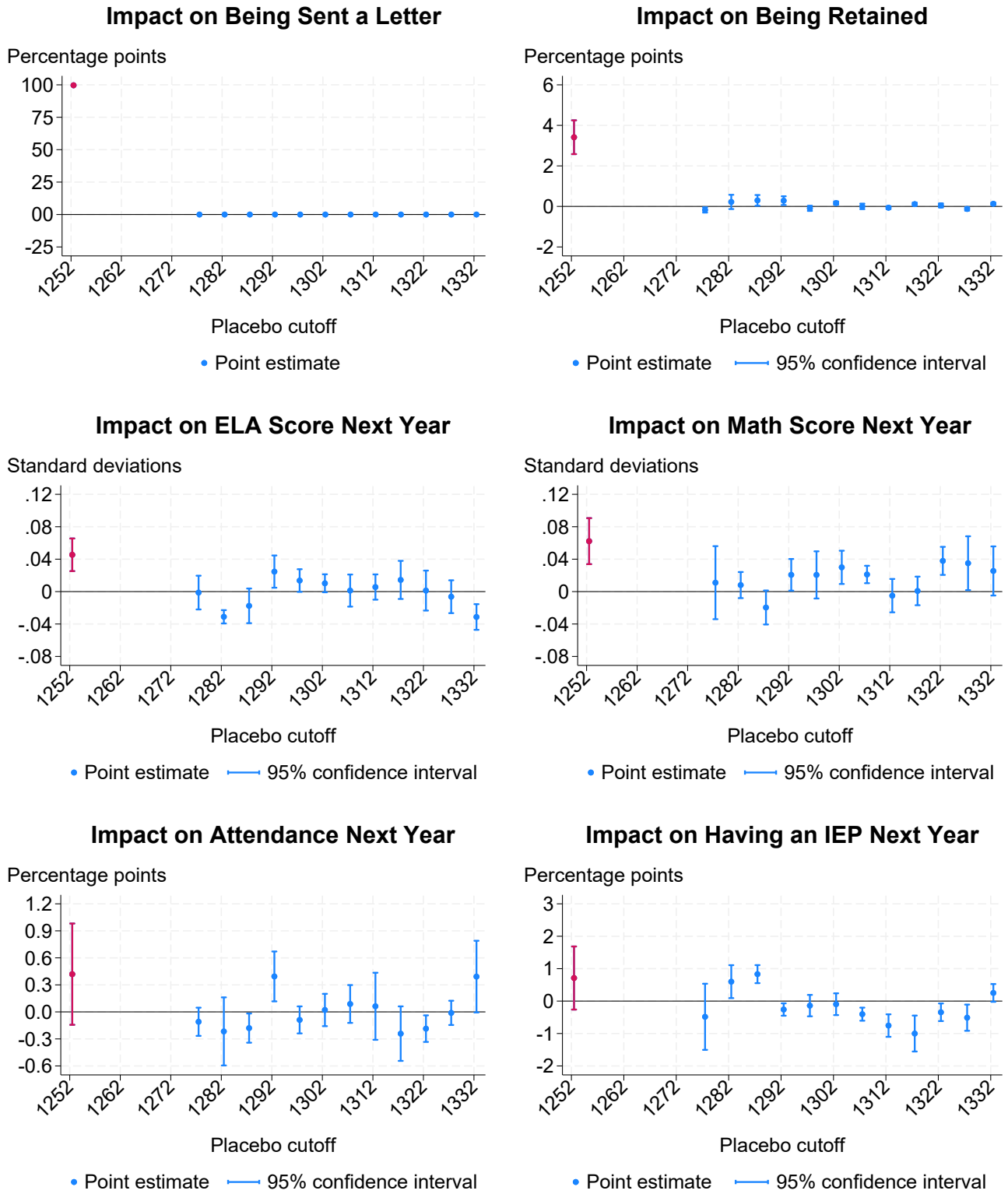
Notes: The impact estimates printed on each plot are estimated as described in Section V.A. The numbers in parentheses below the impact estimates are standard errors (clustered at the level of the running variable). We denote statistical significance by *** for p<0.01, ** for p<0.05, and * for p<0.1.

Figure A8: Impact Estimates with Alternative Bandwidths



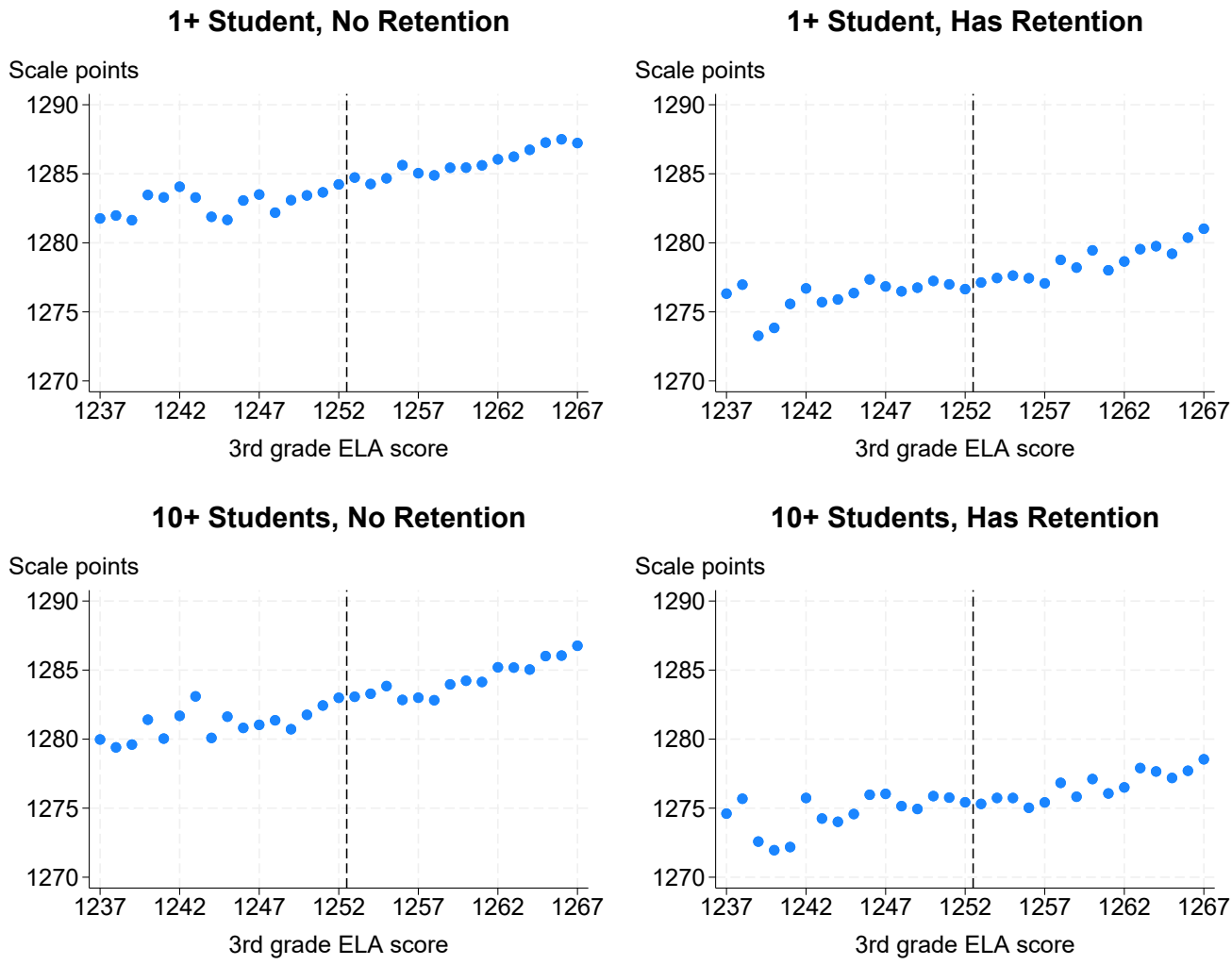
Notes: Each estimate in these plots comes from an RD regression using an alternative bandwidth. For each value X on the x-axis, we re-estimate impacts using the estimation procedure described in Section V.A, except that we impose a bandwidth of (-X,X). Standard errors are clustered at the running variable level.

