Program on Education Policy and Governance Working Papers Series

The Effects of Early Grade Retention on Student Outcomes over Time: Regression Discontinuity Evidence from Florida

Guido Schwerdt [1] Martin R. West [2]

PEPG 12-09

Harvard Kennedy School
79 JFK Street, Taubman 304
Cambridge, MA 02138
Tel: 617-495-7976 Fax: 617-496-4428
www.hks.harvard.edu/pepg/

[1] Ifo Institute, Poschingerstraße 5, 81679 M¨unchen, Germany. E-Mail: schwerdt@ifo.de, Tel: +49 89-9224-1310 Fax.: +49 89-9224-14606
[2] Appian Way, Gutman 454, Cambridge, MA 02138, USA. Email: martin west@gse.harvard.edu, Tel: +1 617-496-4803 Fax: +1 617-496-3095

The Effects of Early Grade Retention on Student Outcomes over Time: Regression Discontinuity Evidence from Florida

Guido Schwerdt* Martin R. West†

August, 2012

Abstract

A growing number of American states and school districts require students to meet basic performance standards in core academic subjects at key transition points in order to be promoted to the next grade. We exploit a discontinuity in the probability of third grade retention under a Florida test-based promotion policy to study the causal effect of retention on student outcomes over time. Regression discontinuity estimates indicate large short-term gains in achievement among retained students and a sharp reduction in the probability of retention in subsequent years. The achievement gains from retention fade out gradually over time, however, and are statistically insignificant after six years. Despite this fade out, our results suggest that previous evidence that early retention leads to adverse academic outcomes is misleading due to unobserved differences between retained and promoted students. They also imply that the educational and opportunity costs associated with retaining a student in the early grades are substantially less than a full year of per pupil spending and foregone earnings.

JEL-Code: H52, I21, I28
Keywords: Educational production, public schools, grade retention

We are grateful to the Florida Department of Education for providing the primary dataset for this study. We thank Matthew Chingos, Paul Peterson, Ludger Woessmann, and seminar participants at CESifo, Harvard University, and Mathematica Policy Research for helpful comments. Any errors are our own.

*Ifo Institute, Poschingerstraße 5, 81679 München, Germany. E-Mail: schwerdt@ifo.de, Tel: +49 89-9224-1310 Fax.: +49 89-9224-14606

†Appian Way, Gutman 454, Cambridge, MA 02138, USA. Email: martin_west@gse.harvard.edu, Tel: +1 617-496-4803 Fax: +1 617-496-3095

1 Introduction

Should students who fail to meet basic performance standards in core academic subjects be retained in the same grade? Roughly 10 percent of American students are retained at least once between kindergarten and eighth grade, with the incidence of retention concentrated among low-income students and traditionally disadvantaged minorities (Planty et al., 2009). Retaining students in the same grade is costly in terms of additional per pupil spending and foregone earnings, if students (as intended) spend an additional year in full-time public education as a result of being held back. Yet consensus is lacking as to whether retention yields benefits for students that could offset these costs and, if so, under what conditions.

Proponents of policies encouraging the retention of low-performing students contend that these students stand to benefit from an improved match of their ability to that of their peers and from the opportunity for additional instruction before confronting more challenging material. Critics, meanwhile, warn that retained students may be harmed by stigmatization, reduced expectations for their academic performance on the part of teachers and parents, and the challenges of adjusting to a new peer group. In fact, a large literature suggests that retained students achieve at lower levels, complete fewer years of school, and have worse social-emotional outcomes than observably similar students who are promoted.[1] Because retention decisions typically reflect student characteristics unobserved by the researcher, however, these studies are likely to suffer from severe selection bias.

In this paper, we use statewide administrative data covering Florida public schools in grade 3 to 9 to study the causal effect of third grade retention on future student outcomes up

[1]Influential studies in this area include Jimerson (1999) and Jimerson et al. (2000, 2002), and McCoy and Reynolds (1999). A survey of 47 empirical studies conducted between by Holmes (1989) concluded that retained students performed 0.19 to 0.31 standard deviations worse on various measures of academic achievement than similar students who were not retained. A meta-analysis of post-1990 studies Allen et al. (2009) found that, although most studies indicated negative effects of retention, a subset with more rigorous designs yielded more positive evidence.

to 6 years later. The Florida database has two key advantages for studying the consequences of grade retention. First, Florida since 2003 has required that schools retain third grade students failing to demonstrate basic proficiency on the state reading test unless the student is eligible for one of a specified set of exemptions. While similar policies elsewhere have led to non-linearities in the relationship between test scores and retention probabilities (e.g., Jacob and Lefgren, 2004, 2009), Florida's test-based promotion policy generates a true discontinuity in the probability of retention. We can therefore employ a standard regression discontinuity design to overcome the selection issues plaguing most existing research on this topic.

Second, the Florida database contains vertically scaled test scores in reading and math that make it possible to compare the achievement of students tested in different grades. The ability to make this comparison is essential because the counterfactual condition for students who are retained is to have been immediately promoted to the next grade. While often reported in the literature, same-grade comparisons conflate any effect of retention on achievement with the effect of being a year older at the time the relevant test is administered. As we demonstrate below, they will also be biased if students on the margin of retention have experienced prior grade retentions or other educational interventions with effects on achievement that fade out over time or are delayed.

It is important to note that the Florida policy required that retained students be given the opportunity to attend a summer reading program prior to the next school year and that they be assigned to a "high-performing" teacher and receive intensive reading interventions during that year. Our estimates of the policy's impact will therefore capture the combined effect of retention and these additional measures and may not be directly comparable to those of some previous studies of retention. Requirements that retained students receive remedial interventions are typical of test-based promotion policies currently in use and under consideration in other states and school districts, however, giving our results

considerable policy relevance.

Due to the availability of exemptions for students scoring below the promotion cut-off, as well as to the voluntary retention of some higher-scoring students, our regression discontinuity design is fuzzy and yields estimates local to students who comply with the policy. From a policy perspective, however, this local average treatment effect is likely to be the most relevant parameter. Teachers granting a low-scoring student an exemption or recommending that a student with higher test scores be retained presumably do so because they have strong views as to whether retention would be beneficial for the student in question. In the case of compliers, in contrast, the fact that retention occurs only as a result of the test-based promotion policy implies that local educators are uncertain about whether retention is desirable. Moreover, because the retention policy is based on reading scores alone, we can exploit variation the math achievement of compliers to provide suggestive evidence that our estimates are generalizable to a broader population in terms of third grade achievement.

Our analysis confirms that students who are retained in third grade under Florida's test-based promotion policy experience substantial short-term gains in both math and reading achievement. After two years, retained students outperform their same-age peers who were promoted by 0.42 standard deviations in reading and by roughly half as much in math. These positive effects fade out over time, however, becoming statistically insignificant in both subjects within five years. We also find that retention reduced the probability that a student would be retained in each of the four subsequent years. In contrast, we find no effects of third grade retention on student absences or special education placement rates.

These findings contribute to an emerging literature using quasi-experimental research designs to study the effects of retention in U.S. public schools.[2] Jacob and Lefgren (2004,

[2]In addition to the studies discussed in the text, Eide and Showalter (2001) use variation in kindergarten entry ages across states as an instrument for retention and conclude that retention increases high school completion and earnings for white students, although their results are not statistically significant. In a

2009) exploit a non-linearity in the relationship between test scores and retention probabilities in third, sixth, and eighth grades to study the impact of retention on achievement and high school completion of Chicago students. They find that retention and mandatory summer school had a small positive short-term effect on achievement for third graders but not for sixth graders. They also find that retention increased dropout rates for eighth graders but not for sixth graders. In a prior study of the Florida policy, Greene and Winters (2007) find that third grade retention improved student achievement after two years.[3] Taken as a whole, this evidence suggests that retention in higher grade levels may have detrimental effects on future student outcomes, but that early grade retention may be beneficial. We confirm that third grade retention in Florida improves student achievement in the short-run, while also showing that these initial benefits fade out over time.

Our evidence that third grade retention reduces the probability of retention in subsequent grades highlights an additional consequence of policies that increase retention rates in early grades and helps to clarify their costs. Specifically, we show that many of the students retained as third graders as a result of Florida's test-based promotion policy would otherwise have been retained in a subsequent grade. To the extent that later grade retention is in fact less beneficial, students who are retained earlier rather than later may particularly benefit from the policy. Overall, our results indicate that after six years students retained in third grade are, on average, only 0.74 grade levels behind their non-retained peers. The cost to the individual student due to foregone earnings of being retained in the early grades is therefore likely to be substantially less than a full year.

The paper proceeds as follows. Section 2 develops a statistical model of education production with potential grade retention that motivates our approach to studying reten-

comparative setting, Manacorda's (2012) regression discontinuity analysis finds that retention in junior high school increases dropout rates for Uruguayan students.

[3] In a follow-up paper Winters and Greene (2012) present evidence on medium-run effects of the Florida policy based on same-grade comparisons. As we discuss in Section 2, same-grade comparisons identify the combined effects of retention, age and years of schooling, but fail to identify the isolated effect of grade retention, which is the focus of our analysis.

tion effects. In Section 3 we describe the Florida policy and our data. Section 4 presents our identification strategy and provides graphical evidence supporting its validity, while Section 5 presents our findings concerning the effects of third grade retention on student outcomes. Section 6 concludes.

2 A Statistical Model of Education Production with Grade Retention

To motivate our empirical approach to identifying the causal effect of grade retention, we incorporate retention effects into a simple education production function that may describe the process by which our data are generated:

$$Y_{iag} = \sum_{t=1}^{a} \alpha_t + \sum_{t=6}^{a} \sum_{h=1}^{g} (\lambda + \beta_h) G_{ith} + \sum_{t=6}^{a-1} \sum_{h=1}^{g-1} \tau_{h(a-t)} I_{ith} + \nu_{iag} \quad (1)$$

where Y_{iag} is a measure of the achievement of student i in grade g at age a that can be decomposed into the cumulative effects of age, α_t, schooling, $\lambda + \beta_h$, the isolated effects of grade retention, $\tau_{h(a-t)}$, and an error term, ν_{iag}, capturing individual heterogeneity and error in the measurement of the student's "true" achievement when tested at age a in grade g. Note that the effect of schooling consists of an average effect of a year of schooling, λ, that is constant across grade levels and a grade-specific deviation from this average effect, β_h. The latter is introduced to allow for differential learning gains across grades. The history of grade levels attended at any age between age 6 and age a is captured by the set of indicators, G_{ith}, that take the value one if student i attended grade h at age t. Similarly, I_{ith} indicates whether student i was retained in grade h at age t. Note that we allow the effects of being retained, $\tau_{h(a-t)}$, to vary by grade level and to fade out over time.

This model serves to clarify the choice between same-grade and same-age comparisons to study retention effects, a point of debate in the literature (see, e.g., Allen et al., 2009). For

simplicity (and to correspond to our empirical analysis below), consider a study designed to estimate the effect of retention in grade 3 (i.e.,$I_{it3} = 1$) on student achievement. At least in theory, the outcome of interest can be defined as achievement when students first reach grade 4 (same-grade) or as achievement one year after potential grade 3 retention (same-age).

