CESifo WORKING PAPERS

10986
2024

March 2024

Does One Plus One Always Equal Two? Examining Complementarities in Educational Interventions

Umut Özek

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10986*

Does One Plus One Always Equal Two? Examining Complementarities in Educational Interventions

Abstract

Public policies targeting individuals based on need often impose disproportionate burden on communities that lack the resources to implement these policies effectively. In an elementary school setting, I examine whether community-level interventions focusing on similar needs and providing resources to build capacity in these communities could improve outcomes by improving the effectiveness of individual-level interventions. I find that the extended school day policy that targets lowest-performing schools in reading in Florida significantly improved the effectiveness of the third-grade retention policy in these schools. These complementarities were large enough to close the gap in retention effects between targeted and higher-performing schools.

JEL-Codes: I200, I280.

Keywords: educational interventions, complementarities, disadvanted communities.

Umut Özek
RAND Corporation
Arlington / VA / USA
uozek@rand.org

INTRODUCTION

Can educational interventions influence student outcomes by improving the effectiveness of other interventions? This is an important policy question because students are increasingly exposed to multiple interventions at the individual-, classroom-, or school-level either concurrently or consecutively with the rise of test-based accountability in the United States over the past three decades (Ravitch et al., 2022). And while there is extensive literature about the individual effects of these interventions on student outcomes, very little is known about how they interact with one another. Understanding these interactions is critical from a public policy perspective as they can create “free” gains or unintended “losses” and have direct implications about the cost effectiveness of these policies.

I examine this question using two interventions in Florida that target low-performing students and schools in reading: (1) third-grade retention policy (enacted in 2002) that requires students to repeat third grade and receive instructional support unless they score above the lowest achievement level on the third-grade reading test and (2) the extended school day (ESD) policy (enacted in 2012) that requires the lowest-performing schools in the state (selected according to an index of school-level reading accountability measures) to extend the school day by an hour in order to provide additional literacy instruction. Several recent studies examine the individual effects of these policies and show that both policies improve student test scores in the short term.[1] In this study, I ask whether the ESD policy improved the effectiveness of the grade retention policy (i.e., the effects of the retention policy on the educational outcomes of targeted

[1] For the effects of Florida’s third-grade retention policy on student outcomes, see Figlio and Ozek (2020); Greene & Winters (2012); Greene & Winters (2007); Schwerdt et al. (2017). For the effects of Florida’s extended school day policy, see Figlio et al. (2018).

students) in targeted schools using student-level administrative data from a large, urban school district (LUSD) in the state.

These two interventions present an interesting case from a public policy perspective. A common issue with targeted interventions in education is that they place disproportionate burden on schools serving traditionally marginalized students that often lack the resources to implement these interventions effectively.[2] This was also the case for the third-grade retention policy examined in this study: 40 percent of the third graders flagged for retention in the LUSD (during the time frame examined in this study) were enrolled in schools that fall into the top poverty quartile (as proxied by the share of subsidized meal eligible students) in the district, and nearly one-third of third graders in these schools were flagged for retention. This study examines whether supplementing these individual-level interventions with school-level interventions that focus on similar needs and provide resources to build capacity in these communities could generate complementarities. More specifically, I ask whether the resources provided by the ESD policy to improve reading instruction in disadvantaged schools (e.g., additional instruction time in literacy, reading coaches) had a spillover effect on the effectiveness of the third-grade retention policy in those schools.

To identify the complementarities, I use a difference-in-differences (DiD) in regression discontinuity (RD) design, making use of a plausibly exogenous expansion of the ESD policy from 100 schools to 300 schools in 2014/15 school year. In essence, in the preferred specification I compare (1) the difference in the effect of being flagged for retention (estimated using an RD design) in schools that were designated as ESD schools due to the expansion (i.e.,

[2] For example, due to socioeconomic and racial/ethnic disparities in test scores driven by structural inequities in American society coupled with differential enforcement of educational interventions (e.g., LiCalsi et al. 2019), students in disadvantaged school settings are significantly more likely to be identified for targeted interventions (e.g., Figlio and Ozek, 2024).

schools that were ranked between 100 and 300) in the year after (2014/15) versus before (2013/14) the expansion with (2) the same difference in elementary schools whose reading accountability index ranking fell right above the 300 cutoff (and hence were not designated as ESD). I check the robustness of this two-by-two DiD in RD design with an event study in RD approach, comparing the retention effects in treatment and comparison schools in the years before and after the expansion of the ESD program.

I find that the ESD policy significantly improved the effectiveness of the retention policy. In particular, the results reveal no significant effect of being flagged for retention on student test scores in reading in ESD schools in the year before the designation (perhaps driven by lack of educational resources or how effectively existing resources were used), yet a significant positive effect in the year after. In contrast, I find no significant change in the effect of being flagged for retention in schools above the ESD cutoff. DiD in RD estimates suggest that the ESD policy increased the effect of being flagged for retention by 23 percent of the standard deviation (0.23σ) in following year reading scores in targeted schools. Using an unexpected pause in the retention policy, I also provide a falsification exercise showing that these effects are not driven by the differential effect of ESD designation on the lowest performing third graders in reading.

There are several mechanisms that might explain these complementarities. The first channel is the additional resources ESD schools receive to improve reading instruction. For example, as discussed below, nearly all ESD schools in Florida reported hiring additional staff for the additional hour of instruction including reading coaches and teachers (Folsom et al., 2016), which could have improved the fidelity of implementation for the retention policy. These additional resources could have also enabled schools to retain more retention-eligible students who would benefit from an additional year of instruction (and supports) without sacrificing the quality/quantity of support for the retained students. Indeed, I find an increase in retention policy enforcement in ESD schools; in other words, ESD designation increased the effect of being flagged for retention on the likelihood of being retained. Second, the longer school day could have allowed these schools to provide the instructional supports for retained students more effectively. Along similar lines, the ESD designation may have led to increased focus on reading instruction (and adoption of more effective instructional strategies in reading) in schools, which might also improve the effectiveness of retention. Indeed, using a DiD in fuzzy RD design, I show that ESD designation improved the effect of being retained on following year reading scores by 0.56σ. This interaction effect is roughly equivalent to the difference in the effect of retention in treatment and comparison schools before the ESD designation. So, in essence, the ESD program closed the gap in retention effects between the most disadvantaged schools and others in the district.

This study contributes to two strands of literature in economics of education. First, it complements the existing literature about the effects of educational interventions on student outcomes in general, and the literature on the effects of K-12 remediation programs in particular. While there is extensive literature on the effectiveness of individual programs/policies, to the best of my knowledge, this study presents the first evidence on how universal education policies implemented at scale interact with one another.[3] Second, it adds to the literature on the importance of school funding (or resources) for student outcomes. There is growing evidence suggesting that school spending has profound effects on student outcomes (for a recent meta-

[3] There is a recent study that examines complementarities between educational investments that are implemented concurrently in Tanzania using a randomized controlled trial (Mbiti et al., 2019). They show that teacher incentives (based on student test scores) and unconditional grants to schools are only effective when implemented simultaneously. A related strand of research examines dynamic complementarities between educational investments implemented consecutively (e.g., Malamud et al., 2016; Johnson & Jackson, 2017).

analysis, see Jackson & Mackevicius, 2024). This study sheds light on an important mechanism through which resources targeting disadvantaged school settings may improve student outcomes.

POLICY BACKGROUND

Florida's Third-Grade Retention Policy

Enacted in 2002 as part of the broader "Just Read, Florida!" initiative, Florida's third-grade retention policy requires students to meet the Level 2 benchmark or higher (the second lowest of five achievement levels) on the statewide reading test in order to be promoted to fourth grade. While test-based grade retention has been a popular proposal long before Florida's policy to improve the outcomes of low-performing students, this initiative has been highly influential and provided a blueprint for many other early grade retention policies nationwide (Cummings & Turner, 2020).

The legislation requires that schools provide substantial instructional support for retained students in the following school year. While students flagged for retention are eligible to participate in summer school at the end of the year similar to other retention policies (e.g., Chicago Public Schools, New York City, Louisiana)[4], under Florida's retention policy, schools are required to develop academic improvement plans for retained students that specifically address their needs, to assign these students to high-performing teachers (based on student performance and performance appraisals), and to provide a minimum of 90 minutes of reading instruction each day.

[4] Under Florida's policy, schools and districts are required to create a reading camp schedule that facilitates intensive reading intervention lasting between six to eight weeks, four days per week, and six hours per day. For these sessions, schools and districts are encouraged to (1) choose qualified teachers and reading coaches with reading certification or endorsement and reading coaches and (2) provide reading instruction utilizing a research-based sequence of reading instruction and small group differentiated instruction in order to meet individual student needs.

There are several “good cause exemptions” that allow students to be promoted to fourth grade despite failing to score at the Level 2 benchmark or above. For example, students are eligible for an exemption (1) if they have certain disabilities and have been already retained once until third grade; (2) if they have received intensive reading remediation for two years and have already been retained twice between kindergarten and third grade; (3) if they have been in the English learner program for less than two years; (4) by demonstrating that they are reading at a level equal to or above a Level 2 on the statewide reading test by performing at an acceptable level on an alternative standardized reading assessment approved by the State Board of Education; or (5) by demonstrating proficiency through a teacher-developed portfolio. That said, the policy has affected a significant share of third graders in the LUSD: Between 2005/06 and 2017/18 school years, 19 percent of third graders were flagged for retention and 7 percent had to repeat third grade.

There is extensive literature examining the effects of grade retention in Florida on student outcomes. The overarching conclusion is that retention (coupled with instructional support) significantly improves student outcomes in the short term, but the effects on long-term outcomes are less clear. For instance, Schwerdt et al. (2017) find that retention increases student test scores in reading by 23 to 58 percent of the standard deviation (0.23σ) in one year (0.30σ in math), and by 0.49σ after two years (0.10σ in math), yet these short-term benefits fade out rapidly after three years. However, retained students under Florida’s retention policy significantly outperform their promoted peers when they reach the same grade level (Figlio and Ozek, 2020; Greene & Winters, 2012; Greene & Winters, 2007; Schwerdt et al., 2017). For example, third-grade retention increases eighth-grade test scores by toughly 0.20σ in reading and 0.13σ in math (Schwerdt et al., 2017). Retained students are also less likely to be retained in a later grade and no more or less likely to graduate from high school (Schwerdt et al., 2017). While test-based grade retention has been a popular proposal long before Florida's policy to improve the outcomes of low-performing students, this initiative has been highly influential and provided a blueprint for many other early grade retention policies nationwide (Cummings & Turner, 2020), partly driven by these positive findings.

Florida's Extended School Day Policy

In 2012, the state of Florida passed legislation requiring the lowest-performing elementary schools – selected based on a school-level accountability index in reading – to extend the school day by an hour to provide literacy instruction.[5] The legislation also imposes several restrictions on how the additional hour is to be utilized: ESD schools are required to provide literacy instruction based on research; instruction must be adapted for student ability; instruction should include phonemic awareness, phonics, fluency, vocabulary, and comprehension; students must have guided practice; and students must read material from social studies, science, and math classes. In 2012/13 and 2013/14 school years, 100 lowest-ranked schools based on this index were identified as ESD schools, which was later expanded to the lowest 300 schools in 2014/15. In the first year of the program, 7 elementary schools in the LUSD (out of 123 elementary schools in the district that year) were identified as ESD schools whereas this number jumped to 23 schools in the first year after the expansion and has remained above 20 schools annually since then.