Figure A9: Impact Estimates with Placebo Cutoffs



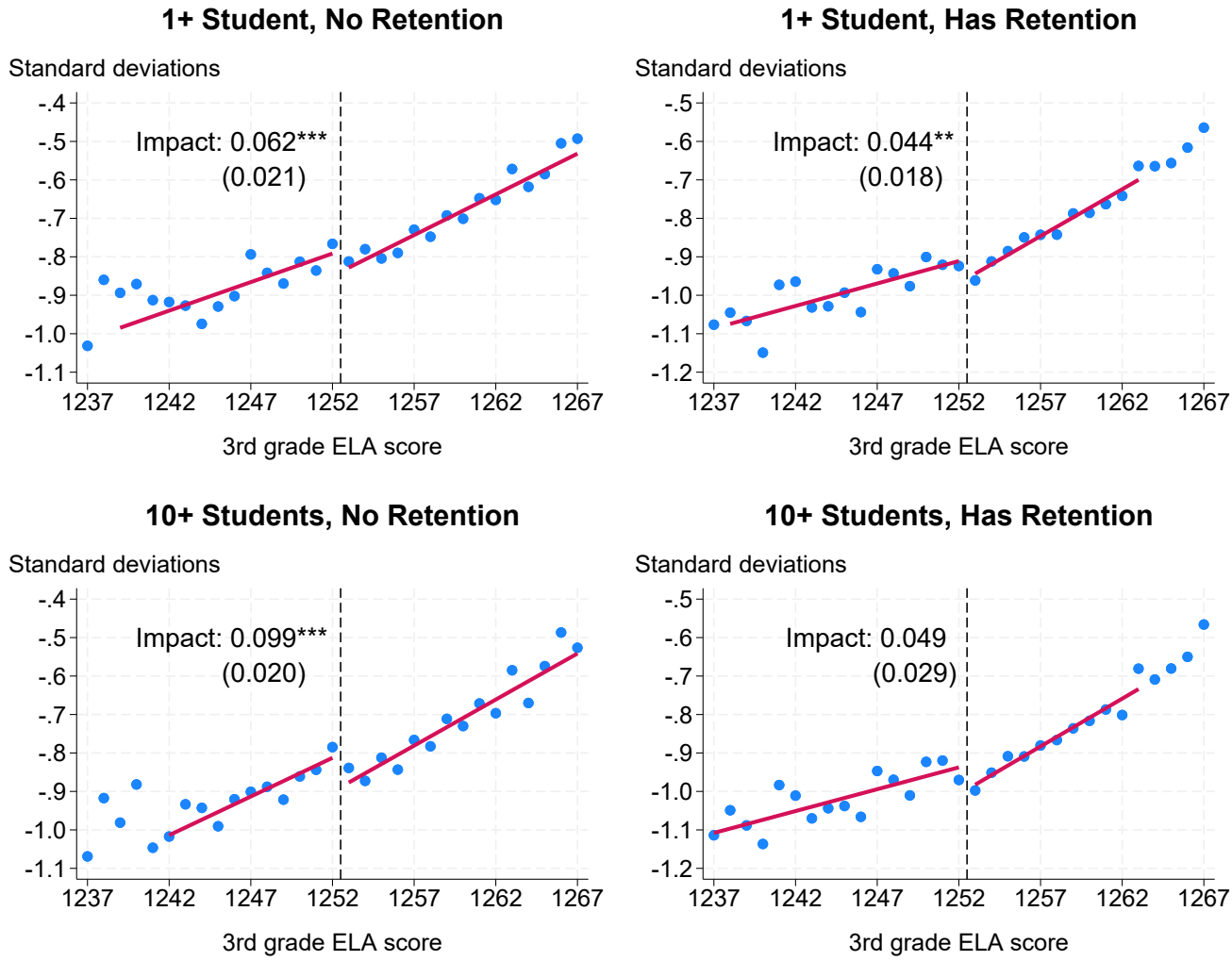
Notes: Each estimate in these plots comes from an RD regression using the associated x-axis value as the (potentially placebo) retention cutoff. Our primary estimates using the true cutoff (1252) are shown in red. The placebo cutoffs range (highly evenly) from the 31st percentile of the M-STEP distribution to the 93rd percentile. We do not estimate regressions with placebo cutoffs less than 1252 due to small sample sizes at the bottom of the distribution. We do not use values between 1252 and 1277 as placebo cutoffs so that scores near the true cutoff do not influence the placebo estimates. For each placebo model, we limit the sample to observations within 25 points of the placebo cutoff. We then estimate models using the procedure developed by Calonico et al. (2018, 2020, 2014, 2019), allowing bandwidths to vary for each outcome and on each side of the cutoff. Standard errors are clustered at the running variable level.

Figure A10: Continuity in Predicted ELA Score Through the Cutoff, by Presence of Retention in District-Year



Notes: The samples are restricted to district-years with at least 1 student (10 students) on each side of the cutoff within 18 points. A district-year is classified as "no retention" if no student within 18 points on either side of the cutoff is retained in third grade. Only the "1+ Student, Has Retention" sample has a statistically significant discontinuity in predicted ELA score, although it is extremely small (-0.039 scale points, $p = 0.077$) and in the opposite direction of our estimated impacts.

Figure A11: ITT Impact Estimates for ELA Test Scores Next Year, by Presence of Retention in District-Year



Notes: The impact estimates printed on each plot are estimated as described in Section V.A. The numbers in parentheses below the impact estimates are standard errors (clustered at the level of the running variable). The samples are restricted to district-years with at least 1 student (10 students) on each side of the cutoff within 18 points. A district-year is classified as "no retention" if no student within 18 points on either side of the cutoff is retained in third grade. We denote statistical significance by *** for p<0.01, ** for p<0.05, and * for p<0.1.

Figure A12: Retention Letter Template in School Year 2022-23

STATE OF MICHIGAN

[DATE]

Dear Parent/Guardian:

We are writing in regard to your child, [INSERT STUDENT'S NAME].

In 2016, the Michigan Legislature passed Public Act 306 of 2016, MCL 380.1280f (commonly referred to as the Read By Grade Three Law), which requires schools to identify students in kindergarten through third grade who are struggling with reading and writing and provide them with additional support services. The law also requires schools to retain a student in the third grade if the student's English language arts score is 1252 or below, indicating more than one year behind on reading according to the Michigan Student Test of Educational Progress (M-STEP). This year's class of third graders will be the last students affected by the retention aspects of the law because Public Act 7 of 2023 repeals the retention requirement as of the 2023-24 school year. For now, retention is still a requirement.

According to the Michigan Department of Education, [INSERT STUDENT'S FIRST NAME]'s M-STEP score may require them to repeat 3rd grade. However, you have the right to request a good cause exemption within 30 days of this letter if you wish for your child to continue to the 4th grade this fall. You also have the right to request a meeting with your child's school to discuss the retention requirement and the standards and processes for a good cause exemption.

We understand that this may be difficult news to hear, but your child's school is ready and willing to help you understand your options. **Please contact the school office right away to discuss the best path forward for [INSERT STUDENT'S FIRST NAME].**

Even if your child is behind in reading, the law permits a good cause exemption that would allow your child to advance to the 4th grade, if:

- You requested a good cause exemption and your child's school determines that being promoted to the 4th grade is in the best interest of the child;
- Your child has an accommodation plan in place such as an individualized education program (IEP) or 504 Plan;
- English is not your child's first language and they have been in an English Language Learner Program for fewer than three years;
- Your child received intensive reading interventions for two or more years, and was previously retained in kindergarten, first or second grade; or,
- Your child has been in their current school for fewer than two years and did not receive an appropriate individual reading improvement plan.

The State of Michigan supports you in making a decision that is right for your family. More information and Spanish and Arabic translations of this letter can be found at: **mi.gov/earlyliteracy**.

Para recibir una copia de esta carta en español, contacte a la escuela de su hijo/a que está en tercer grado o visite al siguiente enlace: mi.gov/earlyliteracy.

للحصو ل على نسخة من هذه الرسالة باللغة العربية ، يرجى الاتصال بإدارة مدرسة الطالب للصف الثالث الإبتدائي أو الذهاب إلى الموقع التالي:mi.gov/earlyliteracy.

Notes: This is a template for the letter the state of Michigan sent to guardians of third graders flagged for retention in the 2022-23 school year.