Consider first the same-grade comparison. The expected achievement in grade 4 at age A of students not retained in grade 3 is given by:

$$\begin{aligned} E[Y_{iag}|a = A, g = 4, I_{i,A-1,3} = 0] &= \sum_{t=1}^{A} \alpha_t + \beta_4 + \lambda + \sum_{t=6}^{A-2}\sum_{h=1}^{3}(\lambda + \beta_h)E[G_{ith}|\cdot] \\ &+ \sum_{t=6}^{A-2}\sum_{h=1}^{2}\tau_{h(A-t)}E[I_{ith}|\cdot]. \end{aligned} \tag{2}$$

If we assume that students retained in grade 3 cannot be required to repeat the grade twice, their expected achievement in grade 4 is:

$$\begin{aligned} E[Y_{iag}|a = A+1, g = 4, I_{i,A-1,3} = 1] &= \sum_{t=1}^{A+1} \alpha_t + \beta_4 + \beta_3 + 2\lambda + \sum_{t=6}^{A-2}\sum_{h=1}^{3}(\lambda + \beta_h)E[G_{ith}|\cdot] \\ &+ \sum_{t=6}^{A-2}\sum_{h=1}^{2}\tau_{h(A+1-t)}E[I_{ith}|\cdot] + \tau_{3(2)}. \end{aligned} \tag{3}$$

Differencing Equations (3) and (2) yields:

$$\Delta^{grade} = -\alpha_{A+1} - (\beta_3 + \lambda) + \Theta + \tau_{3(2)} \tag{4}$$

where $\Theta = \sum_{t=6}^{A-2}\sum_{h=1}^{3}\tau_{h(A+1-t)}E[I_{ith}|\cdot] - \sum_{t=6}^{A-2}\sum_{h=1}^{3}\tau_{h(A-t)}E[I_{ith}|\cdot]$.

The first term in Equation (4) captures the effect of being of age $A+1$ instead of A, while the second term reflects the average effect of an additional year of schooling, λ, plus the grade 3-specific deviation from this average effect, β_3. The third term captures the

effects of potential grade retention in grades 1 or 2, which is zero only if $\tau_{h(A-t)} = \tau_{h(A+1-t)}$. That is, any effects of prior retentions cancel out only if they do not fade out over time. Finally, the equation's last term, $\tau_{3(2)}$, represents the isolated effect of grade 3 retention on achievement two years later.

The same-grade comparison represented by Equation (4) therefore identifies the isolated effect of grade 3 retention, $\tau_{3(2)}$, only in the absence of any grade 3 specific year-of-schooling effect ($\beta_3 = -\lambda$) effect and age effect ($\alpha_{A+1} = 0$) and if any effects of prior grade retentions do not fade out. Although they are not explicitly modeled here, the potential implications of prior grade retentions readily extend to other interventions that affect student achievement prior to grade 3 and fail to persist fully over time. Even if the use of a (quasi-)experimental identification strategy insures that the incidence of such interventions is orthogonal to grade 3 retention, the fact that outcomes are measured at different time points for retained and non-retained students would influence the estimates of retention effects. The decay of achievement impacts is a pervasive pattern in the literature on educational production, suggesting that empirical estimates of retention effects based on same-grade comparisons are likely to be poor proxies of the isolated effects of grade retention even in the absence of grade-specific year-of-schooling and age effects.

In contrast, the same-age approach compares the expected achievement at age A of students who were in grade 3 at age $A-1$ to that of students who were not retained. For non-retained students, this expectation is again given by Equation (2). For retained students, it is:

$$\begin{aligned} E[Y_{iag}|a = A, g = 3, I_{i,A-1,3} = 1] &= \sum_{t=1}^{A} \alpha_t + \beta_3 + \lambda + \sum_{t=6}^{A-2} \sum_{h=1}^{3} (\lambda + \beta_h) E[G_{ith}|\cdot] \\ &+ \sum_{t=6}^{A-2} \sum_{h=1}^{2} \tau_{h(A-t)} E[I_{ith}|\cdot] + \tau_{3(1)}. \end{aligned} \tag{5}$$

First-differencing equations (5) and (2) yields:

$$\Delta^{age} = \beta_3 - \beta_4 + \tau_{3(1)}. \quad (6)$$

Equation (6) shows that a same-age comparison identifies the isolated effect of retention in grade 3 on achievement in the following year plus any effect of having attended grade 4 rather than having attended grade 3 for a second time. Such a grade-specific effect could arise due to differences between grades 3 and 4 in curricula, instructional time, or average teacher quality. Attending grade 3 a second time rather than attending grade 4 in the following year is however a direct consequence of being retained. In other words, $\beta_3 - \beta_4$ is part of the desired treatment effect. Δ^{age} therefore represents a meaningful causal effect of grade retention despite the fact that the two terms on the right-hand-side are not separately identifiable.

Despite its clear advantages in terms of isolating the causal effect of grade retention, implementing the same-age comparison approach requires an achievement measure that places students tested in different grades on a common scale.[4] Our analysis exploits the fact that Florida is one of a small number of states that provides vertically equated developmental scale scores for students tested at each grade level included in its statewide accountability program.

We provide evidence below that the achievement gains made by typical students on this scale are not uniform across grades. Thus, our estimates of Δ^{age} may vary with the number of years since treatment for at least two reasons: true fade out of retention effects and grade-specific effects on achievement conditional on the number of prior years of schooling. For example, our estimates of the effects of grade retention in grade 3 may decline over time if $\beta_h < \beta_{h+1}$ even if $\tau_{3(a)} = \tau_{3(a+1)}$. To back out an approximate estimate of the extent of true fade out of retention effects over time, we rescale the developmental

[4]National longitudinal studies tracking a grade cohort of students over time typically meet this requirement, but often lack credibly exogenous sources of variation in the probability of retention.

scale scores under the assumption that achievement gains are uniform across grades 3 to 10. We explain this rescaling in more detail at the end of the next section.

3 Institutional Setting and Data

In 2002, Florida's legislature mandated that third grade students scoring below level two (of five performance levels) on the Florida Comprehensive Assessment Test (FCAT) reading test be retained unless they qualify for one of six "good cause exemptions."[5] The Florida policy's exclusive focus on third grade reading distinguishes it from test-based promotion measures in Chicago and New York City, which include retention gates based on reading and math achievement at multiple grade levels. This focus reflects a common belief among educators that acquiring basic reading proficiency by third grade is essential for subsequent performance across disciplines, as well as the fact that third is the lowest grade included in the state testing program.

Students scoring below the level two cutoff may be granted an exemption from the policy if they fall into any of the following categories: students with disabilities whose Individualized Education Plan indicates that the state test is an inappropriate measure of their performance; students with disabilities who were previously retained in third grade; Limited English proficiency (LEP) students with less than two years of instruction in English; students who were retained twice previously; students scoring above the 51st percentile nationally on another standardized reading test; and students demonstrating proficiency through a portfolio of work. Since the 2004-05 school year, retained students have also been given the opportunity for a midyear promotion to fourth grade if they demonstrate mastery of necessary skills. In light of these exemptions, calling the Florida policy "test-based promotion" may be a misnomer. It would be more precise to say that,

[5]The description of the Florida program in this section is based on Office of Program Policy Analysis & Government Accountability (2006).

for students not in special education, a low test-score shifts the burden of proof such that educators need to make an affirmative case that the student should be promoted. Across the first six cohorts of third graders impacted by the policy, a slight majority (52.2 percent) of students failing to meet the promotion standard received an exemption.

Even so, the policy sharply increased the number of students held back in third grade. The number of Florida third graders retained jumped to 21,799 (13.5 percent) as the policy was implemented in 2003, up from 4,819 (2.8 percent) the previous year. The number of Florida students retained in third grade fell steadily over the next five years, reaching 9,562 (5.6 percent) in 2008, primarily due to a reduction in the number of students failing to meet the promotion standard.

As noted above, the policy includes several provisions intended to ensure that retained students acquire the reading skills needed to be promoted the following year. First, retained students must be given the opportunity to participate in their district's summer reading camp. Schools must also develop an academic improvement plan for each retained student and assign them to a "high-performing teacher," as determined by student performance data and satisfactory performance appraisals. Finally, during their retained year, retained students must receive intensive reading interventions including ninety uninterrupted minutes daily of research-based reading instruction.[6] A lack of detailed information on the take-up and implementation of these additional requirements makes it impossible to disentangle their separate effects.

The data for our analysis are drawn from the Florida Department of Education's PK-20 Education Data Warehouse and contain information on all Florida students attending public schools in grades 3 to 10 from the 2000-01 through 2008-09 school years. We identify retained students based on the grade level of the state tests taken in adjacent years.[7] Our data extract includes the school each student attends and its location; student characteris-

[6]Since 2004-05, the uninterrupted ninety minute reading block has been mandatory for all K-5 students.
[7]Students receiving mid-year promotions after 2004-05 will therefore be recorded as not being retained.

tics such as ethnicity, gender, special education classification, English proficiency, and free lunch eligibility; annual measures of absences; and annual FCAT math and reading test scores.

Table 1 documents the structure of our data on student cohorts impacted by the test-based promotion policy. The first relevant cohort (which we will refer to as the 2003 cohort) entered third grade for the first time in the 2002-03 school year and can be followed for an additional six years after potential grade 3 retention, at which point promoted students who were not retained in a later grade should have reached ninth grade. [8] The right-most column indicates that roughly 13 percent of the 2003 cohort were retained as third graders; six years later, the vast majority of these students were enrolled in eighth grade, but some were in grade seven (indicating that they had been retained a second time) or in grade nine (indicating that they had subsequently skipped a grade level). Among students not retained in third grade, most had progressed to ninth grade but a substantial number (five percent of the original cohort) were in eighth grade. The differing patterns of grade progression observed for students retained and promoted as third graders motivate our analysis below of the causal effect of third grade retention on the probability of being retained in a subsequent grade. The five additional student cohorts included in our analysis enter third grade in later years and can therefore be tracked for progressively shorter time period. The left-most column shows that, on average, 8 percent of all students in our sample were retained in grade 3.

The first relevant cohort (which we will refer to as the 2003 cohort) entered third grade for the first time in the 2002-03 school year and can be followed for an additional six years after potential grade 3 retention, at which point promoted students who were not retained in a later grade should have reached ninth grade. [9]

[8] All cohorts are defined by the spring of the school year that students are observed in grade 3 in the Florida data for the first time.

[9] All cohorts are defined by the spring of the school year that students are observed in grade 3 in the

Table 2 provides summary statistics for our pooled sample including the 2003-2008 cohorts. The first column reports mean characteristics (measured in third grade) for all students; columns 2 and 3 in turn include all students scoring below the cutoff and all students who were retained in third grade; and column 4 includes students who were retained in third grade despite exceeding the cutoff. Students' raw third grade test scores in reading and math have been standardized by subject and year to have a mean of zero and a standard deviation of one. Naturally, students scoring below the cutoff and retained students perform at low levels. For example, retained students score 1.43 standard deviations below the average student in reading and 1.22 standard deviations below the average student in math. Students scoring below the cutoff and retained students are quite similar with respect to their observable characteristics. In contrast, the relatively few voluntarily retained students are better performing on average, more likely to be white, and substantially younger than the average retained student. They are also more absent more frequently as third graders, perhaps suggesting the importance of behavioral indicators to voluntary retention decisions.