The state allocates roughly $300,000 per school (approximately $800 per student) annually for the implementation of the program, which can be supplemented by funding from the

[5] In particular, all elementary schools in the state are ranked according to the sum of points for "reading performance" and "annual learning gains in reading". Reading performance is determined by the percentage of students in the school who scored a "satisfactory" (also known as Level-3) on the statewide reading test, and annual learning gains are determined by the percentage of students that make adequate gains in reading achievement levels.

school district if necessary (Figlio et al., 2018).[6] Corbett (2015), Folsom et al. (2016), and Folsom et al. (2017) show that ESD schools mainly use this funding (1) to train teachers who provided the additional instruction; (2) to hire additional staff (e.g., reading coaches, teachers, paraprofessionals, or volunteers) for the extended school day (nearly all ESD schools report hiring additional staff in the first year of ESD designation); and (3) to cover other expenses related to the extended school day (e.g., existing staff salaries, facilities). Figlio et al. (2018) examine the causal effect of ESD designation on student reading scores using an RD design in the first year of the program. They find that students enrolled in schools whose reading accountability scores fell immediately below the ESD cutoff (and hence were required to provide additional instruction) score roughly 0.05σ better in reading compared to students enrolled in schools just above the cutoff in the first year of the designation.

DATA AND EMPIRICAL FRAMEWORK

In this study, I make use of detailed longitudinal, student-level administrative data from a LUSD in Florida. These data cover all students enrolled in grades K–12 between 2005/06 and 2017/18 school years and include reading and mathematics scores of all tested students.[7] In addition to these test scores, the data set includes demographic information on students, such as race, gender, free-or-reduced price lunch (FRPL) eligibility, English learner status, exceptional/special education status, country of birth, language spoken at home, student age, and schools attended. Given the main objective of the two policies under investigation, I use test scores in reading as the main outcome of interest in the analysis. Specifically, I use the following

[6] To put this number into perspective, state funding for the ESD program roughly corresponds to 10 percent of the annual per-pupil spending in ESD schools.

[7] These include Florida Comprehensive Assessment Test (FCAT) scores in reading and mathematics for all students in Grades 3–10 until the 2011–12 school year; FCAT 2.0 scores in reading for Grades 3–10 and in mathematics for Grades 3–8 between 2012–13 and 2014–15; and Florida Standards Assessment (FSA) scores in reading for Grades 3–10 and in mathematics for Grades 3–8 since 2014–15.

year developmental scale scores (that are comparable across grades) of first-time third graders that are standardized at cohort level as the primary outcome of interest.[8] In what follows, I describe how I estimate the causal effects of these two policies on student outcomes individually (and thus replicate the findings from prior studies in this context), and how the ESD designation influenced the effectiveness of the retention policy.

Effects of Third-Grade Retention on Student Outcomes

To estimate the effects of grade retention on student outcomes, I rely on an RD design. Formally, let $S_{i,t-1}$ denote the difference between the third-grade reading score of student *i* who entered third grade for the first time in year *t-1* and the retention cutoff—with negative values indicating scores below cutoff—and $B_{i,t-1}$ denote an indicator for students below the cutoff. In this setting, the effect of failing the third-grade test (and being flagged for retention) on student outcomes is given by:

$$\mu = lim_{S_{i,t-1}\uparrow 0}E\left[Y_{it}\middle|S_{i,t-1}\right] - lim_{S_{i,t-1}\downarrow 0}E\left[Y_{it}\middle|S_{i,t-1}\right] \qquad (1)$$

where Y_{it} is the reading score of student *i* in year *t*. I estimate μ using the following equation and OLS:

$$Y_{it} = \gamma + \mu B_{i,t-1} + k\left(S_{i,t-1}\right) + k\left(S_{i,t-1}\right) * B_{i,t-1} + \upsilon_{it} \qquad (2)$$

where $k\left(S_{i,t-1}\right)$ is a function of the relative test score. In particular, I estimate this model using the linear specification and a bandwidth of 25 points based on the range of bandwidths

[8] That is, the outcome of interest represents where each student stands compared to other students in the same third grade cohort in the following year (note that some of these students will be repeating the third grade whereas others will be in fourth grade in the following year). For example, consider the set of first-time third graders in 2012/13 school year. The standardized following year reading score of a student represents where that student's developmental reading score in 2013/14 fell on the distribution of developmental test scores of students in the 2012/13 third-grade cohort in 2013/14.

suggested for various outcomes by the bandwidth selection procedure in Calonico et al. (2017).[9] In the analysis, I focus on the two cohorts of students who entered third-grade for the first time in 2012/13 and 2013/14: These are the cohorts that I use to examine the complementarities between the retention and ESD policy.

In this setting, μ provides the causal effect of being flagged for retention on student outcomes provided that all student attributes (other than the likelihood of being flagged for retention) are smooth around the cutoff. While this condition cannot be definitively proven, I conduct several tests. First, I examine whether the baseline characteristics of students are continuous by replacing the outcome of interest (Y_{it}) in Equation (2) with student baseline characteristics (e.g., third-grade math score, race/ethnicity, subsidized meal eligibility) and examine the discontinuity in these attributes at the retention cutoff. Table 1 presents the results of this falsification exercise examining the pseudo effects of being flagged for retention on the third-grade outcomes and attributes of students. Only two of the seven estimated coefficients are statistically distinguishable from zero and only one of the estimates imply a discontinuity larger than 10 percent of the dependent variable mean at the cutoff.

Second, I check for the possibility of selection variable manipulation as noted in McCrary (2008) using the density test developed by Frandsen (2017) for discrete running variables, even though this is very unlikely in this context since standardized test scores are assessed without any teacher, student, or principal involvement. Panel (A) in Figure 1 portrays the distribution of third-grade reading scores of students in the sample and provides evidence that the distribution is continuous and that manipulation of the running variable is not an issue. I reject the hypothesis

[9] In this analysis, I present the Eicker-Huber-White heteroskedasticity-robust standard errors suggested by Kolesár and Rothe (2018) for RD designs using discrete running variables, noting that the results are robust to clustering at the running variable level as suggested by Lee and Card (2009).

on discontinuity in the density of the distribution at the cutoff, with a p-value of 0.547 (Frandsen, 2017).

Because not all students who score below the retention cutoff are eventually retained under Florida's policy due to exemptions, I also use a fuzzy RD design to estimate the effect of being retained on student outcomes in a two-stage least squares (2SLS) framework where I instrument for grade retention ($R_{i,t-1}$) using $B_{i,t-1}$ as follows:

$$R_{i,t-1} = \omega_0 + \omega_1 B_{i,t-1} + k(S_{i,t-1}) + k(S_{i,t-1}) * B_{i,t-1} + \rho_{i,t-1} \tag{3-1}$$

and the fitted value of $R_{i,t-1}$ is used in a second stage:

$$Y_{it} = \tau_0 + \tau_1 \widehat{R_{i,t-1}} + k(S_{i,t-1}) + k(S_{i,t-1}) * B_{i,t-1} + \theta_{it}. \tag{3-2}$$

In addition to the assumptions outlined for the sharp RD design described in equation (2), this fuzzy RD design requires that being flagged for retention only impacts student outcomes through its effect on retention likelihood (i.e., exclusion restriction). Schwerdt et al. (2017) and Figlio and Ozek (2020) find evidence supporting the validity of this assumption by examining the effects of scoring below the retention threshold before the retention policy took effect. While I am unable to conduct a similar exercise due to data limitations, it is important to note that μ provides a lower-bound for the causal effect of retention on student outcomes.

Effects of ESD Designation on Student Outcomes

Estimating the causal effect of ESD designation is less straightforward in this context for several reasons. For example, unlike Figlio et al. (2018), I am unable to employ an RD design to estimate the causal effect of ESD designation on student outcomes because ESD rankings (that are used to determine the designation) use all elementary schools in the state while I am only able to use data from a single district in Florida. As such, there are very few elementary schools observed in any given year around the ESD cutoff in the district. Further, an event study approach is also problematic given the implementation of the ESD policy for two reasons. First, the set of schools identified as ESD changes every year.[10] More importantly, as noted in Figlio et al. (2018), schools that were identified as ESD could voluntarily continue with an extended school day even though their reading accountability score fell above the cut-point (and hence they were not required to implement the policy).[11] And Florida Department of Education (FLDOE) only provides information about schools below the cutoff (that are required by law to extend school day), and no information, to the best of my knowledge, is available about voluntary participants (schools that keep the longer school day even though they are no longer designated as ESD). This makes it hard to identify treatment and comparison schools in the later years of the program, especially for schools that were identified as ESD in prior years.

Given these challenges, I rely on a two-by-two difference-in-differences design to estimate the effects of ESD designation on student outcomes, focusing on a plausible exogenous policy change that expanded the program from 100 to 300 schools in 2014/15. Formally, I estimate the following equation using the same two cohorts (students who entered third grade for the first time in 2012/13 and 2013/14) and OLS to obtain the causal effect of ESD designation on student reading achievement:

$$Y_{it} = \alpha + \gamma_1 Post_t * Treat_{s(i)} + \gamma_2 Treat_{s(i)} + \gamma_3 Post_t + \gamma_4 X_{i,t-1} + \gamma_5 S_{s(i),t} + \varepsilon_{it} \quad (4)$$

where $Post_t$ is an indicator for the cohort who entered third grade for the first time in 2013/14 (and hence were exposed to the expansion in year *t* with the 2012/13 cohort serving as the comparison group); $X_{i,t-1}$ is the third-grade attributes of these students (e.g., test scores,

[10] There were two exceptions to this rule when the state used the same set of schools due to (1) the adoption of a new standardized test in 2015/16 and (2) the lack of test data in the first year of the Covid-19 pandemic.
[11] For example, as reported in West & Vickers (2014), in 2013/14 school year, 30 schools (out of the 83 surveyed schools that were identified as ESD schools in 2012/13, the first year of the program) chose to continue with an extended school day even though they were not required.

race/ethnicity, English learner and special education status); $S_{s(i),t}$ is a vector of school-by-year level covariates; and $Treat_{s(i)}$ is an indicator for students who were enrolled in treatment schools.

The main identification assumption in this framework is that there are no time-varying school-level factors (e.g., teacher and/or principal turnover) that may simultaneously lead to the ESD designation and the observed differences in student test scores. I follow several approaches to address this issue. First, in the analysis I focus on ESD schools that would not have been designated as ESD in the absence of the expansion (LUSD schools that were ranked between 100 and 300 in 2014/15 school year): These schools are more likely to be designated as ESD because of the (plausibly exogenous) expansion rather than other school-level shocks. Figure 2 provides a visual portrayal of the treatment and comparison schools used in the analysis and presents the distribution of LUSD elementary school rankings based on the statewide reading accountability index in in 2014/15 school year with a bin width of 50. Treatment schools in the main analysis (used both to estimate the main effects of ESD designation and the complementarities) are LUSD elementary schools whose ESD ranking fell between the gray and red lines (i.e., schools whose ESD ranking in 2014/15 fell between 100 and 300). Comparison schools in the main specifications are all non-ESD schools in 2014/15 (schools whose rankings fell above the red line in Figure 2). Based on these definitions, there are 14 schools in the treatment group and 107 schools in the comparison group.

Second, in some specifications I control for school-by-year level covariates including accountability measures that were used to identify ESD schools.[12] Because ESD designation

[12] These include student outcomes and covariates listed in Table 2 for all students in the school averaged at the school-by-year level, school-by-year accountability measures used under Florida's school accountability system including % of students by achievement level, % of student making adequate gains, % of low-performing students

depends on the reading performance of the school relative to other elementary schools in the state, controlling for these accountability measures implies that the variation in ESD designation is driven by (1) the ESD expansion and/or (2) changes in the reading performance of other schools, both of which are plausibly exogenous to time-varying factors (other than ESD designation) in treatment schools.