B Additional Tables and Figures (Separately by Cohorts)

Table B1: Summary Statistics for Third Graders, 2021 and 2022 Cohorts

	2021 Cohort			2022 Cohort		
	All First-Time Third-Graders	Analysis Sample	Analysis Sample Within Bandwidth	All First-Time Third-Graders	Analysis Sample	Analysis Sample Within Bandwidth
Student Characteristics						
Female	0.486	0.490	0.442	0.488	0.491	0.449
White	0.634	0.720	0.559	0.627	0.636	0.423
Black	0.222	0.148	0.304	0.226	0.220	0.431
Hispanic	0.084	0.075	0.101	0.087	0.086	0.113
Asian	0.046	0.043	0.021	0.048	0.046	0.019
NA, NH, or PI	0.013	0.014	0.015	0.012	0.012	0.013
Economically disadvantaged	0.562	0.514	0.780	0.566	0.561	0.817
Has an IEP	0.153	0.140	0.271	0.162	0.148	0.273
English learner (EL)	0.092	0.084	0.127	0.091	0.087	0.120
Previously retained	0.033	0.025	0.053	0.029	0.027	0.055
Same district 2+ years	0.088	0.084	0.102	0.112	0.105	0.137
Math M-STEP score	-	1,293.0	1,262.2	-	1,292.4	1,260.8
ELA M-STEP score	-	1,293.1	1,257.5	-	1,292.4	1,257.3
ELA M-STEP score below 1253	-	0.050	0.233	-	0.058	0.250
Retained in third grade	0.011	0.006	0.021	0.011	0.009	0.030
Other Characteristics						
Third-graders in district-year	533	389	378	527	520	721
City	0.251	0.170	0.259	0.252	0.247	0.387
Rural	0.173	0.209	0.194	0.172	0.174	0.137
Suburb	0.456	0.478	0.416	0.456	0.459	0.381
Town	0.120	0.144	0.131	0.119	0.121	0.095
Charter school	0.123	0.109	0.189	0.123	0.122	0.194
School ED share	0.560	0.522	0.668	0.563	0.559	0.722
School IEP share	0.157	0.154	0.162	0.163	0.160	0.164
School EL share	0.082	0.074	0.095	0.085	0.084	0.100
School White share	0.634	0.709	0.592	0.631	0.635	0.466
School Black share	0.184	0.121	0.230	0.184	0.180	0.345
School Hispanic share	0.086	0.076	0.090	0.087	0.087	0.106
School Asian share	0.037	0.035	0.026	0.037	0.037	0.023
School NA/NH/PI share	0.005	0.006	0.007	0.005	0.005	0.005
Neighborhood BA+ share	0.169	0.176	0.132	0.169	0.169	0.122
Neighborhood median HH income	$60,001	$62,685	$49,467	$60,018	$60,273	$45,712
Districts	818	785	739	812	786	757
Students	104,995	72,236	13,087	103,284	96,207	18,198

Notes: The final column restricts the sample to students scoring within ±15 points of the retention cutoff. "NA, NH, or PI" stands for Native American, Native Hawaiian, or Pacific Islander. An IEP (individualized education program) is a document that outlines the programs and services a student with disabilities will receive to help them realize their educational goals, sometimes referred to as "special education." Neighborhood characteristics are calculated at the Census block group level. Income is converted to real 2015 dollars.

Table B2: ITT Impact Estimates

	2021 Cohort		2022 Cohort	
	Control Mean	Impact Estimate (SE)	Control Mean	Impact Estimate (SE)
Analysis Sample				
Enrolled next year	0.994	-0.002*	0.992	0.004***
		(0.002)		(0.002)
Took M-STEP next year	0.940	0.001	0.942	0.000
		(0.006)		(0.007)
Took M-STEP in fourth grade	0.941	-0.001	0.938	0.000
		(0.008)		(0.006)
Conditional on Enrollment Next Year				
Retained (next year)	0.019	0.032***	0.032	0.037***
		(0.004)		(0.008)
Attendance rate (next year)	0.888	0.002	0.880	0.006**
		(0.004)		(0.004)
Has an IEP (next year)	0.331	0.015**	0.322	0.002
		(0.007)		(0.008)
Switched schools (next year)	0.183	0.025***	0.170	-0.009
		(0.009)		(0.009)
Switched districts (next year)	0.107	0.010	0.094	0.003
		(0.010)		(0.008)
Same-Age Test Score Impacts				
Math score (next year)	-0.698	0.057**	-0.873	0.050**
		(0.031)		(0.031)
ELA score (next year)	-0.801	0.034*	-0.909	0.057**
		(0.018)		(0.026)
Same-Grade Test Score Impacts				
Math score (fourth grade)	-0.883	0.068**	-1.103	0.062***
		(0.034)		(0.027)
ELA score (fourth grade)	-1.063	0.039**	-1.188	0.071***
		(0.018)		(0.025)

*** p<0.01, ** p<0.05, * p<0.1

Notes: We estimate models using the procedure developed by Calonico et al. (2018, 2020, 2014, 2019), allowing bandwidths to vary for each outcome and on each side of the cutoff. Standard errors are clustered at the running variable level. Control group means are calculated using students with ELA M-STEP scores within 2 points above the cutoff. An IEP (individualized education program) is a document that outlines the programs and services a student with disabilities will receive to help them realize their educational goals, sometimes referred to as "special education."

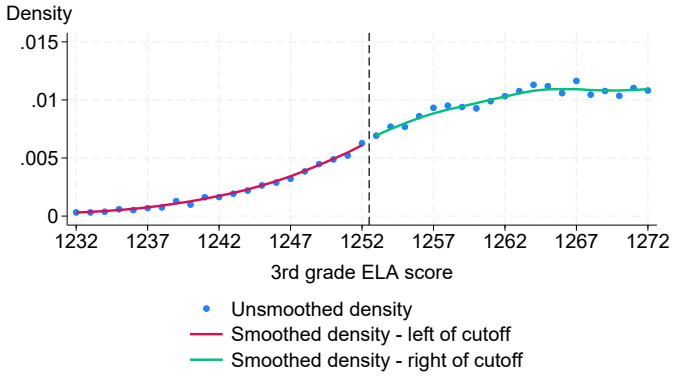
Table B3: Tests for Covariate Continuity Through the Cutoff

	2021 Cohort			2022 Cohort		
	Control Mean	Point Estimate	P-value	Control Mean	Point Estimate	P-value
Composite						
Predicted ELA M-STEP score	1,283.0	-0.7	0.115	1,279.4	-0.2	0.321
Student-Level Characteristics						
Female	0.438	0.006	0.850	0.435	0.018	0.073
White	0.528	0.007	0.963	0.381	0.014	0.458
Black	0.326	-0.007	0.831	0.480	-0.016	0.097
Hispanic	0.109	-0.006	0.717	0.109	0.001	0.899
Asian	0.022	0.005	0.300	0.020	0.003	0.162
NA, NH, or PI	0.016	0.000	0.682	0.010	0.001	0.115
Economically disadvantaged	0.807	0.021	0.061	0.831	0.012	0.025
Has an IEP	0.284	-0.003	0.490	0.270	-0.008	0.737
English learner (EL)	0.118	0.010	0.348	0.111	0.004	0.509
Previously retained	0.065	0.000	0.942	0.052	0.008	0.083
Same district 2+ years	0.117	-0.019	0.077	0.134	0.009	0.479
Other Characteristics						
City	0.275	0.031	0.016	0.435	-0.023	0.009
Rural	0.189	-0.011	0.577	0.133	-0.001	0.840
Suburb	0.399	-0.014	0.160	0.350	0.029	0.004
Town	0.137	-0.002	0.950	0.081	-0.003	0.684
Charter school	0.191	0.010	0.506	0.203	-0.010	0.659
School ED share	0.675	0.031	0.000	0.746	0.001	0.858
School IEP share	0.161	0.002	0.223	0.162	-0.003	0.339
School EL share	0.090	0.004	0.194	0.098	0.004	0.316
School White share	0.583	-0.017	0.326	0.430	-0.001	0.759
School Black share	0.238	0.016	0.510	0.382	0.001	0.761
School Hispanic share	0.089	0.003	0.162	0.106	0.004	0.409
School Asian share	0.027	-0.003	0.141	0.022	0.001	0.463
School NA/NH/PI share	0.006	0.000	0.410	0.005	0.000	0.470
Neighborhood BA+ share	0.129	-0.002	0.508	0.118	-0.005	0.078
Neighborhood median HH income	48,071	-989	0.036	43,990	-375	0.104

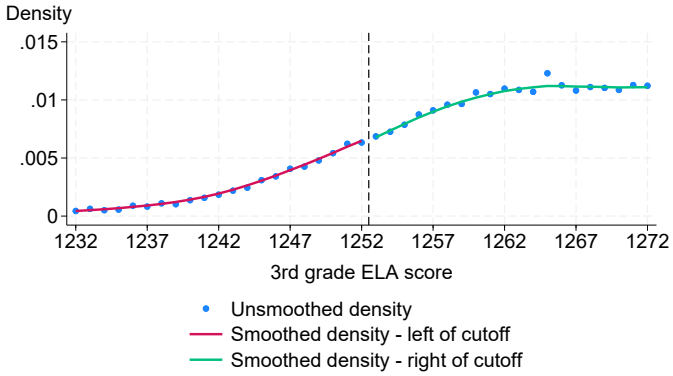
Notes: The sample includes first-time third-grade students with valid M-STEP scores the following year. We estimate models using the procedure developed by Calonico et al. (2018, 2020, 2014, 2019), allowing bandwidths to vary for each outcome and on each side of the cutoff. Standard errors are clustered at the running variable level. Control group means are calculated using students with ELA M-STEP scores within 2 points above the cutoff. "NA, NH, or PI" stands for Native American, Native Hawaiian, or Pacific Islander. An IEP (individualized education program) is a document that outlines the programs and services a student with disabilities will receive to help them realize their educational goals, sometimes referred to as "special education." Neighborhood characteristics are calculated at the Census block group level. Income is converted to real 2015 dollars.