In addition to raw test scores, our data extract includes vertically equated Developmental Scale Scores (DSS) intended to support comparisons of student achievement across grade levels. During the 2000-01 school year, when the FCAT testing program was expanded to include reading and math in all grades three through ten, a special data collection scheme incorporated the use of common items administered to students across multiple grades. Specifically, operational items from each grade's test were also included on the test administered to the higher and lower adjacent grade. These common items provide a basis for the use of Item Response Theory (IRT) methods to place results from each grade's test on a common scale.[10]

Figure 1 plots average DSS scores in reading and math by grade for all students in the

Florida data for the first time.

[10]See Hoffman et al. (2001) for technical details on the construction of the developmental scale scores.

pooled dataset. The DSS scores have an across-grade, student-level standard deviation of 364 points in reading and 305 points in math. The jagged trajectory evident in both subjects indicates that average achievement gains as measured by developmental scale scores vary considerably by grade. For example, math gains are very small in grade six while reading gains are particularly pronounced in grade four. This variation could reflect imperfections in the vertical scaling process. Alternatively, it could reflect true differences in the average rate of learning in Florida public schools across grades. For example, the small math gains in grade six likely reflects the fact that most Florida students transition into a middle school in grade six, which Schwerdt and West (2011) show has a causal impact on their achievement growth. To the extent that retention simply delays students from experiencing a grade in which their own achievement growth is likely to be smaller, policymakers arguably would want to incorporate this information into the metric used to compare their achievement to that of promoted students.

The variation in achievement gains by grade motivates our construction of an alternative vertical scaling of reading and math achievement, which is also plotted in Figure 1. Specifically, we subtract from each student's DSS score the grade-specific mean score and then add the predicted value for each grade from a linear regression of mean scores on grade level. These rescaled scores increase linearly from grades three to ten by construction. The estimated slope coefficients, which indicate the average annual rate of achievement growth between third and tenth grade, are 80 DSS points in reading and 83 DSS points in math.

Using these rescaled scores as the outcome measure when estimating the impact of retention on student achievement enables us to back out an approximate estimate of the extent of true fade out of retention effects over time. In terms of the statistical model presented in section 2, we treat the estimated slope parameters as a measure of λ and the difference the between average and predicted DSS score for grade h as an approximation of β_h. The assumption of linear achievement growth underlying the rescaling is admittedly

arbitrary, and point estimates based on rescaled scores do not necessarily represent an unbiased estimate of the isolated retentions effects, $\tau_{3(a)}$. However, comparing estimates based on rescaled scores across years should be informative about the rate at which retention effects fade out over time.

4 Empirical Strategy

Empirical strategies that rely on a selection-on-observables assumption will fail to provide unbiased estimates of the effect of early grade retention on future student outcomes if students are selected for retention based on factors unobserved by the researcher that influence educational outcomes. We address this concern by taking advantage of Florida's test-based promotion policy, which leads to a discontinuous relationship between third grade reading test scores and the probability of grade retention. This discontinuity generates plausibly exogenous variation in retention, which we exploit to identify the causal effect of grade retention on future outcomes.

4.1 Graphical Evidence

Our identification strategy hinges on the assumption that Florida's test-based promotion policy generates exogenous variation in third grade retention which we can use standard regression discontinuity methods to exploit. We first present graphical evidence of the existence of a discontinuity in the relationship between a student's third grade reading test scores and the probability of being retained. We then discuss potential threats to the validity of regression discontinuity studies and provide additional graphical evidence demonstrating that these threats are not applicable in this setting (c.f., Lee and Lemieux, 2010). Unless otherwise noted, all figures are based on the pooled data set of students in the 2003-2008 cohorts.[11]

[11] Cohort-specific graphs are available from the authors upon request.

Figure 2, which plots the share of students retained as a function of third grade reading scores (measured relative to the test score cutoff), provides visual evidence of the discontinuity in retention probabilities. The data points represent the share of students retained for each possible score on the third grade reading test, with each marker's size proportional to the number of students receiving that score. The solid line represents predicted values from separate local linear regressions on either side of the cutoff. For students 35 or more points ($>$ 1 standard deviations) below the cutoff, retention probabilities are relatively stable at just under 0.6. The probability of retention then declines as test scores increase, with retention probabilities immediately to the left of the cutoff approaching 0.3. Retention probabilities drop sharply to less than 0.05 at the cutoff, however, and approach zero 50 points above it.

Figure 3 displays the same relationship for the two cohorts of students in our data extract entering third grade immediately prior to the introduction of the test-based promotion policy. Note that the probability of retention for students in these cohorts rarely exceeds 20 percent, even for very low-scoring students. More importantly, the probability of retention is essentially continuous around the cutoff, indicating that the discontinuity evident in Figure 2 was in fact generated by the policy change.

While Figure 2 is based on the full distribution of third grade reading test scores, we limit our regression discontinuity analysis of the causal effects of retention to a narrower sample of students within a 10 test-score-point bandwidth on either side of the cutoff. Figure 4 illustrates the discontinuity within this more restricted sample, again plotting the fraction of students retained by third grade reading test scores measured relative to the cutoff. Local linear regressions on either side of the cutoff suggest an approximately linear relationship between test scores and retention probabilities in the cutoff region. However, the slope of this relationship clearly differs for students below and above the cutoff. We make use of this observation below when specifying the functional relationship between the

forcing variable (reading test scores) and the retention indicator in our empirical model.

A common concern with regression discontinuity analyses is the possibility of precise manipulation of the forcing variable around the cutoff (c.f., Urquiola and Verhoogen, 2009). In this setting, for example, one might worry that teachers were able to manipulate students' reading scores to push them just above the promotion cutoff. The fact that the FCAT reading test is objectively scored without teacher input makes this possibility unlikely, however, and Figure 5 confirms that the overall distribution of reading test scores shows no evidence of a heaping of observations around the cutoff.

The regression discontinuity identification strategy also assumes that there are not discontinuities in other characteristics associated with student outcomes at the cutoff. Figure 6 addresses this issue by plotting the mean value of the observable student characteristics available in our data against third grade reading test scores. In addition to examining each characteristic individually, we also use a probit model to generate a predicted retention probability for each student based on all available background characteristics (except reading scores). The figure confirms the absence of discontinuities in observed student characteristics at the test-score cutoff used to inform retention decisions.

Finally, we confirm that attrition from the Florida database in subsequent years also does not vary discontinuously at the promotion cutoff. Even in the absence of sorting around the cutoff based on prior characteristics, differential attrition could occur if, for example, being retained in third grade made students more likely to leave the Florida public schools. Figure 7 therefore plots attrition rates against third grade reading scores around the cutoff. To enhance legibility, the figure plots attrition rates after two, four, and six years only; the patterns after three and five years are similar.[12] Attrition rates increase as expected with the number of years since potential third grade retention, but

[12]Because we identify students as having been promoted or retained in third grade based on the grade in which they are observed the following year, attrition rates one year after potential retention are zero by construction.

they appear to be unrelated to third grade reading scores and there is no evidence of a discontinuity at the promotion cutoff.[13]

4.2 Estimation

Because only a subset of students scoring below the cutoff in reading test scores were actually retained, our empirical analysis takes the form of a fuzzy regression discontinuity design which can be implemented via instrumental variables (IV) estimation. In our preferred specification we estimate the causal effect of early grade retention on future student outcomes in a two-stage least squares model. The first stage is given by the following equation:

$$\begin{aligned} retain &= \gamma_1 below + \gamma_2 below \times LEP + \gamma_3 below \times SpEd \\ &+ \gamma_4 below \times forcevar + \gamma_5 forcevar + \Gamma X + \epsilon, \end{aligned} \tag{7}$$

where $retain$ indicates retention in grade 3, $below$ indicates that the student scored below the promotion cutoff on the grade 3 reading test, LEP identifies students with limited English proficiency in grade 3, $SpEd$ indicates whether students are classified as special education students in grade 3, $forcevar$ measures student achievement on the grade 3 reading test, X is a vector of student demographic characteristics including the student's math achievement in grade 3, and ϵ is a standard zero-mean error term. Note that we model the relationship between reading scores and the retention indicator as a linear relationship with a break in trend at the cutoff, because of the graphical evidence of this relationship in Figure 4.

[13] In addition to the graphical analyses in figure 6 and 7, we used each student characteristic and attrition in each year after potential third grade retention as the outcome variable in regressions with the same specification and bandwidth as our preferred regression discontinuity model. The results (available upon request) confirm the absence of any statistically significant breaks in the relationship between reading scores and these outcomes at the promotion cutoff.

The corresponding second stage of our 2SLS model is given by:

$$y = \delta_1 retained + \delta_2 below \times forcevar + \delta_3 forcevar + \Delta X + \eta, \quad (8)$$

where y denotes the student outcome of interest. Note that we achieve identification of δ_1 by instrumenting for grade retention in grade 3 (*retained*) with the indicator for being below the cutoff for promotion to grade 4 (*below*) and the interactions with LEP and special education status. As noted above, we estimate the 2SLS model for the sample of students within ten test score points on either side of this cutoff. We select this bandwidth based on the optimal bandwidth algorithm developed by Imbens and Kalyanaraman (2009) and demonstrate the robustness of our results to alternative bandwidths below. In order to compare our preferred IV results with conventional estimates of the effects of retention based on a selection-on-observables assumption, we also estimate Equation (8) using OLS. To maximize comparability across the two designs, we also limit the OLS specification to the regression discontinuity sample.

5 Results

Table 3 reports results from estimating the first-stage model in Equation (7) for each cohort of students separately and for the pooled sample. For purposes of comparison, we also present results for the two cohorts of students in our data who were not impacted by the policy. Note that all estimations are based on our preferred discontinuity sample within a 10 test-score-point bandwidth around the cutoff. Despite this narrow bandwidth, we still have between 9,981 and 15,687 students in each post-2002 cohort and a total of nearly 75,000 students in the pooled sample.

The first row of Table 3 presents estimates of the jump in the probability of retention at the promotion cutoff for non-special education, non-LEP students. Consistent with Figure 3, the first two columns confirm that there was essentially no such jump in the two

years immediately preceding the policy's introduction.[14] In contrast, each of the cohort-specific estimates for students impacted by the policy is positive and highly statistically significant, with F-statistics on the excluded instruments exceeding 100. Point estimates of the jump in retention probabilities at the cutoff range from 0.22 to 0.40, with the largest estimate observed for the initial 2003 cohort and the two smallest estimates observed for the 2007 and 2008 cohorts. This suggests that compliance with the retention requirement was relatively weak (a pattern which is arguably consistent with the availability of good cause exemptions) and appears to have declined over time. The overall first stage effect for the pooled sample nonetheless indicates an increase of 0.31 in the probability of retention for typical students scoring immediately below the cutoff, relative to students scoring one point higher. The results also indicate that the increase in retention probabilities for students just missing the cutoff was smaller for special education and, to a lesser extent, LEP students. This is as expected given that students in these groups were eligible for additional good cause exemptions from the retention requirement.