Finally, in some specifications, I restrict the comparison schools to those right above the ESD cutoff, which makes it less likely that differential trends in school-level factors driving the results is an issue because schools around the ESD cutoff that are more likely to be similar along time-varying attributes. To be consistent, I use the two cohorts who entered third grade for the first time in 2012/13 and 2013/14 and examine their following year reading scores in both the retention and the ESD analysis. This implies that the ESD analysis, in a nutshell, corresponds to a comparison of following year reading scores of students in the latter cohort (students who were exposed to the ESD expansion in the following year) with students in the former cohort in treatment schools with the same difference in comparison schools.[13]

Effects of ESD Designation on Retention Effects

The empirical strategy to estimate the complementarities between the two policies is basically a combination of these two empirical approaches. In particular, I rely on a DiD in RD design where the first difference in DiD corresponds to the change in retention effects between 2013/14 and 2014/15 school years in treatment schools and the second difference in DiD is the change in retention effects in comparison schools. This is similar to following a two-step

making adequate gains in reading and math, and the reading accountability index that is used to rank students to identify ESD schools.

[13] Focusing on these school years also alleviates the concern about accurately identifying ESD schools in later years. This follows because "voluntary participants" is not a concern: there were no non-ESD schools in LUSD in 2014/15 that were previously designated as ESD in the sample. In particular, no LUSD school was designated as ESD in 2013/14; and all seven "prior-ESD" schools in 2014/15 were once again designated as ESD with the expansion.

approach in which the first step estimates the retention effects (i.e., the effect of being flagged for retention on following year reading scores) in an RD framework for each school and the two cohorts, and then using these estimated effects as outcomes in a two-by-two DiD design (accounting for the error in estimated retention effects), comparing the retention effects in treatment versus comparison schools for the younger versus the older third-grade cohort. Formally, in this DiD in RD approach, I estimate the following equation using OLS:

$$Y_{it} = [\text{Eq } 2] * [\text{Eq } 4]$$

$$= [\gamma + \mu B_{i,t-1} + k(S_{i,t-1}) + k(S_{i,t-1}) * B_{i,t-1} + \upsilon_{it}] * \quad (5)$$

$$[\alpha + \gamma_1 Post_t * Treat_{s(i)} + \gamma_2 Treat_{s(i)} + \gamma_3 Post_t + \gamma_4 X_{i,t-1} + \gamma_5 S_{s(i),t} + \varepsilon_{it}]$$

Rearranging, one can obtain:

$$Y_{it} = \beta_0 + \{\beta_1 B_{i,t-1} + k(S_{i,t-1}) + k(S_{i,t-1}) * B_{i,t-1}\} * Treat_{s(i)} * Post_t +$$

$$\{\beta_2 B_{i,t-1} + k(S_{i,t-1}) + k(S_{i,t-1}) * B_{i,t-1}\} * Post_t +$$

$$\{\beta_3 B_{i,t-1} + k(S_{i,t-1}) + k(S_{i,t-1}) * B_{i,t-1}\} * Treat_{s(i)} +$$

$$\{\beta_4 B_{i,t-1} + k(S_{i,t-1}) + k(S_{i,t-1}) * B_{i,t-1}\} + \beta_5 Treat_{s(i)} + \beta_6 Post_t + \vartheta_{it} \quad (6)$$

In this setting, β_4 provides the retention effect (the effect of being flagged for retention) for the omitted group (the pre-ESD-expansion year in comparison schools); β_3 provides the treatment–comparison difference in retention effects during the pre-ESD-expansion year; β_2 provides the difference in effects over time in the comparison schools: the post-ESD-expansion year effect minus the pre-ESD-expansion year effect; and β_1 (the parameter of interest) provides the differential retention effect for students in the post-ESD-expansion year in treatment schools, which corresponds to the effect of ESD designation on the effect of scoring below the retention cutoff under certain assumptions discussed below. I estimate equation (6) using a linear $k(S_{it})$ and a bandwidth of 25 points in the main specification and check the robustness of findings to different bandwidths (including the optimal bandwidth selection procedure suggested by Calonico et al. [2017]). I cluster the standard errors at the school level given that the treatment is at the school level in the ESD policy. Once again, I use the two cohorts who entered third grade for the first time in 2012/13 and 2013/14 and examine their following year reading scores. In some specifications, I also control for the third-grade characteristics (and outcomes) of students to check the robustness of the findings and improve the precision of the estimates.

There are two important identification assumptions behind this framework: (1) the standard DiD assumptions are met (i.e., there are no differential time-varying school-level factors that simultaneously lead to the ESD designation and the observed changes in retention effects) and (2) the standard sharp RD assumptions outlined above are met (i.e., students around the retention cutoff are identical other than the treatment status). To address (1), I follow similar approaches in this interaction exercise as the main effects of ESD designation: (1) I focus on ESD schools whose ESD rankings fell above the pre-expansion cutoff (100) as treatment schools; (2) I control for school-level accountability measures that are used to identify ESD schools in some specifications; and (3) I check the robustness of the findings to using schools right above the ESD cutoff as the comparison group. Further, I explicitly check for differential trends in an event study in RD framework with the aforementioned caveats of the event study approach in this context in mind. The findings provided in Table 1 and Panel (A) in Figure 1 provides evidence suggesting that assumption (2) is satisfied.

Table 2 compares the third-grade outcomes and characteristics of first-time third graders in 2012/13 and 2013/14 who were enrolled in treatment and comparison schools in the following year. As expected, students in treatment schools have significantly lower third-grade test scores in reading and math, are more likely to be involved in disciplinary incidents, have higher absences, are more likely to be eligible for subsidized meals, are more likely to be classified as English learners, and are more likely to be non-White (descriptive statistics are provided in the first two columns). These discrepancies narrow considerably, yet still exist, when I focus on students around the retention cutoff in these schools (third and fourth columns), and decline even further (and vanish in some cases) when I compare students around the cutoff in treatment schools and comparison schools right above the ESD cutoff (ranked lower than 400 points above the cutoff) in the last two columns.

RESULTS

Effects of Grade Retention on Student Reading Achievement

Panel (B) in Figure 1 presents a graphical depiction of the following year reading scores around the retention cutoff whereas the top panel of Table 3 presents the effects of being flagged for retention (μ) estimated using a bandwidth of 25 points and with and without the baseline student characteristics listed in Table 2. The findings suggest that students whose third-grade reading scores fall right below the retention cutoff score roughly 0.16σ to 0.17σ higher than their peers right above the cutoff. Another parameter of interest in this context is the effect of retention (rather than the effect of being flagged for retention). Panel (C) in Figure 1 plots the likelihood of being retained around the retention cutoff (i.e., the first stage in a fuzzy RD design) and shows that students right below the retention cutoff were roughly 35 percentage points more likely to be repeat third-grade compared to their peers whose third-grade reading scores fell right above the cutoff. Using this first-stage, the fuzzy RD estimates (τ_1) provided in the bottom panel of Table 3 suggest that retention improves student reading scores by 0.41σ to 0.44σ in one year. This is in line with the estimates presented in the prior literature: For example, Schwerdt et al.

(2017) find first-year retention effects of 0.23σ to 0.58σ depending on whether they use rescaled developmental scale scores or not.

Effects of ESD Designation on Student Reading Achievement

Table 4 presents the estimated DiD coefficient of interest (γ_1) in equation (3) obtained using two different sets of comparison schools. The analysis in the left panel uses all schools whose ESD rankings fell above the cutoff in 2014/15 whereas the bottom panel uses LUSD schools whose rankings fell right above the ESD cutoff (within 200 points in 2014/15). Both panels present the estimated effects (1) for all third graders in these schools and (2) for low-performing third graders (i.e., those whose reading test scores fell within 25 points around the retention cutoff – this is the sample used in the retention analysis in Table 3.

The estimated effects are all positive and slightly higher than the RD estimates presented in Figlio et al. (2018) although the coefficient of interest is imprecise in most specifications.[14] In particular, ESD designation increases following year reading scores of students by 0.09σ to 0.17σ (compared to 0.05σ reported in Figlio et al. 2018), yet the estimates are statistically indistinguishable from zero at conventional levels in all but one specification.

Effects of ESD Designation on Retention Effects

I then turn to the interaction effects between the two policies. Figure 3 presents a graphical portrayal of the main finding and illustrates the complementarities between the two interventions. In particular, this figure examines how the effect of being flagged for retention changed in the year after the ESD expansion versus the year before in treatment schools (in Panel A) and in comparison schools (in Panel B). In other words, results in these two panels can be

[14] This could be due to the fact that Figlio et al. (2018) examines the effects in the first year of ESD implementation (2012/13) in an RD framework using statewide data whereas this analysis uses the third and fourth years using a DiD approach in one Florida district.

regarded as the first and second differences in the DiD approach described above without any controls. In both panels, the black circles represent standardized following year reading scores averaged at the running variable level in the year after the policy changes (solid black lines represent the linear fitted lines estimated separately for the left of the cutoff and the right) whereas the gray triangles portray the same averages for the year before the policy changes (dashed gray lines are the linear fitted lines).

Panel (A) suggests that being flagged for retention had no significant effect on following year reading scores in treatment schools in the year before the ESD policy changes (i.e., before these schools were designated as ESD).[15] In contrast, students right below the retention cutoff in treatment schools significantly outperformed their peers on the other side of the cutoff on following year reading tests after the ESD expansion. I observe no such pattern in comparison schools: being flagged for retention had a similar effect on test scores before and after these policy changes in these schools.[16]

Table 5 presents the numbers behind these figures. In particular, Panel (A) repeats the analysis portrayed in Figure 3 and presents the discontinuities in following year reading scores at

---

[15] This is true not only for the year right before the ESD designation, but also in other prior years. For example, when I examine the effect of being flagged for retention between 2005/06 and 2013/14 in schools that were designated as ESD in 2014/15 school year, I find precisely estimated zero effects on following year reading scores (magnitude of the estimated discontinuity at the cutoff is 0.032σ with a p-value of 0.344. In contrast, the same discontinuity is 0.137σ (p-value<0.0001) for LUSD schools that had never been identified as ESD prior to (and including) 2014/15.

[16] A possible explanation behind the "no retention effect" for the lowest-performing schools (that are later designated as ESD) and positive effects in other schools in pre-ESD years is differences in educational resources (and/or how these resources are utilized). In other words, schools that were designated as ESD may not have the resources (e.g., reading coaches, instructional resources) to implement the retention policy effectively. In contrast, higher-performing schools (based on their students' reading performance) may have these resources, which may have contributed to the effectiveness of the retention policy. There is some evidence that show discrepancies in the effects of retention by school characteristics. For example, Schwerdt et al. (2017) show that the effects of retention on reading test scores are smaller (albeit not statistically different) in schools with higher third-grade failure rates (i.e., share of first-time third graders whose third-grade reading scores fell below the retention cutoff). While these are likely not the same schools as schools designated as ESD in this study (ESD schools are more likely drawn from the left tail of the school reading performance spectrum), these findings provide evidence about school-level heterogeneity in retention effects.

the retention cutoff, estimated separately for treatment and comparison schools in the years before and after the ESD expansion without and with student baseline characteristics listed in Table 2 in the first and second columns respectively. The findings in the first column reveal no significant effect of being flagged for retention on following year reading scores in treatment schools before the policy change while being flagged for retention increases test scores by 0.21σ in those schools after ESD designation. In contrast, the effect of being flagged for retention remains virtually unchanged in comparison schools over this time frame (0.161σ before the policy change versus 0.157σ after). These findings are robust to the inclusion of student baseline outcomes and characteristics, providing evidence that treated and comparison students are comparable around the cutoff.