Figure B1: Third-Grade ELA Score Distribution

(a) 2021 Cohort



(b) 2022 Cohort



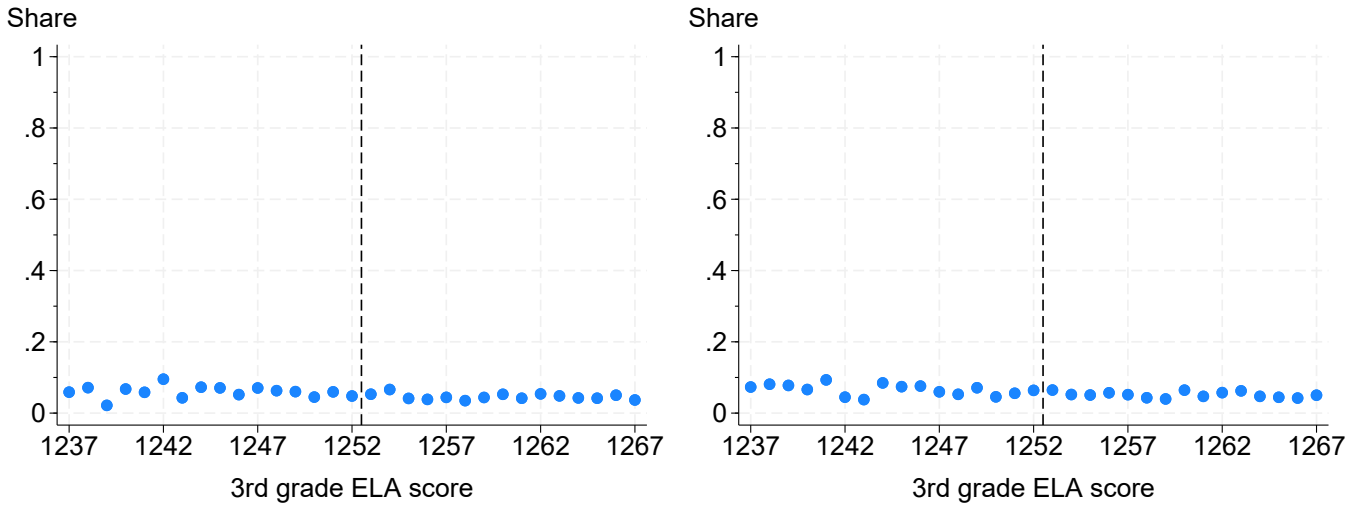
	2021 Cohort		2022 Cohort	
	T-statistic	P-value	T-statistic	P-value
Calonico et al. (2020)	-1.262	0.207	1.633	0.102
McCrary (2008)	-0.672	0.527	0.649	0.541

Notes: The sample includes first-time third-grade students with valid M-STEP scores the following year. In the Calonico et al. (2020) analysis, the bandwidth selection procedure selects 22.5 points for the bandwidth on both sides of the cutoff. In the McCrary (2008) analysis, we use a bandwidth of 5 points on both sides of the cutoff. Inference in the McCrary (2008) analysis is conducted using 1,000 bootstrapped samples. In the figure, the smoothed densities are obtained using the McCrary (2008) method with a 5-point bandwidth.

Figure B2: Share of Third Graders Without an M-STEP Score Next Year

(a) 2021 Cohort

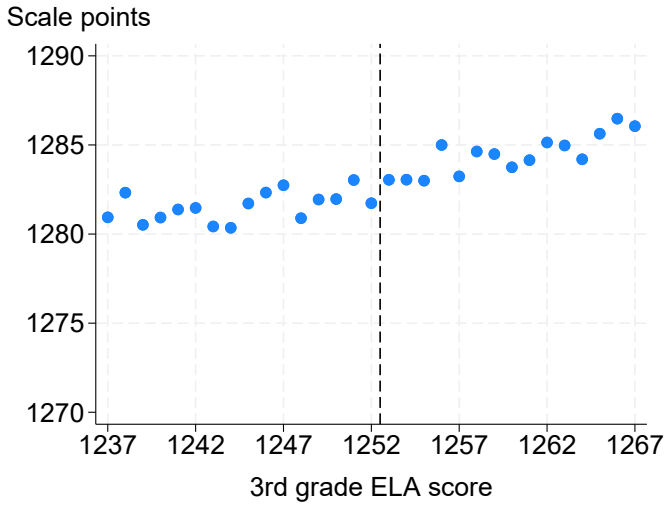
(b) 2022 Cohort



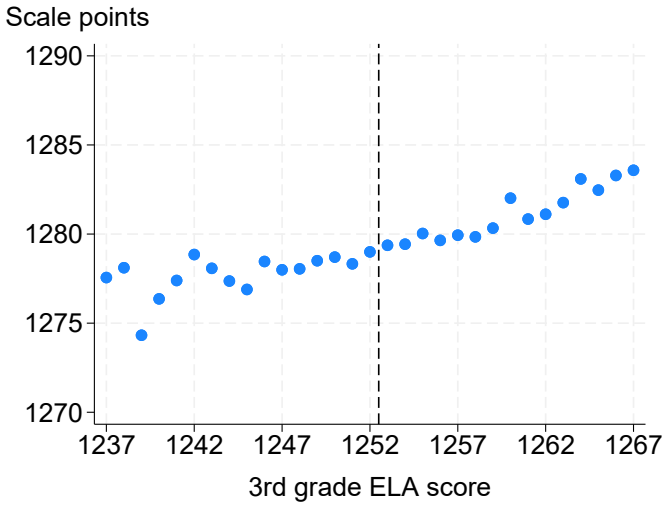
Notes: Each dot gives the share of students, for a given third-grade M-STEP score, that do not have a valid M-STEP score in the following year. This may occur either because a student did not take the test the following year or because they did not enroll in a Michigan public school the following year.