5.1 The Effect of Early Grade Retention on Student Achievement

We begin our discussion of the effects of grade retention on student outcomes with graphical evidence on the reduced form relationship between students' third grade reading test scores and their future achievement. Figures 8 and 9 use local linear regressions estimated separately on each side of the promotion cutoff to depict the relationship between students' third grade reading test scores and their reading and math achievement up to six years after potential third grade retention. In both subjects, we observe students scoring below the promotion cutoff performing at higher levels in the first three years after potential third grade retention. However, these differences dissipate in later years and, in some cases,

[14] Although the results for the 2002 cohort show a statistically significant increase in the probability of retention for students scoring below the cutoff, the cohort-specific estimates while the policy was in place are all more than ten times as large.

appear to turn slightly negative.

Table 4 presents estimates of the effects of third grade retention on reading and math achievement over time. Columns 1 and 2 report OLS estimates from Equation (2) with and without covariates, while columns 3 and 4 report results from our preferred IV model exploiting the discontinuity. As expected, the inclusion of covariates does not notably influence the IV point estimates (although it modestly improves their precision) but substantially alters the OLS results.

Consistent with Figures 8 and 9, the IV estimates indicate that third grade retention substantially improves students reading and math achievement in the short run. Measured relative to the statewide standard deviation in third grade reading DSS scores, reading achievement improves by 26 percent of a standard deviation after one year and by as much as 50 percent of a standard deviation after two years. The estimated impact of retention on math scores is 31 percent of a standard deviation after one year and grows to 36 percent of a standard deviation after three years. These initial benefits fade out in subsequent years, however. The effects of third grade retention on reading scores are reduced in years three and four and become statistically insignificant in years five and six. In the case of math achievement, the estimated effects become slightly negative in years four and five but are statistically insignificant after six years. Appendix Table A-1, which presents the same year-by-year results separately for each cohort, confirms that this apparent fade out in the effects of grade retention over time does not simply reflect smaller impacts of retention on the earliest cohorts whose outcomes we are able to observe for more years.

Relative to our preferred IV estimates, OLS estimates of the effects of third grade retention are always more negative and would suggest a statistically significant negative impact on reading and math achievement after 6 years. The differences across the two sets of results are substantial even after including performance and demographic covariates. In reading after one year, for example, the difference between the OLS and IV point estimates

is more than one third of a standard deviation. Because the sample for both models is limited to students within a narrow range of just five percent of a standard deviation of third grade reading scores, this difference is unlikely to stem from the fact that the IV estimate is local to compliers at the promotion cutoff. We instead conclude that OLS estimates fail to control adequately for unobserved confounding factors and, thus, understate any benefits (and exaggerate any harms) of grade retention.

One unusual aspect of the results in Table 4 is the non-monotonic relationship between the size of the estimated impacts of retention and the time elapsed since the student was retained. The estimated impact is largest after two years in the case of reading achievement and after three years in math. Especially given the overall pattern of fade out, one would expect the impact of retention to be largest in the year the student was retained. This pattern likely stems in part from the grade-to-grade variation in the average achievement gains of Florida public school students as measured by DSS scores. For example, Figure 1 shows that Florida students statewide experience particularly large gains in DSS reading achievement in fourth grade, which promoted students enter immediately and (most) retained students enter one year later. This difference in timing could explain the unexpected growth from year one to year two in the estimated impact of retention on DSS reading achievement. The alternative scaling of the DSS scores discussed above eliminates variation in average achievement gains across grades and thereby allows us to approximate the true rate of fade out over time.

Table 5 presents OLS and IV estimates of Equation (8) based on these rescaled DSS scores. In both reading and math, the magnitude of the estimated impacts now decreases monotonically with distance from treatment. In reading, the impacts are as large as 61 percent of a standard deviation after one year but fade to 14 percent of a standard deviation by year four and are statistically insignificant thereafter. In math, the impacts start at 43 percent of a standard deviation but are statistically insignificant by year four and

become modestly negative after six years. Qualitatively, however, the results concerning achievement impacts of third grade retention do not depend on the test scaling. Both approaches show large positive initial impacts of retention that fade out completely over time.

5.2 The Effect of Early Grade Retention on Grade Progression, Absences, and Special Education Placement

We next present estimates of the effect of third grade retention on subsequent grade progression, absences from school, and special education placement rates. Grade progression is an important outcome to consider when evaluating test-based promotion policies for at least two reasons. First, it has direct implications for retention's costs to both the individual and society. If early grade retention influences the probability that students are retained at higher grade levels, the cost of early grade retention in terms of foregone earnings and additional educational expenditures could be well below a full school year. Second, the effects of retention on outcomes such as student achievement and attainment could vary according to the grade level at which the student is retained. If retention in early grades is more beneficial to students than later retention, test-based promotion policies targeting early grades could benefit students who would eventually be retained by ensuring that they are retained at a younger age.

Figure 10 depicts the reduced form relationship between third grade reading test scores and retention probabilities in each of the next six years after their initial third grade year. The figure indicates that students below the promotion cutoff are substantially less likely to be retained each year from two to five years after potential third grade retention.

Table 6 shows the corresponding estimates of the effect of third grade retention on future retention probabilities for the full sample.[15] The IV estimates confirm that third

[15]Appendix Table A-2 provides estimates of the impact of third grade retention on future retention probabilities by cohort.

grade retention reduces the probability that the student will be retained two years later by 11 percentage points. The effect is smaller in subsequent years, but remains statistically significant and ranges from 3 to 4 percentage points in magnitude in years three to five. The bottom panel of Table 6 makes grade level the outcome variable in Equation (2), thereby providing direct evidence on the differences in the grade progression of retained and promoted students. The IV estimates show that six years after being retained in third grade, students impacted by Florida's test-based promotion policy are only 0.74 grade levels behind comparable peers who were promoted.

The evidence in Table 6 confirms that third grade retention substantially reduced the probability that Florida students at the promotion cutoff would be retained in future grades. Could these differences in the subsequent grade progression of retained and promoted students explain the fade out of test score impacts evident in Table 5? To evaluate this possibility, we assume that (1) the effects of retention on student achievement after one year are in fact fully persistent and (2) that students retained in subsequent grades experience the same short-term benefits, regardless of the grade in which they were retained. We then ask how much of the observed fade out in test score impacts from year one to year two would be explained by the additional gains made by students retained in year two. The results suggest that differences in subsequent retention could account for no more than 35 percent of the observed fade out in reading effects after two years and 22 percent of the fade out in math effects.[16] Additional analyses also confirm that the test score impacts in both subjects fade out event when students who were subsequently retained are excluded from the sample.

Table 7 reports estimates of the effect of third grade retention on student absences

[16]For example, the simple calculation in terms of reading is as follows: True fade out in reading effects between year one and two is given by 225.8 - 154.6 = 71.2 DSS points (see column 4 of Table 5). Fade out resulting from a 11 percentage point reduction in the probability of being retained after two years (see column 4 of Table 6) is given by 0.11* 225.8 = 24.6. Thus, roughly 35 percent of the fade out in reading effects after two years could be explained by effects on future grade retention.

and special education placement. The results generally confirm that retention had no impact on these outcomes for students with third grade reading scores at the promotion cutoff. The lone exception is absences after three years, when the modest improvement in attendance for retained students likely reflects the fact that most of them had not yet made the transition to middle school.[17] Again in contrast to our preferred IV results, the OLS estimates with controls would suggest statistically significant increases in absences in four of six years and increased rates of special education classification in three years.

5.3 Robustness Analysis and Subgroup Results

Table 8 presents the results of alternative specifications of our analysis of the effects of third grade retention on student achievement and the probability of future retention. To consolidate presentation, we combine the data on each outcome across multiple years into two models intended to summarize short-term (after 1-3 years) and longer-term (after 4-6 years) impacts. The achievement results are based on the rescaled DSS scores used in Table 5. The table's first row presents the results from our preferred specification in this summary format; we then examine whether plausible modifications to that specification influence these results.

The next four rows confirm that our preferred results are robust to the use of alternatives to the ten test-score-point bandwidth ranging from five to 25 points on either side of the cutoff.[18] Achievement impacts in both subjects are consistently more positive using wider bandwidths, but the differences are modest in size. No consistent pattern with respect to bandwidth choice is evident in the results for future retention. We next show that the results are not influenced by the exclusion of students at or within one test score point of the

[17]Schwerdt and West (2011) show that the modal Florida student enters middle school in grade six and experiences an increase in absences of roughly one day (relative to students attending K-8 schools) upon making this transition.

[18]These alternatives more than encompass the informal sensitivity test suggested by Nichols (2007) of using twice and half the preferred bandwidth.

promotion cutoff, as could be the case if there were sorting on unobserved characteristics. Finally, the table's last row shows that the results are also essentially unchanged by the use of quadratic terms in modeling the relationship between third grade reading scores and the probability of retention on either side of the cutoff.

Our analysis thus far has focused on the local average treatment effect of third grade retention for all students performing at the promotion cutoff. This approach could conceal important heterogeneities in local treatment effects across subgroups. It also raises the question of whether similar patterns would hold for higher-achieving students were they to be retained. Table 9, which presents results for several key subgroups in the same format as Table 8, provides little evidence of systematic heterogeneity across subgroups based on gender, ethnicity, or age. The one exception is that the short-term and longer-term achievement effects of retention appear to be modestly less positive for black students than for whites or Hispanics, a pattern which may warrant attention in future research on the long-run outcomes of students retained in Florida.

The remaining rows in Table 9 examine whether our estimates of retention effects are local to students at a specific achievement level, exploiting the fact that there is considerable variation in the math achievement of Florida students who are retained on the basis of their reading test scores. Among students in our preferred bandwidth, 20,537 (27 percent) were classified as performing at level one (of five) based on the third grade math test, 26,357 (35 percent) performed at level two, and 29,253 (29 percent) performed at level three or higher. The first-stage results in column (1) show that the increase in the probability of retention at the promotion cutoff was more than twice as large for students performing at level one in math as for students performing at level three or above, suggesting that students' performance in multiple subjects informed whether they were granted an exemption from the retention requirement. Despite this difference, however, the estimated effects of grade retention on achievement in both reading and math are quite similar across all three groups,

providing at least suggestive evidence that the short-term benefits of retention are not limited to students achieving at a specific level.

6 Conclusion

Our analysis exploits a discontinuity in the probability of grade retention under Florida's test-based promotion policy to study the effects of third grade retention on student outcomes up to six years later. Based on same-age comparisons, we find evidence of substantial short-term gains in both math and reading achievement. However, these positive effects fade out over time and become statistically insignificant within five years. We also find that third grade retention substantially reduces the probability of being retained in later grades but has no clear impact on student absences or special education placement rates.

In sum, our analysis provides more favorable evidence on the effects of early grade retention than found in many previous studies - in particular those which do not rely on credible quasi-experimental methods to address unobserved selection into the retention treatment. We show that early grade retention has substantial positive effects on reading and math achievement in the short run, has no detrimental effects on the limited set of outcomes we can measure, and generates educational and opportunity costs well below a full year when subsequent grade progression is taken into account. To the extent that early grade retention is more beneficial than later grade retention (as suggested by the results of Jacob and Lefgren, 2004, 2009), students who were retained in third grade and would have been retained later clearly benefited from the introduction of the Florida policy. However, we also do not provide definitive evidence that early grade retention is beneficial for students in the long run.