Panel (B) in Table 5 presents the results (β_1-β_4) from the DiD in RD approach described in equation (6) without and with school-by-year level covariates. Overall, being flagged for retention increases following year reading scores by 0.16σ, yet this effect was significantly smaller (and virtually zero) in treatment schools before the ESD designation. The estimated parameter of interest, $\widehat{\beta_1}$, suggests that ESD designation increased the effect of being flagged for retention by 0.23σ. This interaction effect is sizable and slightly larger than the effect of being flagged for retention in comparison schools before the ESD policy. In other words, ESD designation closed the gap in the effect of being flagged for retention between lowest-performing schools in reading and others. Once again, these estimates are not sensitive to the inclusion of student baseline characteristics and school-by-year level covariates.

MECHANISMS

Are these interaction effects driven by the differential benefits of the ESD designation on the lowest-performing students in reading instead of ESD designation improving the retention effects in these schools? In other words, it is plausible that ESD designation leads these schools to focus on lowest-performing students in reading or that these students benefit more from the supports and services provided by the designation (although the latter hypothesis is unlikely to explain the discontinuity at the retention cutoff), regardless of the retention policy. If this is case, the interaction effects presented in Figure 3 and Table 5 reflect the heterogeneous effects of ESD designation rather than complementarities between the two policies.

To investigate this possibility, I make use of an unexpected pause in the retention policy in 2014/15 year in LUSD. In that year, FLDOE experienced a delay in the release of state test scores, which play an important role in the third-grade retention decisions.[17] Because of this delay, FLDOE provided flexibility to school districts in the enforcement of the retention policy. While some districts in the state still enforced the policy (albeit at a significantly smaller scale), LUSD decided not to retain students based on their third-grade score at the end of 2014/15: only 57 third graders were retained in the district at the end of that school year (or 0.4 percent of the first-time third graders in the district) compared to 1,067 third graders (7.7 percent) at the end of 2013/14 and 1,247 third graders (8 percent) at the end of the 2015/16 school year. Further, the state did not change the ESD designations at the end of 2014/15 so the ESD schools in 2014/15 were once again identified as ESD in 2015/16.[18]

Table 6 presents the results from this falsification exercise. In the top panel, I provide the results presented in the bottom panel of Table 5 (controlling for school-level covariates) where I use the first-time third graders in 2012/13 as the pre-policy cohort and the 2013/14 cohort as the

[17] See, for example, https://www.orlandosentinel.com/news/education/os-third-grade-reading-retention-scores-20160718-story.html for more information, accessed on 4/14/2023.
[18] This was primarily driven by the test switch from FCAT 2.0 (Florida Comprehensive Assessment Test) to FSA (Florida Standards Assessment) in 2014/15, which prevented the state from calculating student learning gains in reading, a key component in the identification of ESD schools.

post-policy cohort. In the bottom panel I repeat this analysis using 2012/13 cohort as the pre-policy and the 2014/15 cohort, for whom the retention policy was not enforced, as the post-policy cohort. If it is indeed the complementarities between the two policies that is driving the results in Table 5 (rather than heterogeneous effects of the ESD policy for the lowest performing students), one would expect significant interaction effects (β_1) in the top panel and no interaction effects in the bottom panel.

The results provide evidence supporting the complementarities theory. In the top panel, similar to Table 5, I find significant positive effects of ESD designation on the effect of scoring in the lowest achievement level on the third-grade reading test when the retention policy was enforced. When the policy was not enforced, I observe a significant drop in the effect of scoring in the lowest achievement level (as indicated by the coefficient on the interaction between "below the retention cutoff" and "after the policy"), yet this drop is not significantly different in treatment schools compared to comparison schools: $\widehat{\beta_1}$ in the top panel is about 4 times larger than $\widehat{\beta_1}$ in the bottom panel.

Are the results in Table 5 driven by differences in student baseline characteristics between schools, cohorts, and/or students around the retention cutoff? Table 7 presents the results of another falsification exercise where I replace the outcome of interest in Table 5 (following year reading scores) with student baseline characteristics including third-grade math scores, disciplinary incidents, absences, race/ethnicity, and special education status. The results reveal no (statistically or economically) significant effects on these baseline characteristics, providing further evidence about the validity of the main findings.

Are these results driven by the effect of ESD designation on student attrition among students flagged for retention? This could be an issue given the recent evidence suggesting that parents are more likely to exercise school choice and move their children whose older sibling is flagged for retention to another school if their school is designated as low performing (Figlio et al., 2023). Figure 4 addresses this question and examines the likelihood that the student changes schools at the end of third grade in the summer the policy change took effect (summer 2014) and the previous summer (summer 2013) in treatment schools and comparison schools. I find no significant influence of ESD designation on the effect of being flagged for retention on student mobility, with a $\widehat{\beta_1}$ of 0.033 (p-value: 0.684).[19]

Did ESD designation (and the accompanying resources) allow these schools to retain more students who would benefit from an additional year of instruction and supports? Figure 5 provides strong evidence that this was the case.[20] In particular, the top panel in this figure suggests the effect of scoring below the retention cutoff on being retained increased significantly in treatment schools with the policy changes while the change in this effect is much smaller in comparison schools. The DiD in RD estimates reveal that ESD designation increases the effect of being flagged for retention on being retained by 26 percentage points (p-value: 0.018).

Can the rise in enforcement explain the improvement in retention effects in treatment schools? In other words, did the effect of retention improve in these schools or was it the fact additional resources provided by the ESD designation allowed the treatment schools to retain more retention-eligible students (even though the benefits of retention did not significantly change)? Or could a more aggressive retention policy (in the absence of ESD) have led to similar

[19] I also examine the effect of being flagged for retention on the likelihood of being enrolled in an ESD school and find no significant effect.

[20] It is important to note that ESD designation can not affect the running variable in the analysis as students in all cohorts take the third-grade test before the treatment schools are designated as ESD schools. That said, ESD designation can change who gets retained among students flagged for retention as the retention decisions are typically made in the summer after third grade after ESD designations are announced. For example, ESD designation could affect the use of good cause exemptions (especially the use of more “subjective” exemptions such as the teacher-developed portfolio) to promote students who were flagged for retention.

results? To address the last question, I examine the effects of retention during pre-designation years (2005/06 through 2013/14) for treatment schools. In particular, Figure 6 breaks down the pre-ESD years for these schools into high-retention and low-retention years based on the magnitude of the discontinuity in the likelihood of retention at the cutoff (above-median versus below-median first stage during this time frame), and examines whether the observed effect of being flagged for retention was higher in high-retention years.

The results suggest that while the effect of being flagged for retention on the likelihood of being retained differs considerably between high-retention and low-retention years (a discontinuity of 0.39 with p-value<0.001 for the former versus a discontinuity of 0.27 with p-value<0.001 for the latter), the effects of being flagged for retention on following year reading scores are virtually identical (0.057σ with p-value: 0.370 for the high-retention years versus 0.064σ with p-value: 0.188). This provides evidence that simply retaining more students in ESD schools will not improve the effectiveness of retention in these schools without the resources provided by the ESD program.

To what extent does the effect of ESD designation on retention likelihood explain the observed influence of the ESD designation on the effect of being flagged for retention? Table 8 presents more direct evidence on the influence of ESD designation on the effects of being retained, estimated in a 2SLS approach using "scoring below the cutoff" as an instrument for "being retained".[21] The results suggest that ESD designation significantly improved the effects of being retained. In particular, being retained had no significant effect on following year reading scores in treatment schools before ESD designation, yet retention improved test scores by 0.31σ

[21] In particular, I first replace $B_{i,t-1}$ with $R_{i,t-1}$ in equation (6) and then estimate this equation in a 2SLS framework where I instrument for $R_{i,t-1}$ (and its interactions) using $B_{i,t-1}$ (and its interactions).

after the designation. For comparison schools, there was a decline in retention effects after the policy changes from a benefit of 0.54σ to 0.56σ to a benefit of 0.32σ to 0.37σ. Overall, ESD designation improved the effect of being retained by about 0.55σ, roughly equivalent to the gap in retention effects between treatment and comparisons prior to the policy changes.[22]

Is the retention effect improving in ESD schools (relative to non-ESD schools) because the designation changed the composition of retained students? In other words, are the main results driven by ESD schools retaining students who are more likely to benefit from retention? It is difficult to directly test this hypothesis, but in Table 9, I examine the effects of ESD designation on the third-grade outcomes and characteristics of the retained students estimated using equation (4), with all schools above the ESD cutoff serving as the comparison group.[23] The results reveal no significant effect of ESD designation on the composition of retained students.

Finally, could the stigma associated with being labeled as "low-performing" (rather than the ESD resources) have triggered ESD schools to implement different policies/practices that improved the effectiveness of the retention policy? This is certainly possible, but it is important to keep in mind that many of these schools had already been labeled as low-performing under Florida's high stakes school accountability system (see, for example, Rouse et al., 2013 for details about Florida's school accountability system). In particular, all of the ESD schools examined in this study had received a school grade of "C" or lower in the previous year, and roughly half of them received a near-failing ("D") or a failing grade ("F"). As such, it is unlikely that the ESD designation revealed new information about the performance of these schools and led to stigma.

[22] As such, about a third of the ESD influence on retention effect is driven by the decline in retention effects in non-ESD schools before and after the policy changes.
[23] In particular, I replace Y_{it} with the third-grade characteristics and outcomes (listed in Table 2) of retained students in equation (4).

Overall, these findings point to the benefits of retention improving in ESD schools (above and beyond the shifts in the volume or the composition of retained students) to be an important mechanism behind the observed interaction effects. And this improvement is more likely to be driven by the additional resources provided by the ESD designation that were relevant to reading instruction in these schools (e.g., reading coaches, additional hour of literacy instruction) rather than the stigma associated with the designation itself.

DIFFERENTIAL TRENDS, ROBUSTNESS TO ALTERNATIVE SPECIFICATIONS, AND EXTENSIONS

Differential trends could be an issue in the DiD in RD design if retention effects change differently over time in ESD and non-ESD schools prior to ESD designation. I follow two approaches to assess differential trends in this context. First, in the top panel of Figure 7, I examine the robustness of the main findings (1) to different student reading score bandwidths and (2) to restricting the sample of comparison schools to those right above the ESD cutoff. For the former, I use the optimal bandwidths obtained from the optimal bandwidth selection proposed by Calonico et al. (2017), bandwidth of 25, 30, 35, and 40 points.[24] For the latter, I restrict the set of comparison schools included in the analysis to those ranked within 400 to 1500 points (with increments of 100) above the ESD cutoff. The top panel present the β_1 coefficient in Table 5 (i.e., the coefficient on the "flagged for retention*treatment schools*after policy" variable). The estimated coefficients remain virtually unchanged under these alternative specifications, providing evidence suggesting that the results are unlikely to be driven by differential trends.

[24] To obtain the optimal bandwidths, I use the STATA command rdrobust with the following year reading scores as the outcome to obtain the optimal bandwidth. I then restrict the analysis sample to students within these test score bandwidths and estimate equation (1) to obtain the β_1 coefficient in Table 2.

In the bottom panel, I conduct an event study in RD exercise and present the estimated differences in the effect of being flagged for retention on following year reading scores between treatment and comparison schools for each school year between 2009/10 and 2017-18, with the year before the policy change (2013/14) serving as the baseline.[25] The findings suggest no evidence of differential trends in retention effects between treatment and comparison schools before the policy change and imply a significant improvement in the effect of being flagged for retention in treatment schools compared to comparison schools. There is no significant effect of ESD designation in the second year (due to the pause in retention policy as discussed in Table 6), yet these interaction effects re-emerge in the third and fourth year after the designation.