Figure B3: Continuity in Predicted M-STEP Scores Through the Cutoff

(a) 2021 Cohort

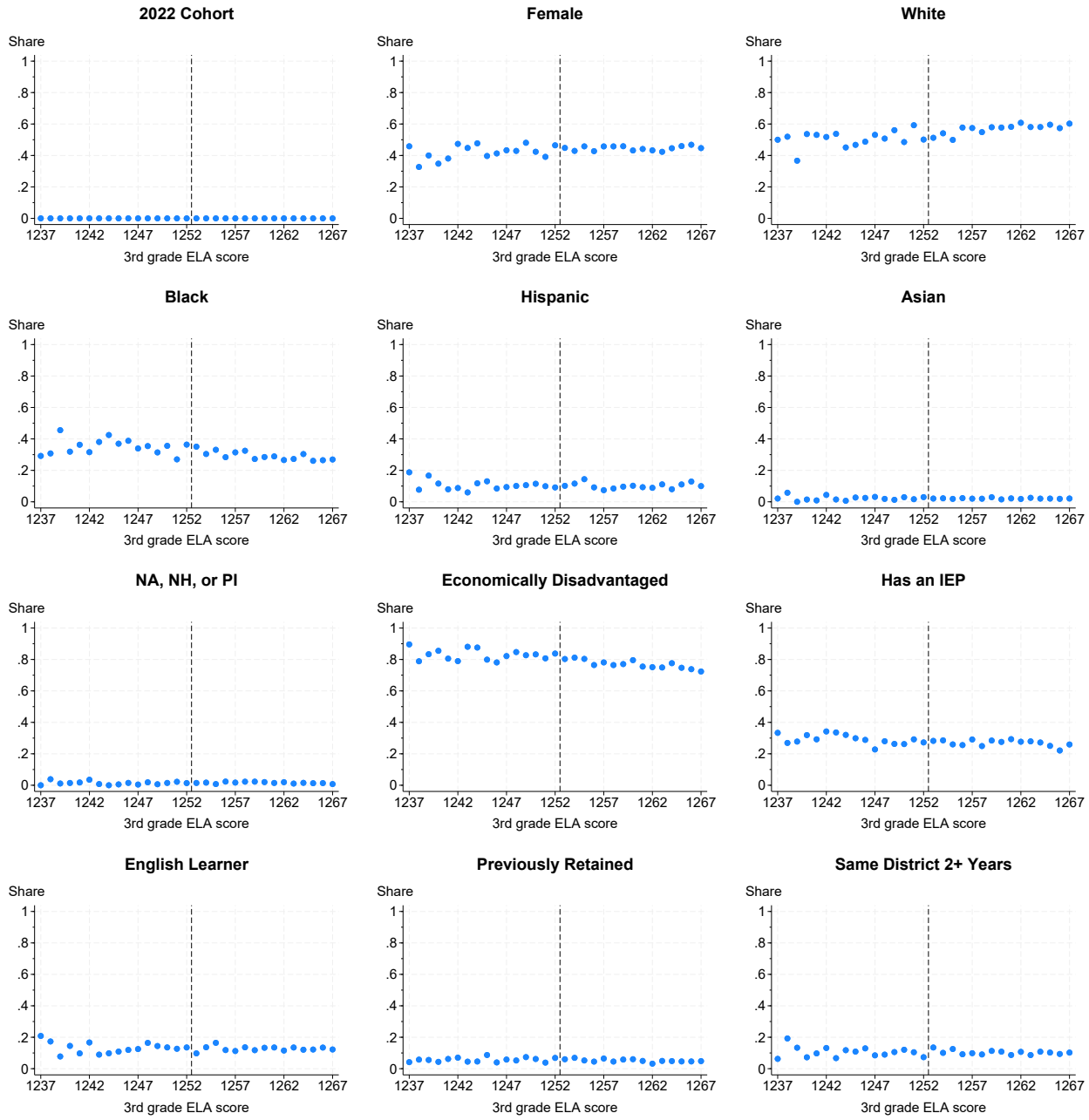


(b) 2022 Cohort



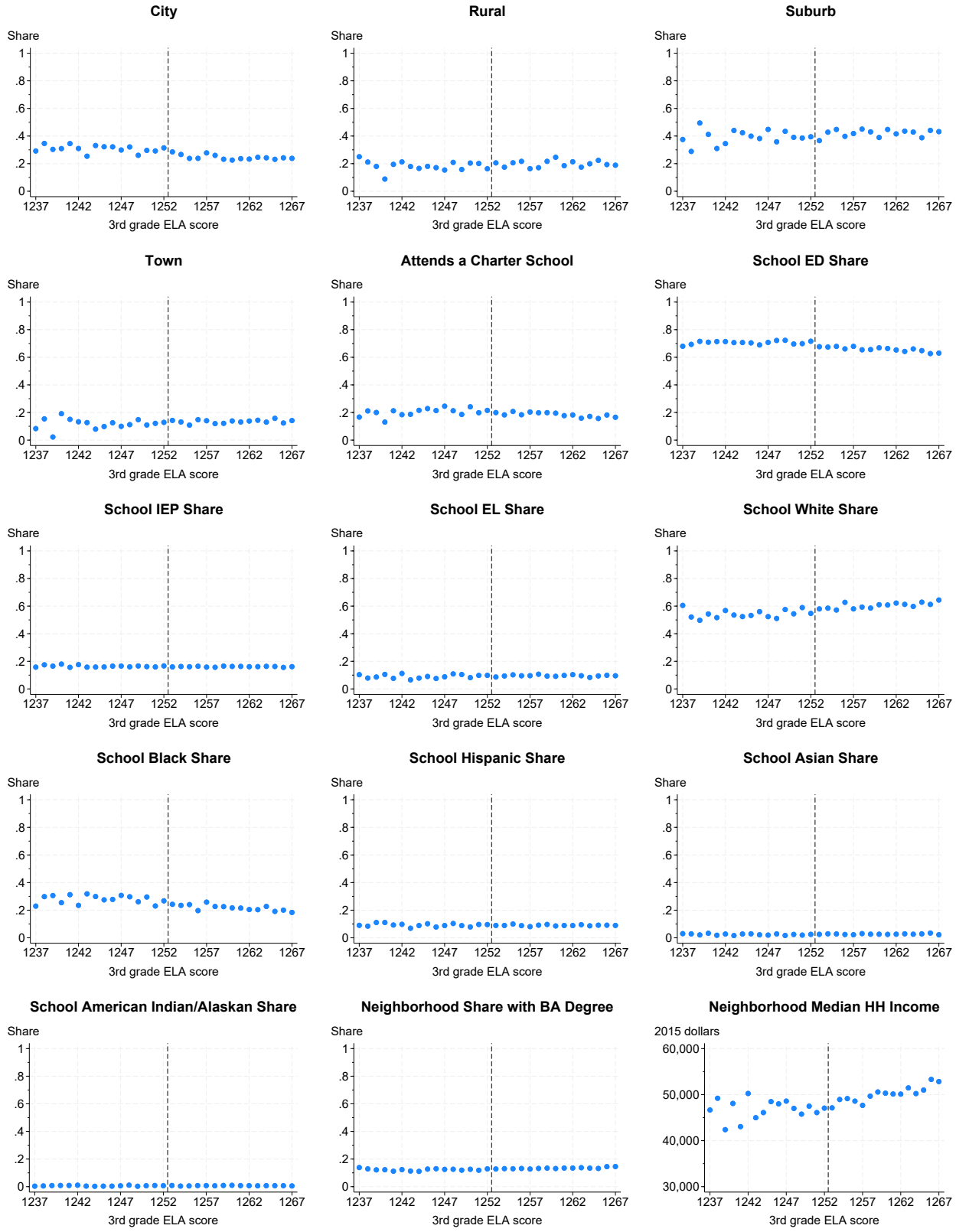
Notes: The sample includes first-time third-grade students with valid M-STEP scores the following year. Each dot gives the average predicted third-grade ELA M-STEP score for students with a given actual third-grade M-STEP ELA score. Predicted scores are computed by regressing third-grade ELA scores on the following covariates: sex, race, economic disadvantage (ED) status, baseline IEP status, baseline English learner (EL) status, being previously retained in-grade, being enrolled in a district for at least two years, attending a charter school, school-level race composition, school-level ED share, school-level IEP share, and school-level EL share.