The fade out of test score impacts is a common pattern in the literature on educational interventions, including those which have been shown to generate lasting impacts on adult outcomes. For example, Chetty et al. (2011) show that kindergarten classroom quality

improves college enrollment and adult earnings despite the complete fade out of short-term test score gains. The same appears to be true of early childhood interventions such as the Perry and Abecederian preschool demonstration projects and the Head Start program (see Almond and Currie [2011] for a review). Whether students retained in Florida will also experience long-run benefits remains uncertain. However, the null effects observed six years after third grade retention imply that retained students are performing at the same level as their promoted peers despite the fact that the latter are closer to expected graduation. To the extent that additional time in school (conditional on achievement) increases, for example, the probability of graduation or post-secondary enrollment, early grade retention could generate benefits that outweigh the opportunity costs. An analysis of the effects of third grade retention on educational attainment should be feasible in Florida within a few years.

The Florida policy we have exploited in this paper to study the effects of early grade retention has recently emerged as a model for policymakers in other states. Arizona, Indiana, Oklahoma, and Ohio enacted test-based promotion policies modeled on Florida's between 2010 and 2012, and similar bills have been introduced in the legislatures of several other states. In light of this current interest, we should emphasize that the effects on retained students are only one component of a comprehensive analysis of these policies' merits. Test-based promotion policies also aim to provide incentives for educators and parents to improve the skills of low-performing students prior to third grade. There are also a variety of potential mechanisms, such as the creation of grade cohorts that are more homogenous in ability, that could influence outcomes of higher-performing students. With few exceptions (e.g., Babcock and Bedard, 2011), the broader consequences of policies influencing retention rates have received little attention and deserve further scrutiny.[19]

[19]Using within-state variation in primary school retention rates from 1960 to 1980, Babcock and Bedard (2011) show that a one standard deviation increase in retention rates is associated with a 0.7 percent increase in mean earnings for adult males.

References

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

Almond, D. and Currie, J. (2011). Human capital development before age five. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 4b, pages 1315–1486. Elsevier.

Babcock, P. and Bedard, K. (2011). The wages of failure: New evidence on school retention and long-run outcomes. *Education Finance and Policy*, 6(3):293–322.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., and Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from project star. *The Quarterly Journal of Economics*, 126(4):1593–1660.

Eide, E. R. and Showalter, M. H. (2001). The effect of grade retention on educational and labor market outcomes. *Economics of Education Review*, 20(6):563–576.

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

Hoffman, R., Wise, L. L., and Thacker, A. A. (2001). Florida comprehensive assessment test: Technical report on vertical scaling for reading and mathematics. Technical report, San Antonio, TX: Harcourt Educational Measurement.

Holmes, C. T. (1989). Grade level retention effects: A meta-analysis of research studies. In Shepard, L. A. and Smith, M. L., editors, *Flunking Grades: Research and Policies on Retention*, pages 16–33. New York: The Falmer Press.

Imbens, G. and Kalyanaraman, K. (2009). Optimal bandwidth choice for the regression discontinuity estimator. NBER Working Papers 14726.

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

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

Jimerson, S. R. (1999). On the failure of failure: Examining the association between early grade retention and education and employment outcomes during late adolescence. *Journal of School Psychology*, 37(3):243–272.

Jimerson, S. R., Anderson, G. E., and Whipple, A. D. (2002). Winning the battle and losing the war: Examining the relation between grade retention and dropping out of high school. *Psychology in the Schools*, 39(4):441–457.

Jimerson, S. R., Egeland, B., Sroufe, L. A., and Carlson, B. (2000). A prospective longitudinal study of high school dropouts examining multiple predictors across development. *Journal of School Psychology*, 38(6):525–549.

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355.

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

McCoy, A. R. and Reynolds, A. J. (1999). Grade retention and school performance: An extended investigation. *Journal of School Psychology*, 37(3):273–298.

Nichols, A. (2007). Causal inference with observational data. *Stata Journal*, 7(4):507–541.

Office of Program Policy Analysis & Government Accountability (2006). Third grade retention leading to better student performance statewide. OPPAGA Report 06-66, http://www.oppaga.state.fl.us/reports/pdf/0666rpt.pdf.

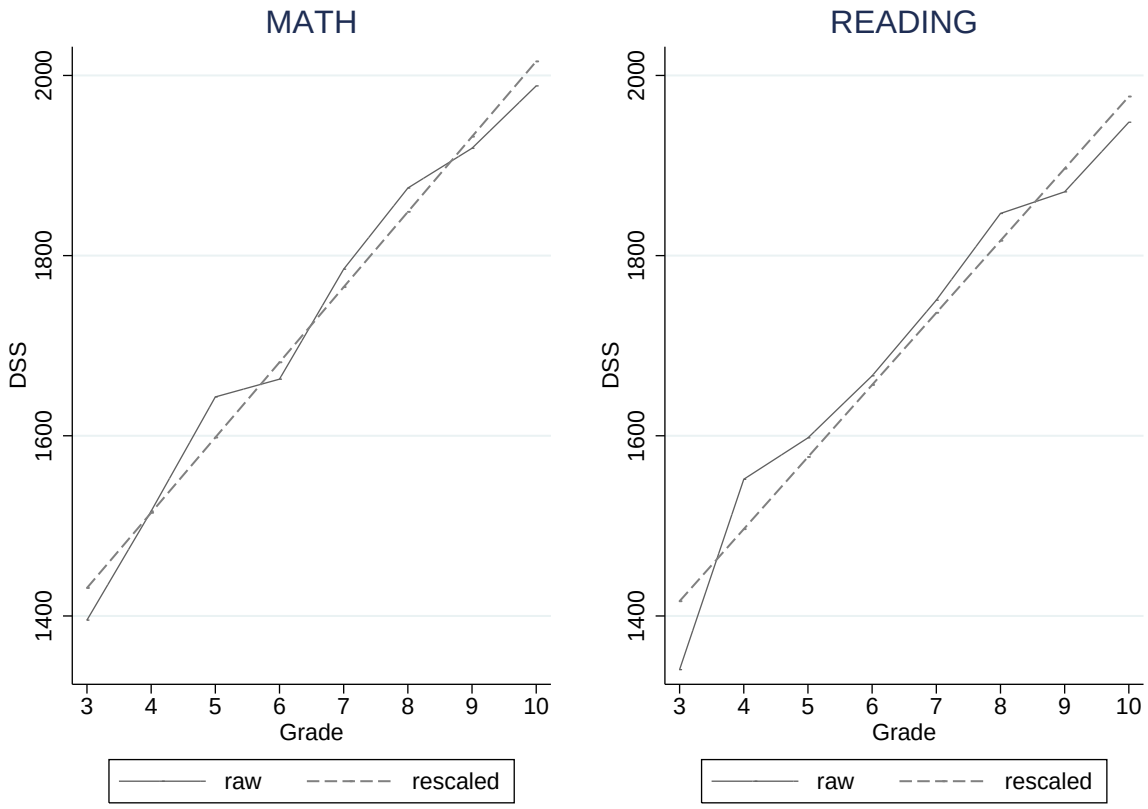
Planty, M., Hussar, W., Snyder, T., Kena, G., Ramani, A. K., Kemp, J., Bianco, K., and Dinkes, R. (2009). *The Condition of Education 2009*. (NCES 2009-081). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Schwerdt, G. and West, M. R. (2011). The impact of alternative grade configurations on student outcomes through middle and high school. PEPG Working Paper 11-02.

Urquiola, M. and Verhoogen, E. (2009). Class-size caps, sorting, and the regression-discontinuity design. *American Economic Review*, 99(1):179–215.

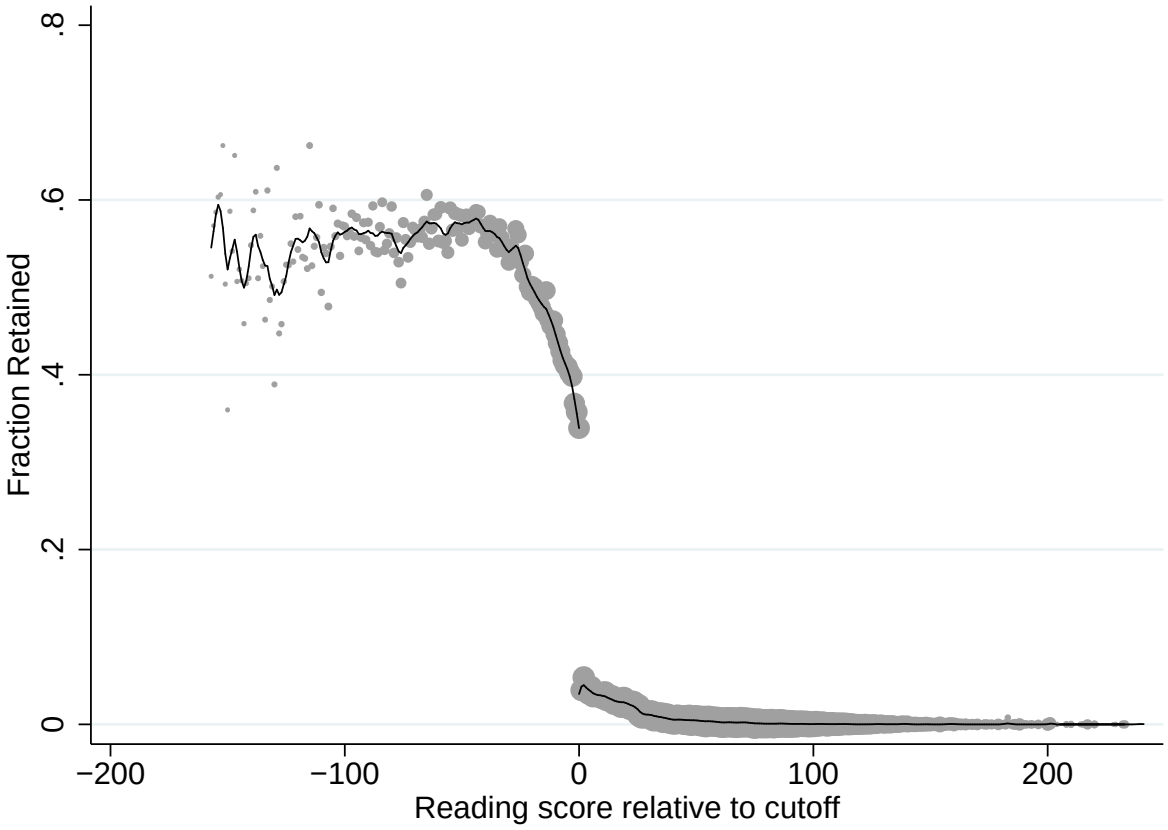
Winters, M. A. and Greene, J. P. (2012). The medium-run effects of Floridas test-based promotion policy. *Education Finance and Policy*, 7(3):305–330.