Finally, I examine the interaction effects on math scores, repeating the analysis in Table 5 using following year math scores. It is important to note that one could still expect complementarities in math (even though both interventions target reading achievement): If ESD designation improves the retention effects on reading achievement, improved reading skills could in turn improve math achievement. In fact, Florida's third-grade retention has been found to improve student test scores in math considerably (at least in the short term) even though it is a reading intervention (e.g., Schwerdt et al. 2017). The findings (available upon request) reveal that the complementarities are smaller in math and statistically indistinguishable from zero at 5 percent level in all cases, but they are still sizable. For example, the estimated effects suggest that ESD designation increases the effect of retention on following year math scores by 0.16σ (p-value: 0.209) to 0.19σ (p-value: 0.100).

CONCLUDING REMARKS

[25] I also repeat the same exercise using the set of first-time ESD schools in 2014/15 and dropping the 3 schools that were previously designated as ESD and whose accountability raking fell between 100 and 300 in 2014/15. The results remain virtually unchanged with no evidence of differential pre-treatment trends and positive effects after the designation (except for the second year).

Public policies targeting individuals based on need typically impose disproportionate burden on communities that lack the resources to implement these policies effectively. In this study, I examine whether other targeted policy interventions that focus on similar needs and provide resources to build capacity in these communities could generate complementarities by improving the effectiveness of individual-level interventions. I address this question using two educational interventions in elementary schools in Florida: the third-grade retention policy that targets lowest-performing third graders in reading and the ESD policy that targets the lowest-performing elementary schools in the state in reading. Both policies have been shown to improve student test scores in reading in the short term: In this study, I explore their interaction.

I find that the ESD policy significantly improved the effectiveness of the retention policy in targeted schools. In particular, using a difference-in-difference in regression discontinuity design and the plausibly exogenous expansion of the ESD policy, I find that being designated as an ESD school increases the effect of being retained on the following year reading scores of students by 0.56σ. To put number into perspective, this interaction effect is roughly equivalent to the gap between the lowest-performing and higher-performing schools in the effectiveness of the retention policy on lowest-performing third graders in reading before the ESD policy took effect. These findings suggest that supplementing individual-level educational interventions with school-level interventions with similar objectives that provide additional resources to schools with highest needs could create significant complementarities, which can be regarded as “free” from a public policy perspective.

One important policy question in this context is whether the same spillover effect could have been achieved by simply providing monetary resources to the lowest-performing schools in the state (rather than providing resources with “strings” and requiring these schools to implement certain interventions such as an additional hour of literacy instruction). The answer to this question depends heavily on how effectively these resources would be used (or whether these resources would even be used to improve reading achievement) in these schools. There is evidence in school autonomy literature suggesting that autonomy is more likely to be an effective lever to improve student outcomes in school settings with effective leadership (Jackson, 2023). Yet, many ESD schools may have lacked effective leadership (as suggested by their low performance), for example, due to higher rates of staff turnover, and hence providing these schools monetary resources without strings may not be as effective. In contrast, the ESD program provides resources while requiring these schools to implement intervention strategies that have been proven effective (e.g., extended instruction time) to improve student outcomes in the area of need (i.e., reading).

Overall, the findings presented in this study highlight the discrepancies in the effectiveness of student-level interventions between different school settings and the importance of providing targeted resources to schools serving disadvantaged student populations that are often disproportionately affected by these interventions. That said, it is important to note that the two policies examined here target the same student need (i.e., improving reading achievement). As such, the complementarities may not be as large in cases where the two interventions target different student needs or the supports provided to these schools by one of the interventions are not as relevant for the other intervention.

Acknowledgements

I thank the anonymous school district for providing the data used in the analysis, and for providing useful feedback. Note that the views expressed are those of the author and do not

necessarily reflect those of the anonymous district or the institution to which the author is affiliated.

REFERENCES

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

Corbett, J. (2015). Florida State Policy Brief. San Francisco, CA: WestEd. Available at: https://files.eric.ed.gov/fulltext/ED559735.pdf.

Cummings, A., & Turner, M. (2020). COVID-19 and Third-Grade Reading Policies: An Analysis of State Guidance on Third-Grade Reading Policies in Response to COVID-19. Education Policy Innovation Collaborative Policy Brief. Available at: https://epicedpolicy.org/wp-content/uploads/2020/10/RBG3-Reading-Policies-FINAL-10-29-20.pdf.

Figlio, D., & Ozek, U. (2024). The Unintended Consequences of Test-Based Remediation. American Economic Journal: Applied Economics, 16(1): 60-89.

Figlio, D., Karbownik, K., & Özek, U. (2023). Sibling Spillovers May Enhance the Efficacy of Targeted School Policies. National Bureau of Economic Research Working Paper 31406.

Figlio, D., Holden, K., & Özek, U. (2020). An extra year to learn English? Early grade retention and the human capital development of English learners. Journal of Public Economics, 186, 104184.

Figlio, D., Holden, K., & Özek, U. (2018). Do students benefit from longer school days? Regression discontinuity evidence from Florida's additional hour of literacy instruction. Economics of Education Review, 67, 171-183.

Folsom, J. S., Petscher, Y., Osborne-Lampkin, L., Cooley, S., Herrera, S., Partridge, M., & Smith, K. (2016). School reading performance and the extended school day policy in Florida (REL 2016–141). Washington, DC: U.S. Department of Education, Institute of

Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from http://ies.ed.gov/ncee/edlabs.

Folsom, J. S., Osborne-Lampkin, L., Cooley, S., & Smith, K. (2017). Implementing the extended school day policy in Florida's 300 lowest performing elementary schools (REL 2017–253). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from http://ies.ed.gov/ncee/edlabs.

Frandsen, B. R. (2017). Party bias in union representation elections: Testing the manipulation in the regression discontinuity design when the running variable is discrete. Advances in Econometrics, 38, 281–315.

Greene, J. P., & Winters, M. A. (2012). The medium-run effects of Florida's test-based promotion policy. Education Finance and Policy, 7(3), 305–330.

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

Jackson, C. K., & Mackevicius, C. L. (2024). What Impacts Can We Expect from School Spending Policy? Evidence from Evaluations in the United States. American Economic Journal: Applied Economics, 16(1): 412-46.

Jackson, C. K. (2023). When Does School Autonomy Improve Student Outcomes? (EdWorkingPaper: 23-808). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/cdj7-rg41.

Johnson, R. C., & Jackson, C. K. (2019). Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending. American Economic Journal: Economic Policy, 11(4), 310–349.

Kolesár, M., & Rothe, C. (2018). Inference in regression discontinuity designs with a discrete running variable. American Economic Review, 108(8), 2277–2304.

Lee, D., & Card, D. (2008). Regression discontinuity inference with specification error. Journal of Econometrics, 142(2), 655–674.

LiCalsi, C., Özek, U., & Figlio, D. (2019). The uneven implementation of universal school policies: Maternal education and Florida's mandatory grade retention policy. Education Finance and Policy, 14(3), 383–413.

Malamud, O., Pop-Eleches, C., & Urquiola, M. (2016). Interactions between Family and School Environments: Evidence on Dynamic Complementarities? NBER Working Paper no. 22112.

Mbiti, I., Muralidharan, K., Romero, M., Schipper, Y., Manda, C., & Rajani, R. (2019). Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania. The Quarterly Journal of Economics, 134(3), 1627–1673.

Muslimova, D., van Kippersluis, H., Rietveld, C. A., von Hinke, S., & Meddens, F. (2020). Dynamic Complementarity in Skill Production: Evidence From Genetic Endowments and Birth Order. Tinbergen Institute Discussion Paper 2020-082/V. Available at SSRN: https://ssrn.com/abstract=3748468 or http://dx.doi.org/10.2139/ssrn.3748468.

Özek, U. (2021). The effects of middle school remediation on postsecondary success: Regression discontinuity evidence from Florida. Journal of Public Economics, 203, 104518.

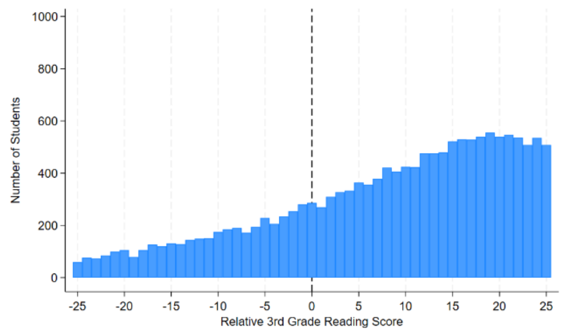
Ravitch, D., Forte, D., Moss, P., & Reville, P. (2022). Policy Dialogue: Twenty Years of Test-Based Accountability. History of Education Quarterly, 62(3), 337–352.

Rouse, C. E., Hannaway, J., Goldhaber, D., & Figlio, D. (2013). Feeling the Florida Heat? How

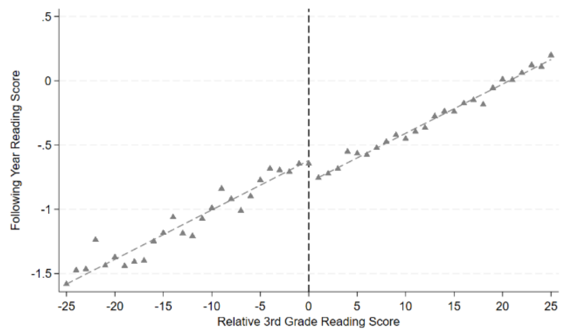
Low-Performing Schools Respond to Voucher and Accountability Pressure. American Economic Journal: Economic Policy, 5(2), 251–281.

Schwerdt, G., West, M. R., & Winters, M. A. (2017). The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida. Journal of Public Economics, 152, 154–169.

(A) Density of third- grade reading scores



(B) Following year reading scores around the cutoff



(C) Retention likelihood around the cutoff

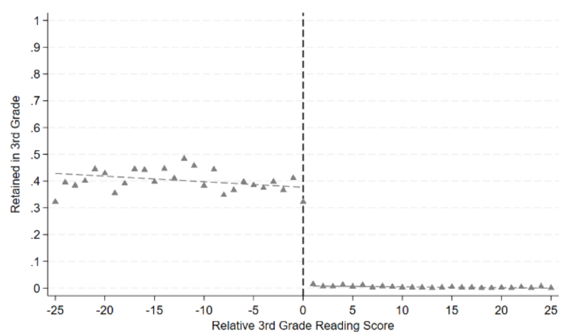


FIGURE 1 Distribution of third-grade reading scores, following year reading scores, and retention likelihood around the cutoff.

Notes: Panel (A) presents the number of students in each reading score bin between 25 points below and above the retention cutoff, which is shown by the vertical line. Panel (B) presents the raw cell means of the following year reading scores for each reading score between 25 points below and 25 points above the retention cutoff. The dashed lines represent the linear fitted lines estimated separately for the left of the retention cutoff and the right. Panel C repeats the same analysis using the retention indicator following year reading scores.