Figure B4: Covariate Continuity Through the Cutoff I, 2021 Cohort



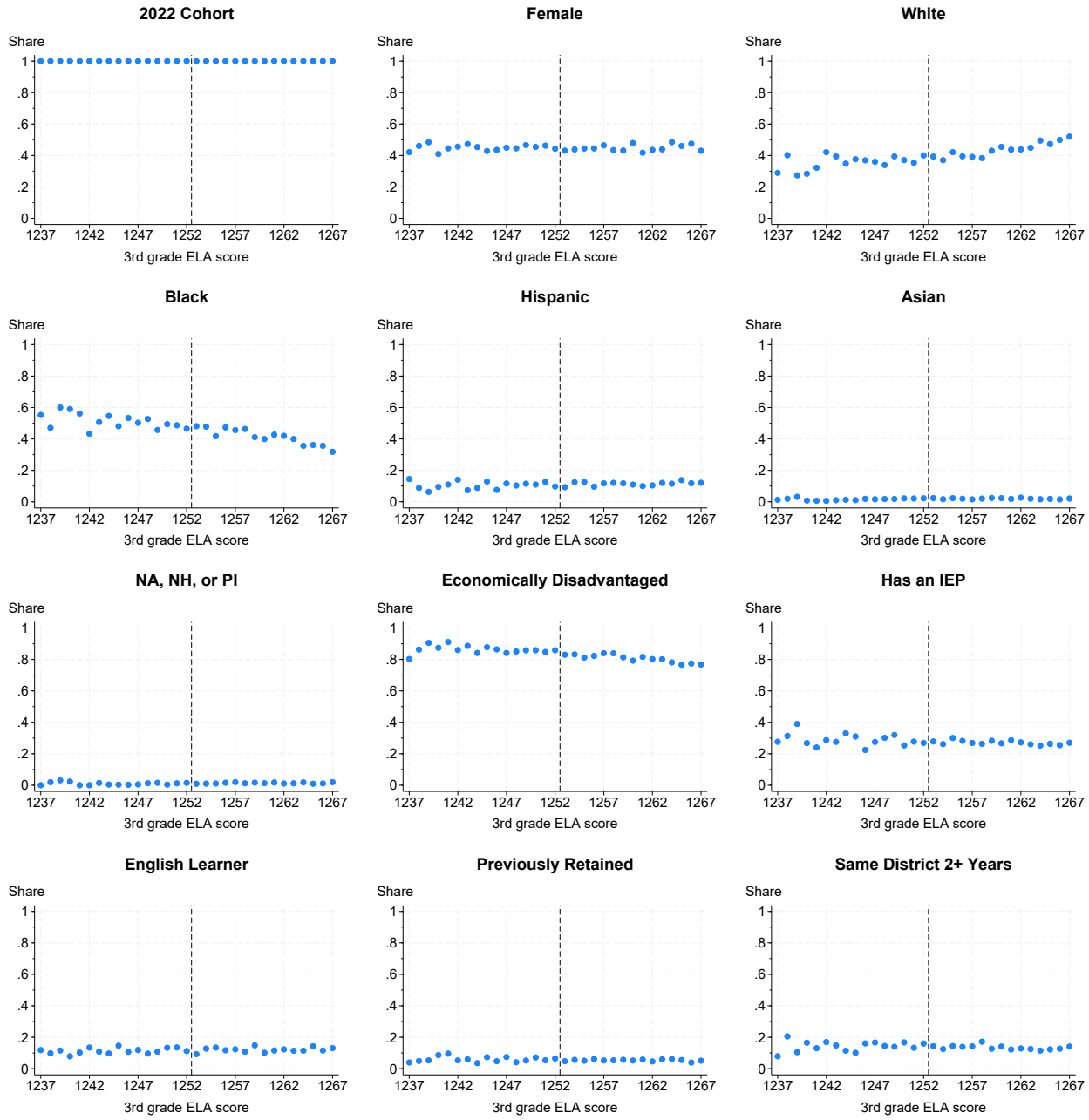
Notes: The sample includes first-time third-grade students in the 2021 cohort with valid M-STEP scores the following year. Each dot gives the share of students with a given characteristics for a given third-grade M-STEP ELA score. "NA, NH, or PI" stands for Native American, Native Hawaiian, or Pacific Islander. An IEP (individualized education program) is a document that outlines the programs and services a student with disabilities will receive to help them realize their educational goals, sometimes referred to as "special education."

Figure B5: Covariate Continuity Through the Cutoff II, 2021 Cohort



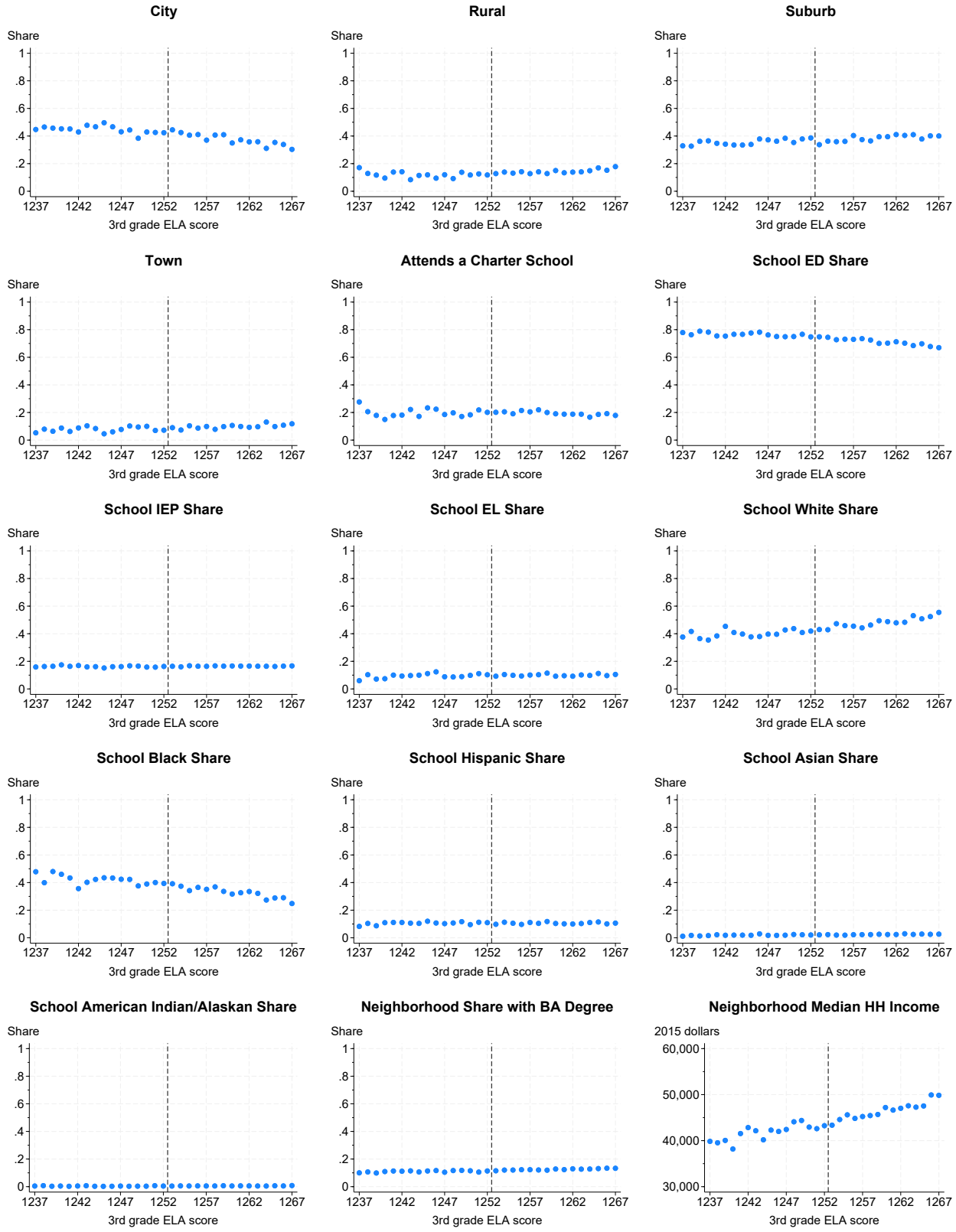
Notes: The sample includes first-time third-grade students in the 2021 cohort with valid M-STEP scores the following year. Each dot gives the share of students with a given characteristics for a given third-grade M-STEP ELA score. An IEP (individualized education program) is a document that outlines the programs and services a student with disabilities will receive to help them realize their educational goals, sometimes referred to as "special education." Neighborhood characteristics are calculated at the Census block group level.

Figure B6: Covariate Continuity Through the Cutoff I, 2022 Cohort



Notes: The sample includes first-time third-grade students in the 2022 cohort with valid M-STEP scores the following year. Each dot gives the share of students with a given characteristics for a given third-grade M-STEP ELA score. "NA, NH, or PI" stands for Native American, Native Hawaiian, or Pacific Islander. An IEP (individualized education program) is a document that outlines the programs and services a student with disabilities will receive to help them realize their educational goals, sometimes referred to as "special education."

Figure B7: Covariate Continuity Through the Cutoff II, 2022 Cohort



Notes: The sample includes first-time third-grade students in the 2022 cohort with valid M-STEP scores the following year. Each dot gives the share of students with a given characteristics for a given third-grade M-STEP ELA score. An IEP (individualized education program) is a document that outlines the programs and services a student with disabilities will receive to help them realize their educational goals, sometimes referred to as "special education." Neighborhood characteristics are calculated at the Census block group level.

C Details on the Principal Survey

As part of a broader project studying Michigan's Read by Grade Three (RBG3) Law, the Education Policy Innovation Collaborative (EPIC) at Michigan State University led a survey of educators throughout the state in the spring/summer of 2020, 2021, 2022, and 2023. The survey instruments were adapted from other surveys on literacy instruction and were fielded online using the Qualtrics platform. EPIC promoted the surveys to potential respondents using several online channels and through multiple Michigan education associations. The surveys contained questions on a wide range of topics and were targeted to several types of educators, including teachers, principals, district superintendents, and intermediate school district (ISD) early literacy coaches. For more information on the design and fielding of the surveys, see EPIC's year-one and year-two reports (Strunk et al., 2022, 2021).

In this paper, we are specifically interested in a survey question posed to principals about the literacy supports their schools provided. In 2022 and 2023, after the retention component of RBG3 was first implemented, principals were asked whether their school provided various literacy supports to three mutually exclusive groups: students who (i) scored below the cutoff and were retained, (ii) scored below the cutoff and were promoted, and (iii) scored just above the cutoff (in the 1253-1271 range) and were promoted. The third group was included because state guidance recommends that these students receive additional support, although it is not mandated. Because this question was asked in spring/summer of 2022 and 2023, principals' answers reflect supports provided to the 2021 and 2022 third-grade cohorts, i.e., the focal cohorts of this paper.