Figure 1: Average Developmental Scale Scores by Subject and Grade



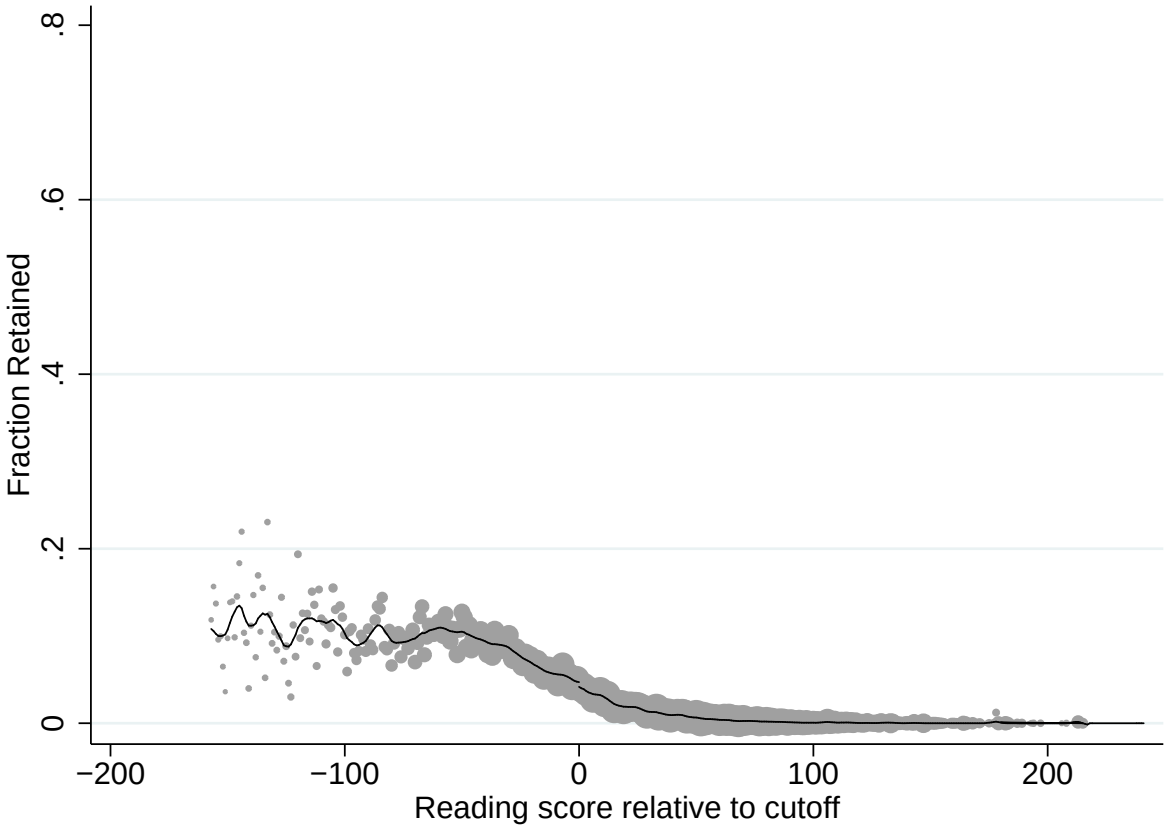
Note: Based on all students in grades 3 to 10 between 2002 and 2009. Rescaled scores stem from predicted values of a linear regression of developmental scale scores on grade levels.

Figure 2: The Relationship between Grade 3 Reading Scores and the Probability of Grade 3 Retention, 2003-2008



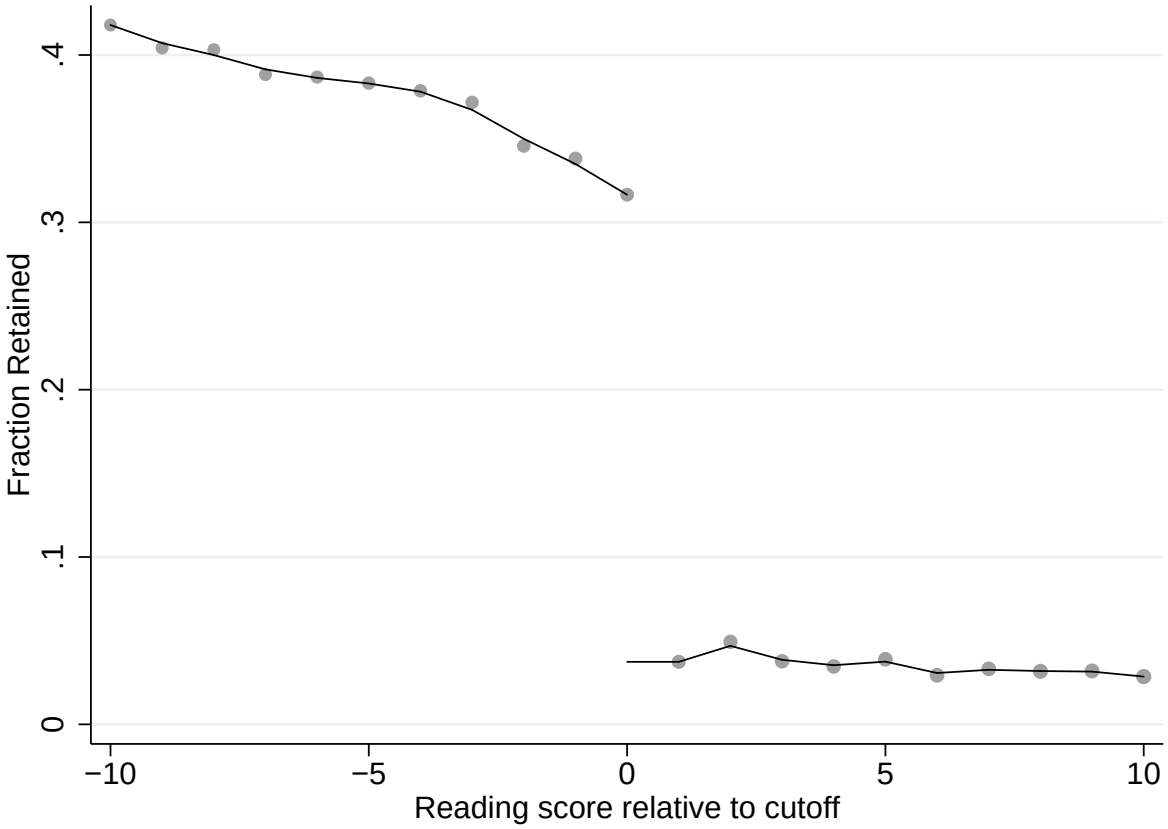
Note: Based on 2003-2008 cohorts. Full sample. Solid line represents predicted values from local linear regressions on both sides of the cutoff. Marker size represents relative group size.

Figure 3: The Relationship between Grade 3 Reading Scores and the Probability of Grade 3 Retention, 2001-2002



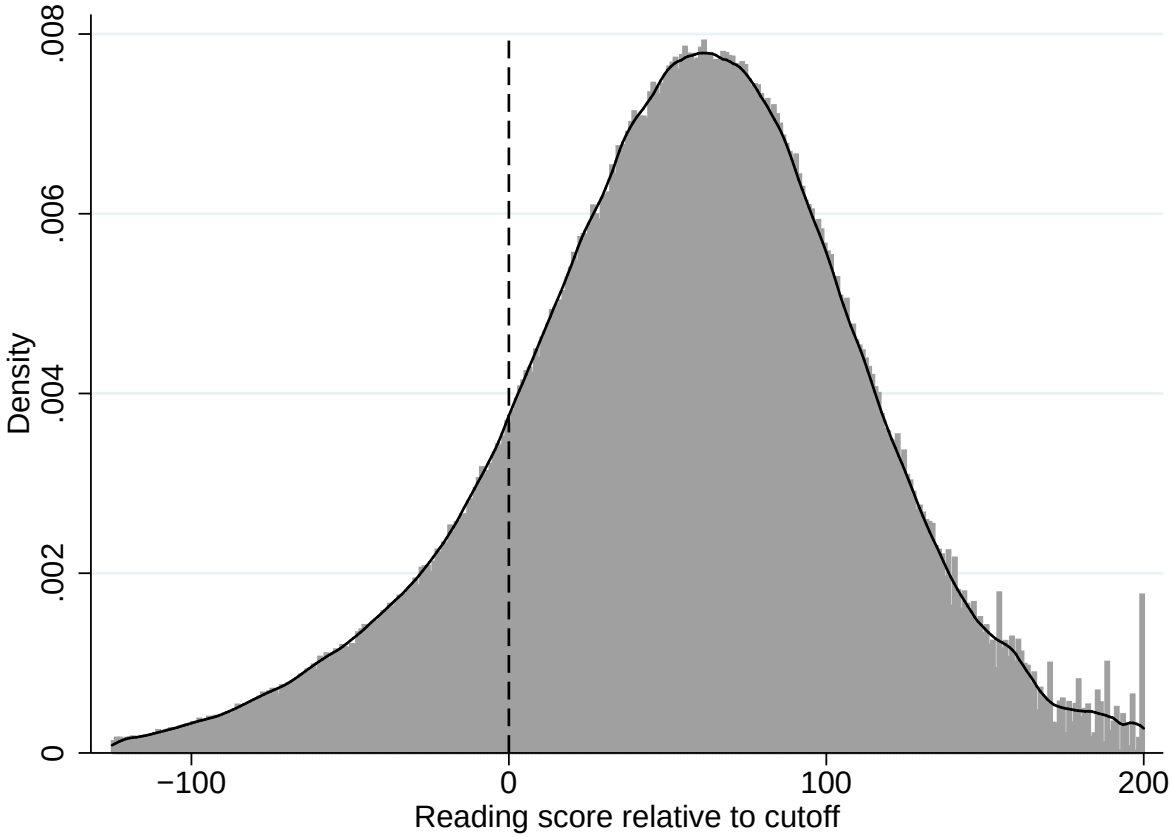
Note: Based on 2001-2002 cohorts. Full sample. Solid line represents predicted values from local linear regressions on both sides of the cutoff. Marker size represents relative group size.