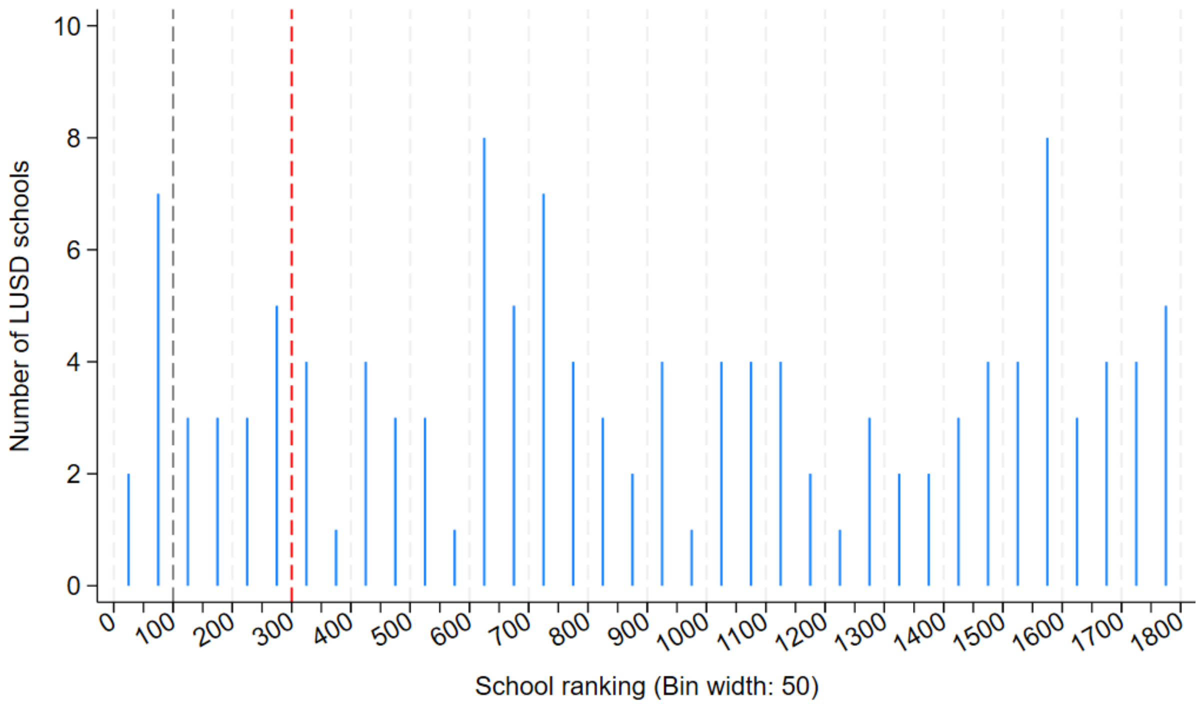
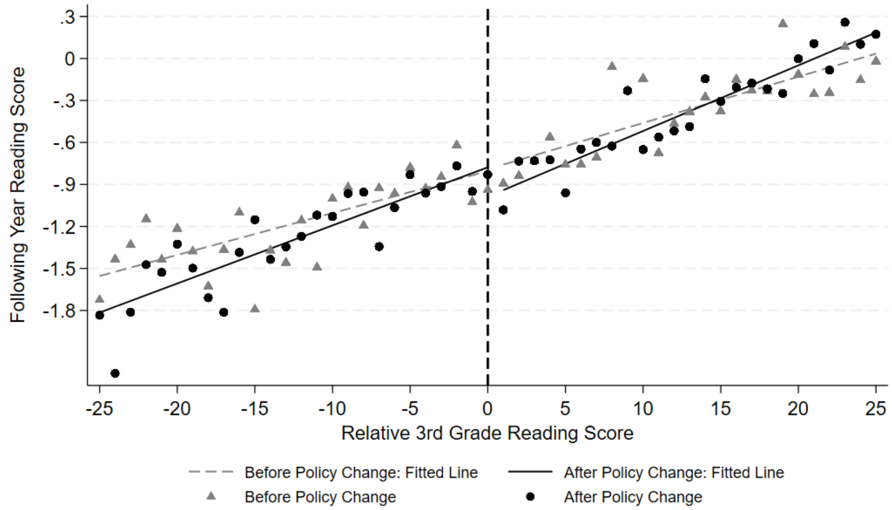


FIGURE 2 ESD rankings of elementary schools in LUSD in 2014/15.

Notes: The figure presents the distribution of LUSD elementary school rankings in the state in 2014/15 school year with a bin width of 50. The red vertical lines represent the ESD cutoff 2014/15 whereas the gray vertical line represents the lower bound for the treatment schools in 2014/15.

(A) Treatment schools



(B) Comparison schools

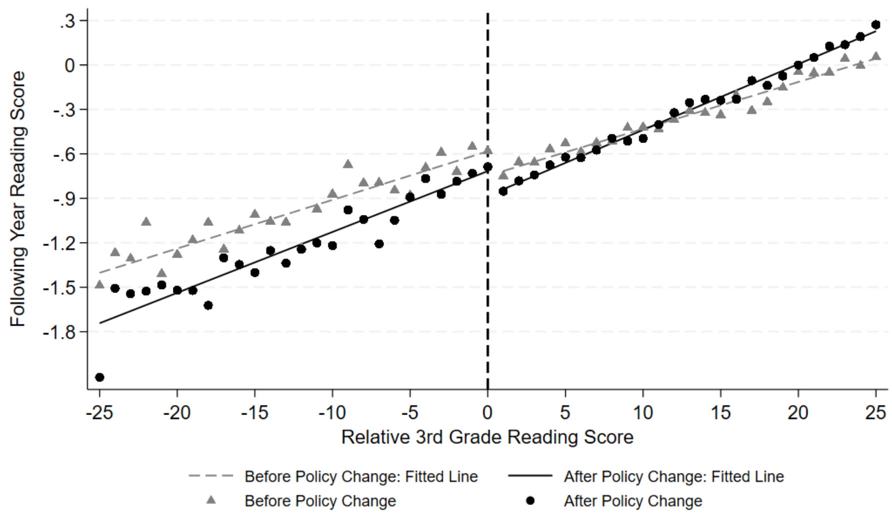


FIGURE 3 Effects of being flagged for retention on following year reading scores: treatment versus comparison schools, after versus before ESD policy change.

Notes: Panel (A) presents the raw cell means of the following year reading scores for each reading score between 25 points below and 25 points above the retention cutoff in treatment schools before (gray triangles) and after (black circles) the policy change. Panel (B) repeats the same analysis in comparison schools. The black solid line and the dashed gray lines provide the fitted lines estimated separately for observations below and above the cutoff, which is shown by the vertical line.

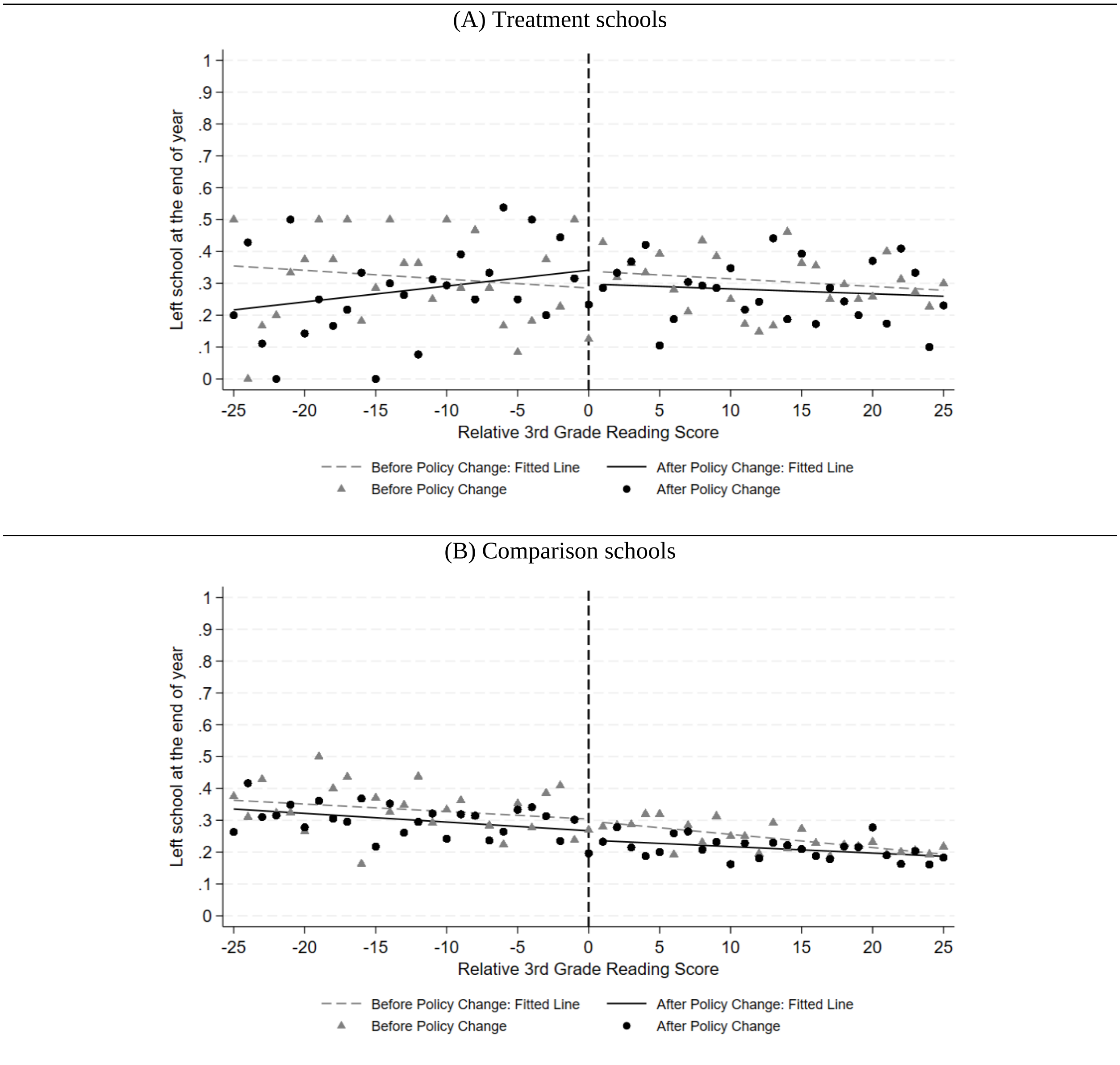
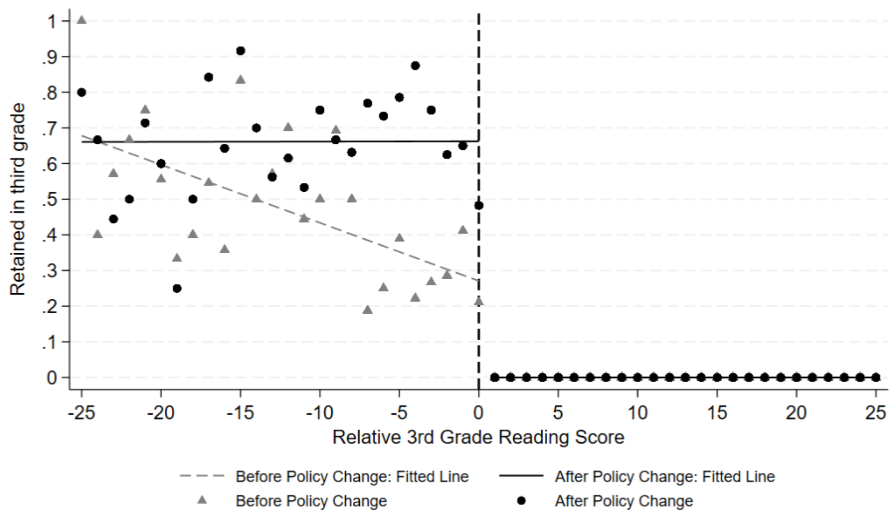


FIGURE 4 Effects of being flagged for retention on student attrition: treatment versus comparison schools, after versus before ESD policy change.

Notes: Panel (A) presents the raw cell means of the likelihood that the student leaves the school at the end of the third grade for each reading score between 25 points below and 25 points above the retention cutoff in treatment schools before (gray triangles) and after (black circles) the policy change. Panel (B) repeats the same analysis in comparison schools. The black solid line and the dashed gray lines provide the fitted lines estimated separately for observations below and above the cutoff, which is shown by the vertical line.