Principals answered this question by checking a box next to each type of support service their school provided (separately for each type of student). If their school did not provide any of the listed supports, they could check a box for "none of the above." In total, 395 principals responded to the question in 2022 and 310 responded in 2023. However, many responses were invalid; nearly half (46%) of the respondents left each checkbox blank for at least one of the three student types. Such a response is internally inconsistent given the "none of the above" option. To maintain comparability across student types, we drop respondents with an invalid response for any of the three types. This restriction leaves 227 complete responses in 2022 and 156 complete responses in 2023. These are the responses we tabulate and present in Table 4.

C.A Comparing the RD and Survey Samples

In Table C1, we investigate the degree to which schools in our survey analysis are representative of schools throughout the state. We begin by linking the survey responses to administrative data from the

Michigan Department of Education, using principals' ID codes as the linking variable. We successfully link 201 principals in the 2022 survey and 140 in the 2023 survey, which is 89% of the responses tabulated in Table 4. 48 schools were surveyed in both years, meaning we have 293 unique schools with survey responses. As Table C1 shows, the schools in our sample are generally very similar to other schools in the state. The biggest differences are that charter schools and economically disadvantaged students are slightly overrepresented in the survey sample. Surveyed schools also come from districts with lower average performance on the M-STEP exam. Lastly, Table C2 shows that the results of the survey do not change much when we restrict the survey sample to responses with linkable IDs.

Although schools in the survey sample appear fairly representative, the implementation and impact of RBG3 might still differ between schools in the survey sample and schools in the RD sample (which is the near universe of Michigan public schools). Given this possibility, we conduct our RD analysis again using only students from the 293 schools that have complete survey responses and can be linked to administrative data. We again use the estimation strategy outlined in Section V.A of the paper. As Table C3 shows, we again find small impacts on being retained and on ELA scores in the following year. The latter estimate is not statistically significant, which is unsurprising given the small sample size, but the point estimate (0.043 SD) is remarkably similar to the point estimate in the main analysis (0.045 SD, see Figure 1 or Appendix Table A3).

C.B Survey Results in District-Years With and Without Retention

Next, we examine how principal responses differ in district-years that did and did not retain students scoring near the retention cutoff. To test this, we re-tabulate survey responses using two subsets of the survey sample (among responses linkable to administrative data). The first subset contains 181 survey responses from 156 district-years with (i) at least one student scoring within 18 points of the cutoff on each side of the cutoff, and (ii) no students within 18 points of the cutoff who are retained. (Note that these districts may have retained students with scores more than 18 points away from the cutoff.) The second subset contains 108 survey responses from 92 district-years with (i) at least one student scoring within 18 points of the cutoff on each side of the cutoff, and (ii) at least one retained student within 18 points of the cutoff.

As Table C4 shows, principals in district-years with and without retention report similar levels of additional literacy support for students. In district-years without retention, students scoring above the cutoff receive less support from schools than students scoring below. The same is true for district-years with retention, but to a lesser extent. Table C5 formally explores comparisons across district-year type. Columns 1, 2, and 3 show some differences in the ways district-years with and without retention provide support,

but the differences are mostly small, very few are statistically significant, and there is no clear pattern. The column on the far right of the table shows a somewhat clearer pattern: scoring below the cutoff results in a bigger boost in literacy supports in district-years without retention than in those with retention. However, the differences are still not large or precise enough for us to reject the null hypothesis that the boost is the same.

C.C Internal Validity Check

One potential concern about the survey is that principals were asked about support services they provided to students who may not have existed in their school. For instance, many schools did not have any students who scored below the cutoff and were retained. In these cases, it is unclear how principals interpreted the survey question or chose to respond.

To ensure our results are not driven by principals answering hypothetical questions, we conduct a validity check in which we limit the sample to school-years that have at least one student in the administrative data in the "below-cutoff, promoted" and "above-cutoff, promoted" (1253-1271) categories. These two groups comprise the vast majority of students near the retention cutoff in the RD analysis. Among the 341 survey responses that are linkable to the administrative data, 209 come from school-years with at least one student in each of the aforementioned categories. Table C6 shows that our results are robust; as in Table 4, students in both categories receive support, but students scoring below the cutoff receive more support.

Table C1: Summary Statistics for Schools in Principal Survey that are Linkable to Administrative Data

	Linkable Principal Survey Schools	All Other Schools	Difference	P-value
School-Level Characteristics				
Charter school	0.150	0.095	0.055	0.002
City	0.198	0.222	-0.024	0.343
Rural	0.290	0.272	0.019	0.494
Suburb	0.355	0.376	-0.021	0.485
Town	0.157	0.131	0.026	0.210
White	0.646	0.645	0.001	0.969
Black	0.188	0.185	0.003	0.853
Hispanic	0.084	0.086	-0.002	0.781
Asian	0.026	0.035	-0.008	0.082
Economically disadvantaged	0.614	0.559	0.055	0.000
Enrollment	405.5	391.9	13.6	0.507
Pupil:teacher ratio	16.5	20.0	-3.5	0.368
Mean M-STEP ELA score (grades 3-8)	-0.084	-0.045	-0.039	0.227
Mean M-STEP math score (grades 3-8)	-0.101	-0.048	-0.053	0.139
District-Level Characteristics				
City	0.205	0.232	-0.027	0.297
Rural	0.280	0.233	0.047	0.071
Suburb	0.358	0.387	-0.029	0.327
Town	0.157	0.148	0.009	0.676
Economically disadvantaged	0.587	0.542	0.045	0.001
Enrollment	4,574.9	6,161.8	-1,586.9	0.010
Instructional spending per pupil	6,528.4	6,603.0	-74.6	0.445
Mean M-STEP ELA score (grades 3-8)	-0.113	-0.058	-0.055	0.038
Mean M-STEP math score (grades 3-8)	-0.120	-0.062	-0.059	0.053
Observations	293	3,269		

Notes: The sample of schools in the "Linkable Principal Survey School" column includes schools with principals who met two criteria: (i) they have logically consistent responses to our survey question on literacy supports for each of the three student groups, and (ii) their staff ID allowed us to link the survey response to school administrative data. In total, 341 survey responses met this criteria. 48 of those responses came from schools that were surveyed in each year of the survey, meaning the sample contains 293 unique schools. Mean M-STEP scores are missing at the school level for 13% of the linkable survey schools and 39% of the other schools.

Table C2: Share of Schools Offering Additional Supports to Students, in Schools Linkable to Administrative Data

	Student Group				
	Below Cutoff Retained (1)	Below Cutoff Promoted (2)	Above Cutoff Promoted (3)	P-value of (1)-(2)	P-value of (2)-(3)
Increased dosage (%)					
Assign to a high-quality teacher	68.6	71.0	66.0	0.182	0.003
Summer reading programs	71.3	69.2	62.5	0.548	0.004
High-dosage tutoring	41.4	36.4	29.3	0.033	0.014
Work with families on home reading	68.0	63.0	58.1	0.017	0.021
Extra instructional time in literacy	70.7	73.0	68.3	0.378	0.026
Before- or after-school literacy interventions	30.8	29.9	27.0	0.677	0.065
Supplemental virtual learning	17.0	17.3	16.4	0.622	0.412
Small group instruction during school	82.1	84.2	82.7	0.205	0.462
Other support (%)					
Literacy intervention curriculum	49.6	50.7	47.5	0.359	0.024
Focus on essential skills	73.9	75.4	72.1	0.191	0.125
Data-driven instruction	81.5	84.5	83.6	0.052	0.770
Total number of supports (#)	6.3	6.3	6.0	0.601	0.001

Notes: These results are from 341 principal surveys (in 296 district-years) that have complete responses for each student group in the table. The sample is limited to principals with IDs we can link to school administrative data. 48 schools were surveyed in both survey years, meaning 293 unique schools are in the sample. We weight each response from a school that was surveyed twice at 0.5 and all other responses at 1.0. "High-dosage tutoring" was only asked in the survey for the 2022 cohort. Items are listed in ascending order based on the last column.