Figure 4: The Relationship between Reading Scores and Grade Retention around the Cutoff



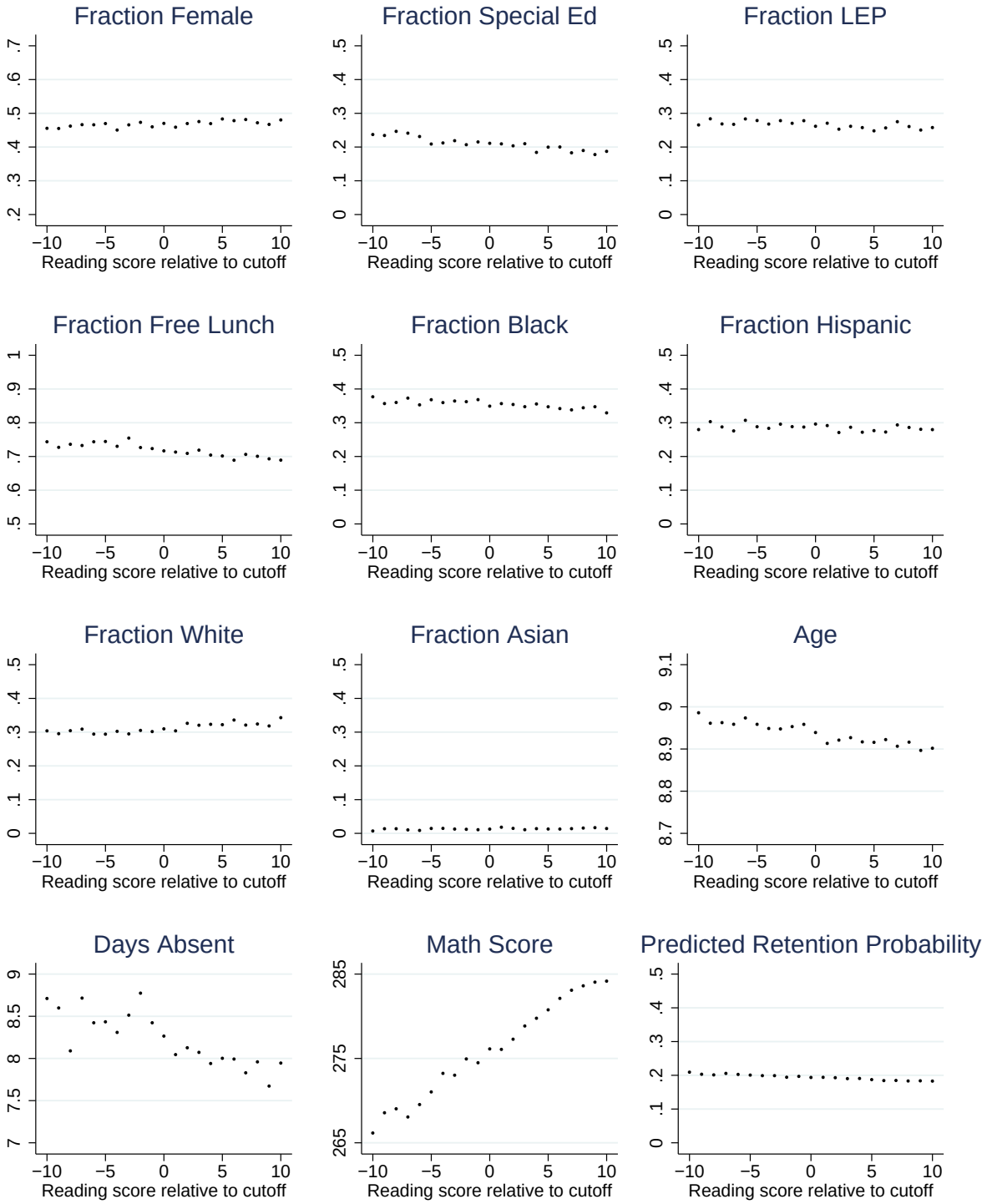
Note: Based on 2003-2008 cohorts. Discontinuity sample with 10-point bandwidth. Solid line represents predicted values from local linear regressions on both sides of the cutoff. Marker size represents relative group size.

Figure 5: Distribution of Reading Scores in Grade 3



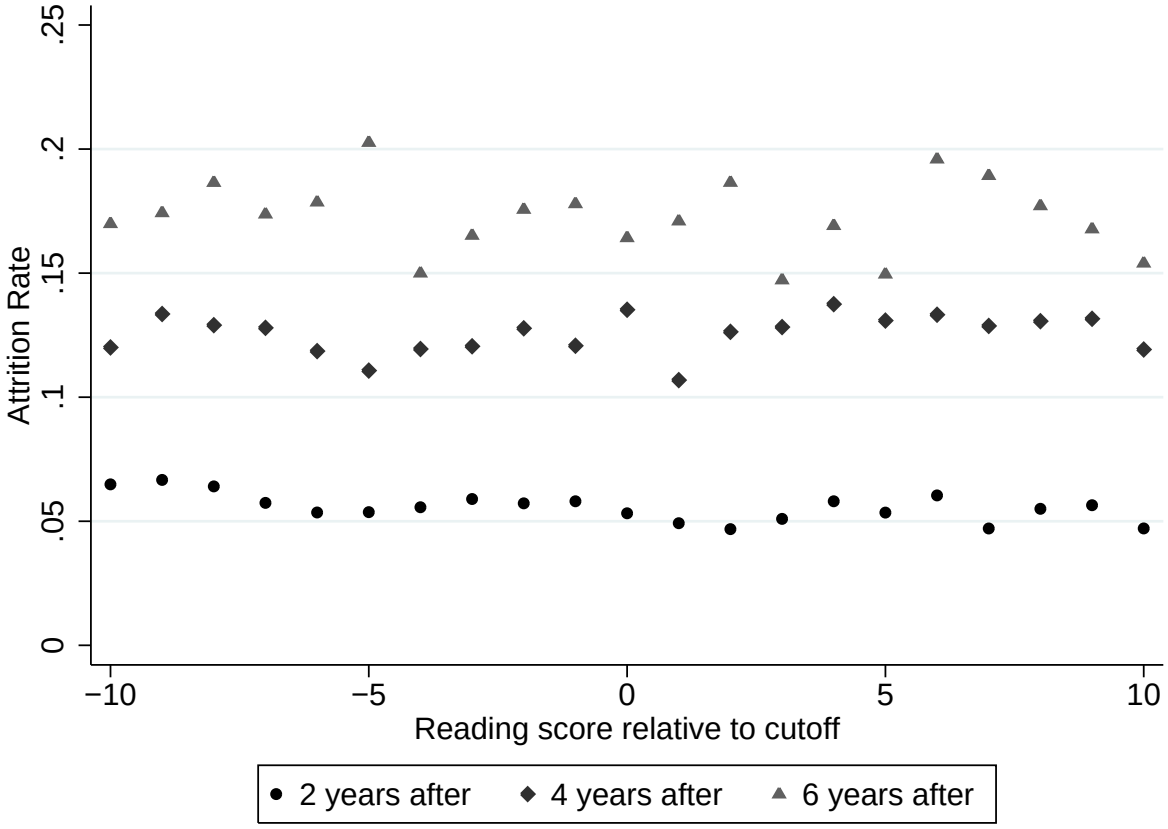
Note: Based on 2003-2008 cohorts. Full sample. Solid line represents kernel density estimates.

Figure 6: The Relationship between Reading Scores in Grade 3 and Student Characteristics



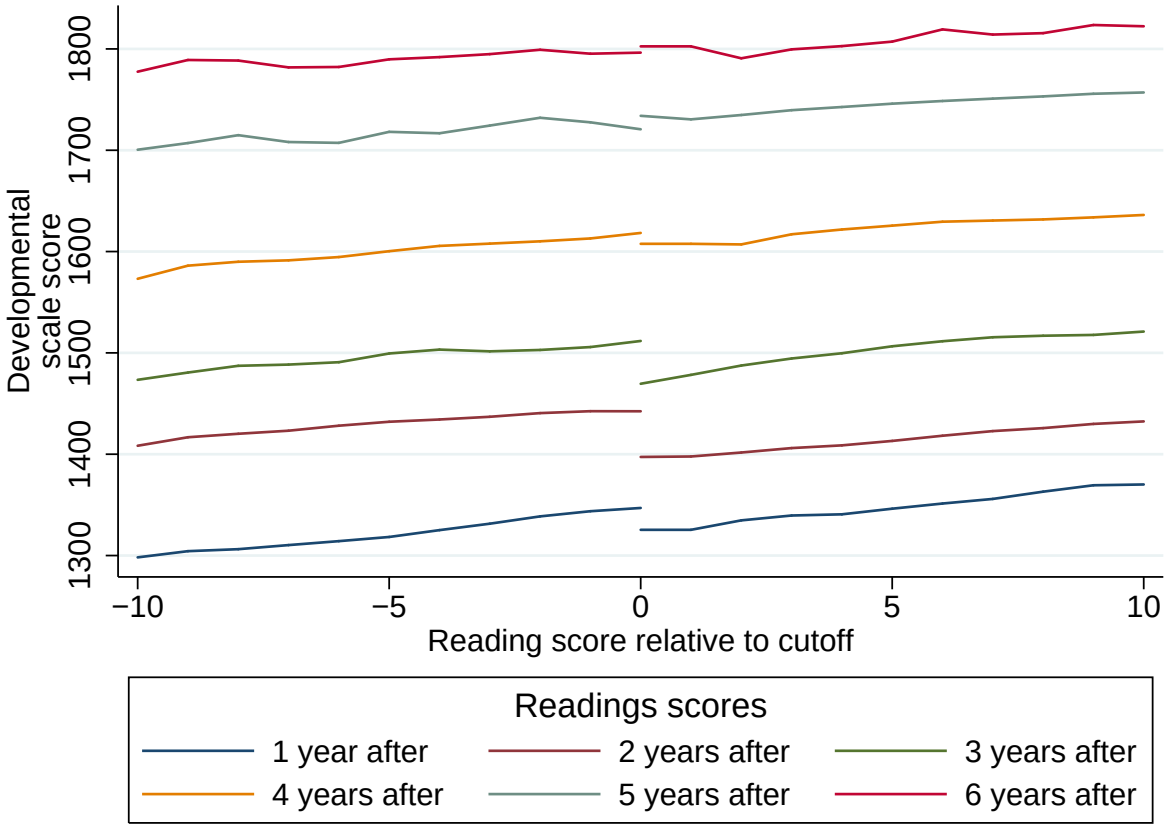
Note: Based on 2003-2008 cohorts. Discontinuity sample with 10-point bandwidth. Predicted retention probability displays predicted values after estimating a probit model that includes all student background variables except for reading scores as explanatory variables.

Figure 7: The Relationship between Reading Scores in Grade 3 and Subsequent Attrition from the Data around the Cutoff



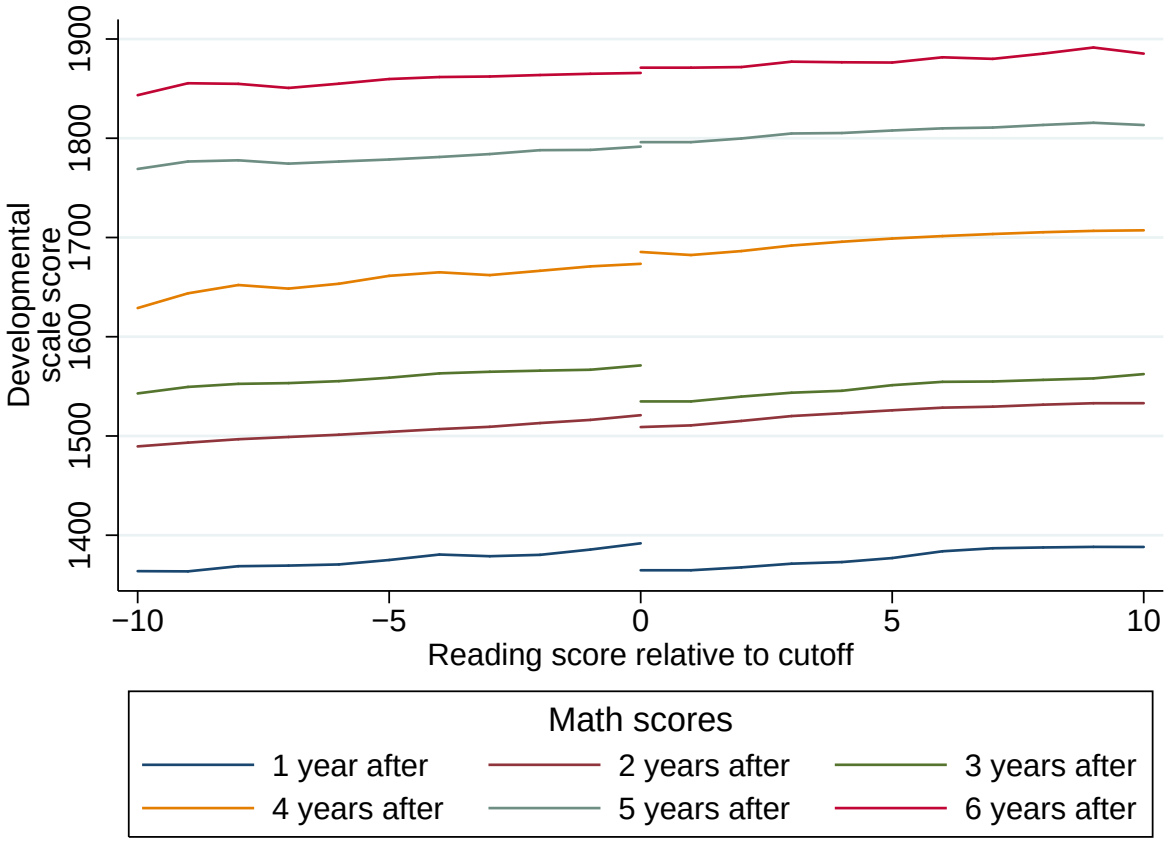
Note: Based on cohorts 2003-2008. Discontinuity sample with 10-point bandwidth.