(A) Treatment schools



(B) Comparison schools

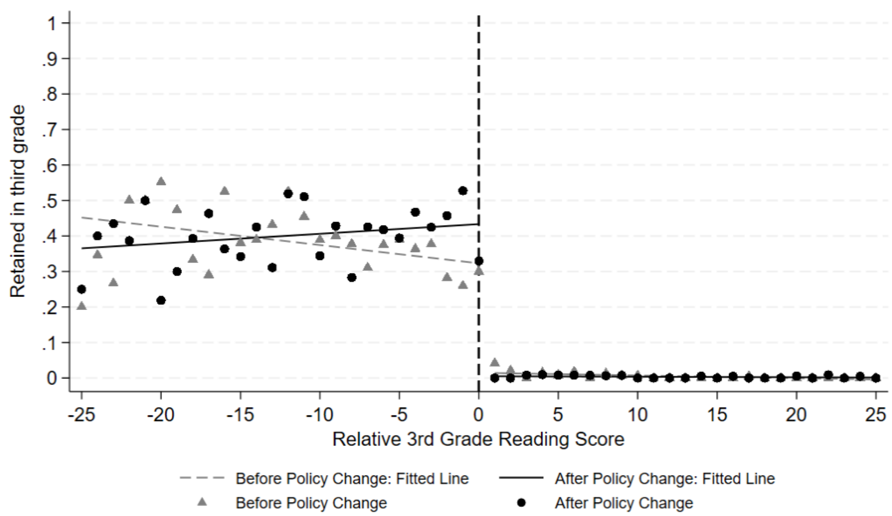


FIGURE 5 Effects of being flagged for retention on being retained: treatment versus comparison schools, after versus before ESD policy change.

Notes: Panel (A) presents the raw cell means of the likelihood of being retained at the end of the third grade for each reading score between 25 points below and 25 points above the retention cutoff in treatment schools before (gray triangles) and after (black circles) the policy change. Panel (B) repeats the same analysis in comparison schools. The black solid line and the dashed gray lines provide the fitted lines estimated separately for observations below and above the cutoff, which is shown by the vertical line.

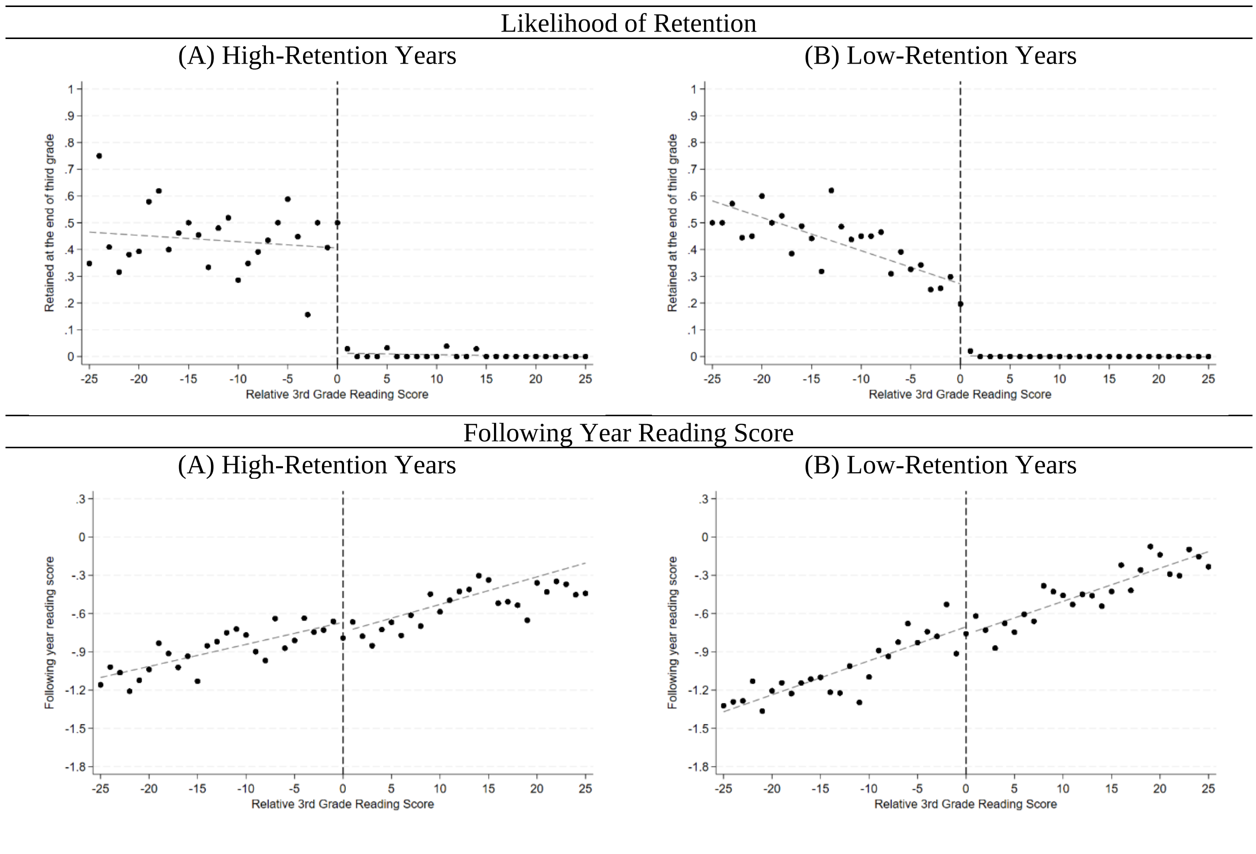
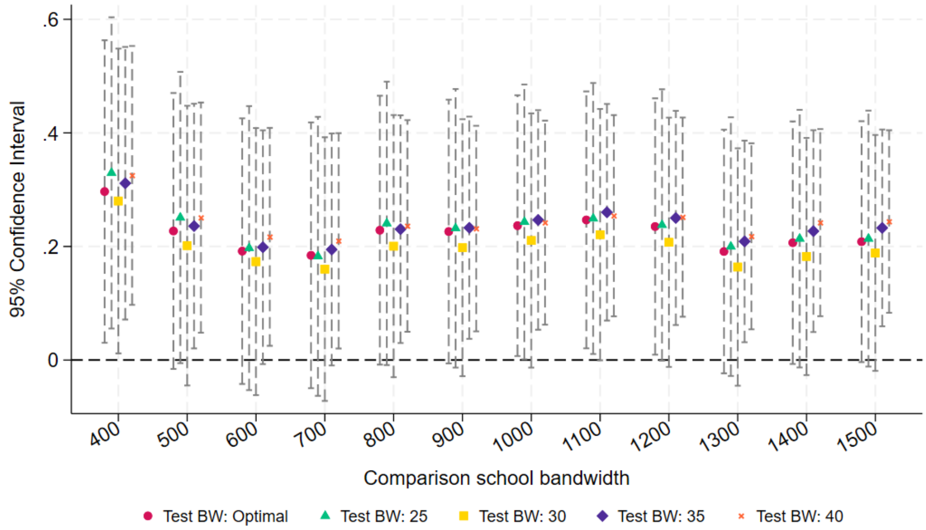


FIGURE 6 Likelihood of retention and following year reading scores in treatment schools, high-retention versus low-retention years between 2005/06 and 2013/14.

Notes: The top panel presents the raw cell means of the retention indicator for each reading score between 25 points below and 25 points above the retention cutoff in high-retention years (i.e., where the discontinuity in retention likelihood at the cutoff is higher than the median) and low-retention years between 2005/06 and 2013/14 school years in treatment schools. The bottom panel repeats the same analysis using following year reading scores. The dashed lines represent the linear fitted lines estimated separately for the left of the retention cutoff and the right.

(A) Sensitivity of complementarities to different reading score bandwidths and alternative comparison schools



(B) Event Study Estimates

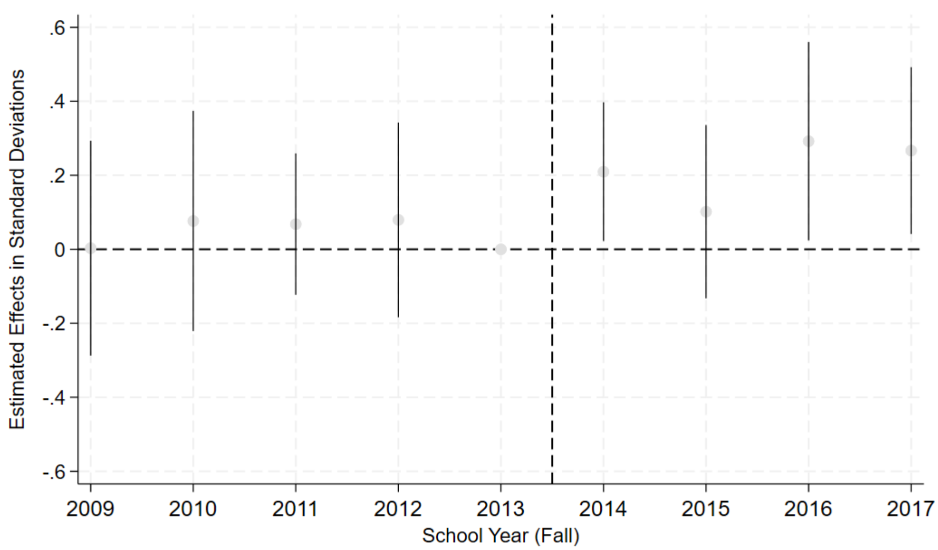


FIGURE 7 Sensitivity checks.

Notes: Panel (A) presents the β_4 coefficient in Table 5 (coefficient on "flagged for retention*treatment schools*after policy") estimated using the test score bandwidth given and the set of comparison schools within the bandwidth given on the x-axis above the ESD cutoff. Panel (B) presents the estimated difference in the effect of being flagged for retention on following year reading scores between treatment and comparison schools for each school year between 2009/10 and 2017-18, with the year before the policy change (2013/14) serving as the baseline. Spikes represent the 95 percent confidence intervals obtained using robust standard errors clustered at the school level. All regressions control for student baseline characteristics listed in Table 2 and school-by-year level covariates.

TABLE 1 Falsification exercise: the pseudo effect of being flagged for retention on student baseline characteristics.

	Third-grade outcomes						
	Math score	Disciplinary incident	% absent days	White	Hispanic	Black	Special education
Flagged for retention	0.017	0.001	-0.003*	0.026**	-0.001	-0.026	0.010
	(0.024)	(0.012)	(0.002)	(0.013)	(0.018)	(0.017)	(0.013)
Control mean at the cutoff		0.133	0.043	0.141	0.395	0.403	0.173
N				13,997			

Notes: The numbers represent the regression discontinuity estimates (μ in equation (2)) obtained using a bandwidth of 25 points in student reading scores and the third-grade outcome/characteristic given in the column as the outcome. Robust standard errors are given in parentheses. *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

TABLE 2 Student characteristics: first-time third graders in 2012/13 and 2013/14 school years.