Table C3: ITT Impact Estimates, Schools in Principal Survey that are Linkable to Administrative Data

	Control Mean	Impact Estimate	Optimal Bandwidth	Observations Within Bandwidth
Analysis Sample (N=26,270)				
Enrolled next year	0.998	-0.008***	(-11,25)	8,405
		(0.003)		
Took M-STEP next year	0.948	0.004	(-11,40)	13,422
		(0.007)		
Took M-STEP in fourth grade	0.950	0.000	(-11,37)	12,421
		(0.008)		
Conditional on Enrollment Next Year (N=26,104)				
Retained (next year)	0.025	0.046***	(-12,21)	7,169
		(0.014)		
Attendance rate (next year)	0.892	0.000	(-14,31)	10,294
		(0.004)		
Has an IEP (next year)	0.330	0.010	(-11,18)	6,207
		(0.012)		
Switched schools (next year)	0.174	-0.017	(-13,36)	12,081
		(0.018)		
Switched districts (next year)	0.097	-0.002	(-12,36)	12,053
		(0.014)		
Same-Age Test Score Impacts (N=25,343)				
Math score (next year)	-0.811	0.101***	(-8,21)	6,718
		(0.027)		
ELA score (next year)	-0.814	0.043	(-14,18)	6,037
		(0.033)		
Same-Grade Test Score Impacts (N=25,323)				
Math score (fourth grade)	-1.030	0.125***	(-8,21)	6,715
		(0.032)		
ELA score (fourth grade)	-1.093	0.047	(-13,16)	5,424
		(0.030)		

*** p<0.01, ** p<0.05, * p<0.1

Notes: This sample is limited to the 293 schools associated with principals who have complete responses to our survey question and whose ID codes allow us to link their survey responses to Michigan administrative data. We include each school in 2021 and 2022. We estimate models using the procedure developed by Calonico et al. (2018, 2020, 2014, 2019), allowing bandwidths to vary for each outcome and on each side of the cutoff. Standard errors are clustered at the running variable level. Control group means are calculated using students with ELA M-STEP scores within 2 points above the cutoff. An IEP (individualized education program) is a document that outlines the programs and services a student with disabilities will receive to help them realize their educational goals, often referred to as "special education."

Table C4: Share of Schools Offering Additional Supports to Students, in District-Years With and Without Retention Near the RD Cutoff

	Student Group				
	Below Cutoff Retained (1)	Below Cutoff Promoted (2)	Above Cutoff Promoted (3)	P-value of (1)-(2)	P-value of (2)-(3)
	Panel A. District-Years Without Retention				
Increased dosage (%)					
Summer reading programs	76.8	70.7	60.8	0.048	0.000
Assign to a high-quality teacher	66.9	72.9	65.2	0.016	0.001
Extra instructional time in literacy	74.0	78.5	71.8	0.088	0.002
Work with families on home reading	69.1	66.3	59.1	0.276	0.003
High-dosage tutoring	38.7	33.3	22.7	0.103	0.004
Small group instruction during school	82.9	88.4	83.4	0.025	0.029
Before- or after-school literacy interventions	31.5	30.4	27.1	0.639	0.109
Supplemental virtual learning	16.6	17.7	14.9	0.481	0.132
Other support (%)					
Focus on essential skills	73.5	77.9	72.4	0.074	0.018
Literacy intervention curriculum	51.9	54.1	49.7	0.286	0.021
Data-driven instruction	82.9	87.8	83.4	0.020	0.032
Total number of supports (#)	6.4	6.6	6.0	0.288	0.000
	Panel B. District-Years With Retention				
Increased dosage (%)					
Summer reading programs	70.4	73.1	65.7	0.408	0.032
Assign to a high-quality teacher	75.9	73.1	68.5	0.441	0.058
Extra instructional time in literacy	73.1	72.2	68.5	0.765	0.207
Work with families on home reading	72.2	64.8	59.3	0.032	0.083
High-dosage tutoring	40.9	36.4	34.1	0.323	0.660
Small group instruction during school	88.0	83.3	81.5	0.096	0.530
Before- or after-school literacy interventions	31.5	29.6	27.8	0.530	0.482
Supplemental virtual learning	17.6	17.6	18.5	1.000	0.320
Other support (%)					
Focus on essential skills	78.7	76.9	70.4	0.530	0.090
Literacy intervention curriculum	50.0	50.0	45.4	1.000	0.096
Data-driven instruction	85.2	86.1	84.3	0.741	0.530
Total number of supports (#)	6.6	6.4	6.0	0.264	0.018

Notes: These results are from 289 principal surveys that have complete responses for each student group in the table. The sample for Panel A includes 181 responses from principals in 156 district-years with (i) at least one student scoring within 18 points of the cutoff on each side of the cutoff, and (ii) no students within 18 points of the cutoff who are retained. The sample for Panel B includes 108 responses from principals in 92 district-years with (i) at least one student scoring within 18 points of the cutoff on each side of the cutoff, and (ii) at least one retained student within 18 points of the cutoff. "High-dosage tutoring" was only asked in the survey for the 2022 cohort. Items are listed in ascending order based on the last column of Panel A.

Table C5: Differences in Additional Support Offered by Schools in District-Years Without vs. With Retention Near the RD Cutoff

	Below Cutoff Retained (1)		Below Cutoff Promoted (2)		Above Cutoff Promoted (3)		Difference in "Discontinuities" (2)-(3)	
	Difference	P-value	Difference	P-value	Difference	P-value	Difference	P-value
Increased dosage (%)								
Summer reading programs	6.4	0.237	-2.4	0.657	-5.0	0.397	2.5	0.561
Assign to a high-quality teacher	-9.1	0.095	-0.2	0.968	-3.3	0.562	3.1	0.351
Extra instructional time in literacy	0.9	0.870	6.2	0.241	3.3	0.556	2.9	0.421
Work with families on home reading	-3.2	0.568	1.5	0.799	-0.1	0.981	1.6	0.681
High-dosage tutoring	-2.2	0.812	-3.0	0.741	-11.4	0.192	8.4	0.183
Small group instruction during school	-5.1	0.228	5.1	0.242	1.9	0.677	3.1	0.400
Before- or after-school literacy interventions	0.0	0.999	0.8	0.892	-0.7	0.897	1.5	0.661
Supplemental virtual learning	-1.0	0.825	0.1	0.985	-3.6	0.434	3.7	0.073
Other support (%)								
Focus on essential skills	-5.2	0.311	1.0	0.838	2.0	0.717	-1.0	0.829
Literacy intervention curriculum	1.9	0.752	4.1	0.497	4.4	0.475	-0.2	0.950
Data-driven instruction	-2.3	0.603	1.7	0.675	-0.8	0.853	2.6	0.474
Total number of supports (#)	-0.2	0.601	0.2	0.551	-0.1	0.827	0.2	0.248

Notes: These results are from 289 principal surveys that have complete responses for each student group in the table. Each principal is in a district-year with at least one student scoring within 18 points of the retention cutoff of each side of the cutoff. The sample is then split into (i) 181 responses are from principals in 156 districts with no students within 18 points of the cutoff who are retained, and (ii) 108 responses from principals in 92 district-years with at least one retained student within 18 points of the cutoff. The first three "difference" columns give share of students who are offered additional literacy services in districts without retention minus the share in districts with retention. The final "difference" column gives the difference between the second and third difference columns. This so-called difference in discontinuities indicates whether the difference in services offered to promoted students below vs. above the cutoff is greater in districts without retention or in districts with retention. "High-dosage tutoring" was only asked in the survey for the 2022 cohort.

Table C6: Share of Schools Offering Additional Supports to Students, in School-Years With Students in Each Category in the Administrative Data

	Student Group		
	Below Cutoff Promoted (2)	Above Cutoff Promoted (3)	P-value of (2)-(3)
Increased dosage (%)			
Summer reading programs	75.6	66.0	0.000
Work with families on home reading	67.9	60.3	0.001
Assign to a high-quality teacher	76.6	70.8	0.001
Extra instructional time in literacy	78.9	73.7	0.004
Small group instruction during school	90.4	86.1	0.020
High-dosage tutoring	35.4	29.2	0.057
Before- or after-school literacy interventions	32.5	29.2	0.071
Supplemental virtual learning	16.7	16.3	0.706
Other support (%)			
Focus on essential skills	83.7	77.5	0.003
Data-driven instruction	93.8	89.0	0.004
Literacy intervention curriculum	58.9	54.1	0.004
Total number of supports (#)	6.9	6.4	0.000

Notes: These results are from 209 principal surveys that have complete responses for each student group in the table. The sample is limited to principals with IDs we can link to school administrative data. Among these principals, we further limit the sample to those in school-years with at least one student in the administrative data in the "below-cutoff, promoted" category and in the "above-cutoff, promoted" (1253-1271) category. "High-dosage tutoring" was only asked in the survey for the 2022 cohort. Items are listed in ascending order based on the last column.