			Students within 25 points around retention cutoff			
					School ranking ≤ cutoff + 400	
	Comparison schools	Treatment schools	Comparison schools	Treatment schools	Comparison schools	Treatment schools
Student third-grade characteristics						
Reading score	0.082***	-0.491	-0.521***	-0.712	-0.610***	-0.712
	(0.980)	(0.905)	(0.550)	(0.586)	(0.558)	(0.586)
Math score	0.081***	-0.417	-0.385***	-0.584	-0.502***	-0.584
	(0.983)	(0.935)	(0.772)	(0.797)	(0.776)	(0.797)
Involved in a disciplinary incident	0.063***	0.138	0.087***	0.151	0.137	0.151
	(0.242)	(0.345)	(0.282)	(0.358)	(0.344)	(0.358)
% absent days	0.039***	0.045	0.042***	0.046	0.045	0.046
	(0.0374)	(0.045)	(0.039)	(0.045)	(0.042)	(0.045)
Subsidized meal eligible	0.623***	0.967	0.760***	0.973	0.918***	0.973
	(0.485)	(0.179)	(0.427)	(0.163)	(0.274)	(0.163)
Special education	0.093*	0.104	0.132**	0.112	0.113	0.112
	(0.290)	(0.305)	(0.338)	(0.315)	(0.317)	(0.315)
Gifted	0.0844***	0.026	0.010*	0.005	0.008	0.005
	(0.278)	(0.159)	(0.101)	(0.074)	(0.088)	(0.074)
English learner	0.148***	0.251	0.233***	0.288	0.249***	0.288
	(0.355)	(0.434)	(0.422)	(0.453)	(0.432)	(0.453)
English non-native	0.306***	0.374	0.363	0.379	0.358	0.379
	(0.461)	(0.484)	(0.481)	(0.485)	(0.479)	(0.485)
U.S. born	0.944***	0.919	0.944***	0.917	0.947***	0.917
	(0.230)	(0.273)	(0.230)	(0.277)	(0.224)	(0.277)
Male	0.517	0.535	0.549	0.545	0.531	0.545
	(0.500)	(0.499)	(0.498)	(0.498)	(0.499)	(0.498)
White	0.329***	0.0630	0.238***	0.045	0.127***	0.045
	(0.470)	(0.243)	(0.426)	(0.207)	(0.333)	(0.207)
Black	0.213***	0.597	0.267***	0.613	0.464***	0.613
	(0.409)	(0.491)	(0.442)	(0.487)	(0.499)	(0.487)
Hispanic	0.375***	0.308	0.437***	0.316	0.373***	0.316
	(0.484)	(0.462)	(0.496)	(0.465)	(0.484)	(0.465)
Number of students	20,274	2,237	10,768	1,653	3,110	1,653
Number of schools	107	14	107	14	29	14

Notes: Standard deviations are given in parentheses. Third-grade reading and math scores are standardized to zero mean and unit variance at the cohort level. *, **, and *** imply that the corresponding student baseline attribute in the comparison schools is statistically different than treatment schools at 10, 5, and 1 percent levels respectively.

TABLE 3 Effects of being flagged for retention and being retained on following year reading scores.

	(I)	(II)
	Reduced-form estimates	
Flagged for retention	0.171***	0.161***
	(0.029)	(0.028)
N	13,997	
	IV estimates	
Retained	0.441***	0.414***
	(0.076)	(0.074)
Student baseline characteristics	No	Yes
N	13,997	

Notes: The top panel presents the effects of being flagged for retention on following year reading scores (μ) estimated using equation (2) and a bandwidth of 25 points using the first-time third graders in LUSD in 2012/13 and 2013/14 school years. The bottom panel presents the effects of being retained on following year reading scores (τ_1) estimated using equations (3-1) and (3-2). Robust standard errors are given in parentheses. Column (II) introduces the student baseline outcomes and characteristics given in Table 2 as covariates. *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

TABLE 4 Effects of ESD designation on following year reading scores.

	Including all schools above the ESD cutoff		Including non-ESD schools right above the ESD cutoff	
	All third graders	Third graders around the retention cutoff	All third graders	Third graders around the retention cutoff
ESD school x post-designation	0.090	0.142*	0.153	0.165
	(0.065)	(0.077)	(0.221)	(0.242)
N	19,791	12,347	3,658	2,801

Notes: The numbers represent the difference-in-differences estimates (γ_1 given in equation 4) obtained using OLS and equation (3). The left panel includes all schools above the ESD cutoff in 2014/15 as comparison schools whereas the right panel only includes schools above the ESD cutoff whose ESD rankings fell between 300 and 500 in 2014/15. Treatment schools in both panels include ESD schools whose rankings fell between 100 and 300 in 2014/15. In each panel, the first column includes all third graders (excluding students who scored in the highest achievement level on prior year reading test and hence were exempt from the longer school day) and the second column includes third graders whose third-grade reading scores fell within 25 points around the retention cutoff. Robust standard errors clustered at the school level are given in parentheses. All regressions control for third-grade student outcomes and characteristics listed in Table 2. *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

TABLE 5 ESD designation and the effect of being flagged for retention on following year reading scores.

	(A) RD estimates	
	(I)	(II)
Treatment schools		
Before policy change *[N= 791]*	-0.020	-0.053
	(0.099)	(0.102)
After policy change *[N= 854]*	0.209***	0.189**
	(0.081)	(0.079)
Comparison schools		
Before policy change *[N= 5,243]*	0.161***	0.169***
	(0.045)	(0.045)
After policy change *[N= 5,224]*	0.157***	0.135***
	(0.033)	(0.032)
	(B) DiD in RD estimates	
Without school-by-year level covariates		
Flagged for retention	0.161***	0.166***
	(0.050)	(0.049)
Flagged for retention*treatment schools	-0.182**	-0.189**
	(0.088)	(0.089)
Flagged for retention*after policy	-0.004	-0.023
	(0.065)	(0.063)
Flagged for retention*treatment schools*after policy	0.234**	0.243**
	(0.094)	(0.100)
With school-by-year level covariates		
Flagged for retention	0.148***	0.156***
	(0.046)	(0.046)
Flagged for retention*treatment schools	-0.168**	-0.173**
	(0.083)	(0.083)
Flagged for retention*after policy	-0.004	-0.025
	(0.063)	(0.061)
Flagged for retention*treatment schools*after policy	0.226**	0.232**
	(0.092)	(0.095)
Student baseline characteristics	Yes	No
N	12,380	

Notes: Panel (A) present the estimated effects of being flagged for retention on following year reading scores using a regression discontinuity design and a bandwidth of 25 points, separately for treatment and comparison schools in the years before and after the ESD policy change. Panel (B) presents the difference-in-differences in regression discontinuity estimates (β_1-β_4 given in equation 6) estimated using a bandwidth of 25 points in student reading scores. Robust standard errors in panel (A) and robust standard errors clustered at the school level in panel (B) are given in parentheses. Column (II) introduces the student baseline outcomes and characteristics given in Table I as covariates. *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

TABLE 6 Falsification exercise: interaction effects without retention.

	(I)	(II)
	Retention policy on for the post-policy cohort	
Flagged for retention	0.148***	0.156***
	(0.046)	(0.046)
Below the retention cutoff*treatment schools	-0.168**	-0.173**
	(0.083)	(0.083)
Below the retention cutoff *after policy	-0.004	-0.025
	(0.063)	(0.061)
Below the retention cutoff *treatment schools*after policy	0.226**	0.232**
	(0.092)	(0.095)
N	12,380	
	Retention policy off for the post-policy cohort	
Flagged for retention	0.143***	0.155***
	(0.046)	(0.046)
Below the retention cutoff *treatment schools	-0.153*	-0.169*
	(0.089)	(0.090)
Below the retention cutoff *after policy	-0.169***	-0.174***
	(0.059)	(0.058)
Below the retention cutoff *treatment schools*after policy	0.068	0.060
	(0.121)	(0.116)
Student baseline characteristics	Yes	No
N	12,590	

Notes: The top panel presents the difference-in-differences in regression discontinuity estimates (β_1-β_4 given in equation 6) estimated using a bandwidth of 25 points, third graders in 2012/13 as the pre-policy and third graders in 2013/14 as the post-policy cohort. The bottom panel repeats the same analysis using third graders in 2014/15, for whom the retention policy was not enforced, as the post-policy cohort. Robust standard errors clustered at the school level are given in parentheses. Column (II) introduces the student baseline outcomes and characteristics given in Table 2 as covariates. All regressions also control for the school-by-year level covariates. *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

TABLE 7 Falsification exercise: ESD designation and the pseudo effect of being flagged for retention on student baseline characteristics.

	Third-grade outcomes						
	Math score	Disciplinary incident	% absent days	White	Hispanic	Black	Special education
Flagged for retention	-0.041	-0.006	0.001	0.020	-0.012	-0.006	0.030
	(0.039)	(0.016)	(0.003)	(0.018)	(0.022)	(0.021)	(0.024)
Flagged for retention*treatment schools	0.061	0.048	0.007	0.014	0.000	-0.014	-0.025
	(0.104)	(0.049)	(0.008)	(0.034)	(0.062)	(0.054)	(0.045)
Flagged for retention*after policy	0.085	0.020	-0.005	-0.032	-0.003	0.026	-0.040
	(0.055)	(0.025)	(0.003)	(0.029)	(0.034)	(0.032)	(0.033)
Flagged for retention*treatment schools*after policy	-0.052	-0.040	-0.007	0.031	0.007	-0.024	0.068
	(0.118)	(0.073)	(0.009)	(0.057)	(0.091)	(0.112)	(0.077)
N				12,380			

Notes: The numbers represent the difference-in-differences in regression discontinuity estimates (β_1-β_4 given in equation 6) obtained using a bandwidth of 25 points in student reading scores and the third-grade outcome/characteristic given in the column as the outcome. Robust standard errors clustered at the school level are given in parentheses. *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

TABLE 8 ESD designation and the effect of being retained on following year reading scores.

	(A) Fuzzy RD estimates	
	(I)	(II)
Treatment schools		
Before policy change *[N= 791]*	-0.071	-0.184
	(0.347)	(0.354)
After policy change *[N= 854]*	0.313***	0.284**
	(0.121)	(0.118)
Comparison schools		
Before policy change *[N= 5,243]*	0.544***	0.562***
	(0.158)	(0.156)
After policy change *[N= 5,224]*	0.369***	0.315***
	(0.080)	(0.075)
	(B) DiD in RD estimates	
Without school-by-year level covariates		
Retained	0.544***	0.556***
	(0.165)	(0.163)
Retained *treatment schools	-0.615**	-0.633**
	(0.301)	(0.302)
Retained *after policy	-0.175	-0.223
	(0.195)	(0.189)
Retained *treatment schools*after policy	0.560*	0.595**
	(0.289)	(0.296)
With school-by-year level covariates		
Retained	0.492***	0.517***
	(0.155)	(0.154)
Retained *treatment schools	-0.563**	-0.575**
	(0.283)	(0.282)
Retained *after policy	-0.159	-0.216
	(0.188)	(0.183)
Retained *treatment schools*after policy	0.530**	0.556**
	(0.270)	(0.272)
Student baseline characteristics	Yes	No
N	12,380	

Notes: Panel (A) present the estimated effects of being retained on following year reading scores using a fuzzy regression discontinuity design and a bandwidth of 25 points, separately for treatment and comparison schools in the year before and after the ESD policy change. Panel (B) presents the difference-in-differences in regression discontinuity estimates obtained using a bandwidth of 25 points in student reading scores, and using scoring below the retention cutoff as an instrument for being retained. Robust standard errors in panel (A) and robust standard errors clustered at the school level in panel (B) are given in parentheses. Column (II) introduces the student baseline outcomes and characteristics given in Table 2 as covariates. *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

TABLE 9 Effects of ESD designation on the composition of retained students.

	Third-grade outcomes						
	Math score	Reading score	Subsidized meal eligible	Hispanic	Black	English learner	Special education
Treatment schools*after policy	0.010	0.014	0.017	-0.037	-0.008	0.024	0.003
	(0.081)	(0.060)	(0.030)	(0.040)	(0.040)	(0.076)	(0.061)
N				1,549			

Notes: The numbers represent the difference-in-differences estimates of the effect of ESD designation on the third-grade outcomes and characteristics of retained students. All regressions control for school-by-year level covariates. Robust standard errors clustered at the school level are given in parentheses. *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.