Figure 8: The Relationship between Reading Scores in Grade 3 and Future Reading Achievement around the Cutoff



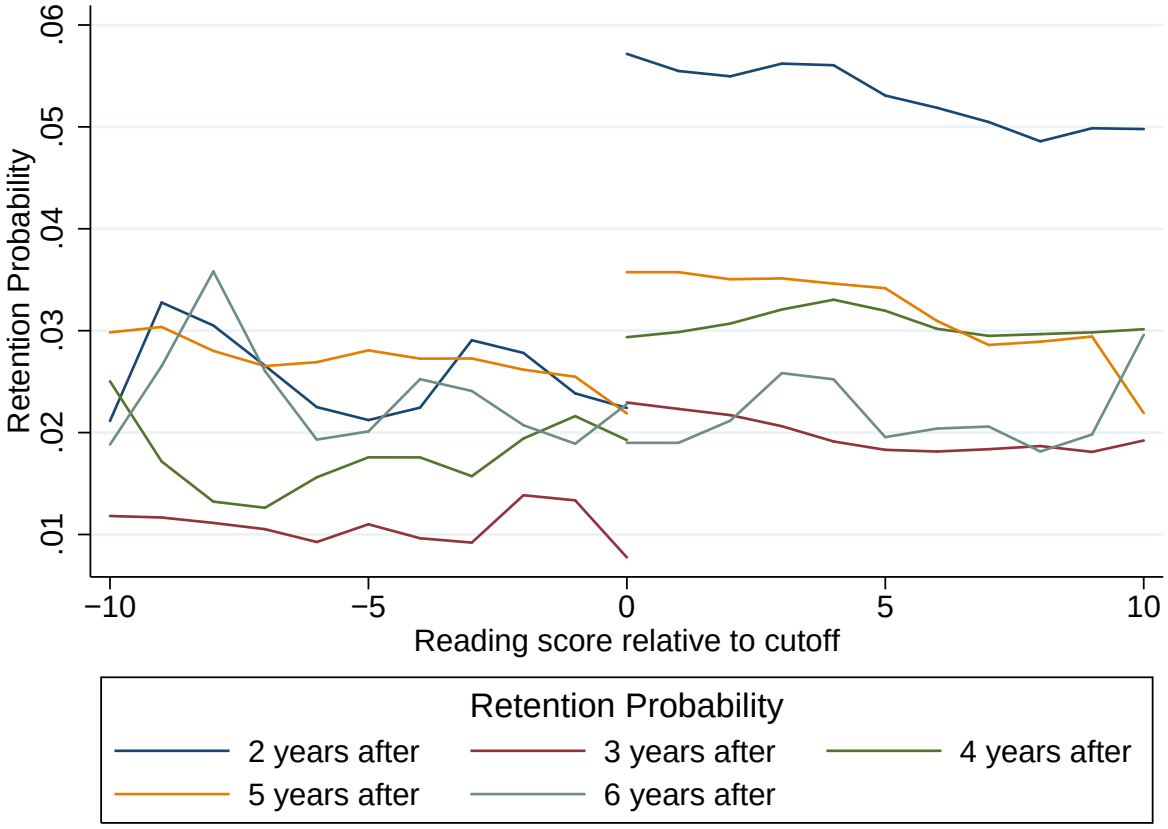
Note: Based on 2003-2008 cohorts. Discontinuity sample with 10-point bandwidth. Solid line represents predicted values from local linear regressions on both sides of the cutoff.

Figure 9: The Relationship between Reading Scores in Grade 3 and Future Math Achievement around the Cutoff



Note: Based on 2003-2008 cohorts. Discontinuity sample with 10-point bandwidth. Solid line represents predicted values from local linear regressions on both sides of the cutoff.

Figure 10: The Relationship between Reading Scores in Grade 3 and Future Grade Retention around the Cutoff



Note: Based on 2003-2008 cohorts. Discontinuity sample with 10-point bandwidth. Solid line represents predicted values from local linear regressions on both sides of the cutoff.

Table 1: Observations by Year and Grade

	Years after potential treatment (= retention in grade 3)					
	1	2	3	4	5	6
Grade 3						
T=1	**0.08**	0.00	-	-	-	-
T=0	0.00	0.00	-	-	-	-
Grade 4						
T=1	0.00	**0.09**	0.00	-	-	-
T=0	**0.92**	0.01	0.00	-	-	-
Grade 5						
T=1	0.00	0.00	**0.09**	0.00	-	-
T=0	0.00	**0.90**	0.02	0.00	-	-
Grade 6						
T=1	-	0.00	0.00	**0.10**	0.01	-
T=0	-	0.00	**0.88**	0.04	0.00	-
Grade 7						
T=1	-	-	0.00	0.00	**0.10**	0.01
T=0	-	-	0.00	**0.86**	0.05	0.00
Grade 8						
T=1	-	-	-	0.00	0.01	**0.11**
T=0	-	-	-	0.00	**0.83**	0.05
Grade 9						
T=1	-	-	-	-	0.00	0.01
T=0	-	-	-	-	0.00	**0.81**
Cohorts	2003-2008	2003-2007	2003-2006	2003-2005	2003-2004	2003
Students	983,308	768,593	578,387	418,680	275,194	134,284

Table 2: Summary Statistics

	Total	Failed Promotion Cuttoff	Retained	Retained, but above Cuttoff
FCAT Math	0.06	-1.13	-1.22	-0.83
FCAT Reading	0.07	-1.46	-1.43	-0.38
Female	0.49	0.42	0.42	0.46
Age	8.84	9.06	8.89	8.77
White	0.48	0.28	0.28	0.50
Black	0.22	0.38	0.40	0.29
Hispanic	0.24	0.31	0.29	0.15
Asian	0.02	0.01	0.01	0.01
Other	0.04	0.03	0.03	0.04
Free or reduced lunch	0.52	0.78	0.79	0.65
Limited English proficiency	0.19	0.30	0.29	0.11
Special Education	0.16	0.37	0.29	0.15
Days absent	7.46	9.10	9.28	10.13
Number of students	983,308	159,866	81,357	4,959

Note: Based on 2003-2008 cohorts. Full sample. Test scores in math and reading are standardized by subject, year, and grade to have a mean of zero and a standard deviation of one.

Table 3: Effect of Reading Performance on the Probability of Grade Retention in Grade 3

	Policy off		Policy on						
Cohorts	2001	2002	2003	2004	2005	2006	2007	2008	2003-2008
Below cutoff	0.006	0.020***	0.396***	0.298***	0.317***	0.372***	0.222***	0.234***	0.310***
	[0.006]	[0.007]	[0.013]	[0.015]	[0.014]	[0.017]	[0.013]	[0.014]	[0.006]
Below cutoff	–0.002	–0.004	–0.017	–0.048***	0.014	0.027	–0.029**	–0.002	–0.014**
× LEP	[0.007]	[0.007]	[0.014]	[0.015]	[0.015]	[0.017]	[0.013]	[0.014]	[0.006]
Below cutoff	–0.000	–0.003	–0.104***	–0.086***	–0.116***	–0.166***	–0.083***	–0.075***	–0.109***
× Special Education	[0.007]	[0.008]	[0.016]	[0.016]	[0.015]	[0.016]	[0.014]	[0.014]	[0.006]
Reading	–0.000	–0.001	–0.001	–0.000	–0.001	–0.001	–0.001	0.001	–0.000
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.000]
Reading	0.000	0.000	–0.010***	–0.009***	–0.007***	–0.006**	–0.007***	–0.008***	–0.008***
× Below cutoff	[0.001]	[0.001]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.001]
Additional covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	No	No	No	No	No	No	No	Yes
Students	17,676	16,516	15,687	12,040	12,435	9,981	12,995	11,536	74,674
R^2	0.018	0.020	0.299	0.209	0.233	0.261	0.161	0.171	0.230
F-statistic on instruments	0.33	2.71	314.47	142.74	184.60	187.64	106.80	104.94	1,014.72
Pr > F	0.80	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00

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

Note: Based on discontinuity sample with 10-point bandwidth. Dependent variable is a dummy indicating retention in grade 3 in all columns. Additional covariates (not reported) include special education status in grade 3, LEP status in grade 3, math scores, gender, age, race, and free or reduced-price lunch status in grade 3. Robust standard errors in parentheses.

Table 4: Effect of Grade Retention on Student Achievement

	Specification			
	OLS		IV	
Dependent Variable	(1)	(2)	(3)	(4)
Reading (SD= 370)				
1 year (n = 74,443)	–60.68***	–41.19***	92.58***	94.58***
	(2.064)	(2.058)	(10.409)	(9.941)
2 years (n = 59,554)	58.18***	76.43***	183.6***	184.5***
	(2.287)	(2.263)	(11.653)	(11.179)
3 years (n = 45,175)	–4.555*	14.09***	98.24***	100.1***
	(2.691)	(2.650)	(12.462)	(11.989)
4 years (n = 35,001)	–53.18***	–35.85***	46.95***	48.29***
	(2.970)	(2.934)	(13.534)	(12.974)
5 years (n = 23,568)	–70.45***	–55.23***	–9.989	–6.667
	(3.180)	(3.135)	(13.842)	(13.201)
6 years (n = 12,912)	–30.21***	–14.74***	15.03	15.39
	(3.852)	(3.779)	(15.759)	(14.916)
Math (SD= 306)				
1 year (n = 74,327)	–1.454	47.85***	94.35***	95.43***
	(2.097)	(1.729)	(10.213)	(8.028)
2 years (n = 59,354)	–58.12***	–15.26***	29.00***	28.64***
	(2.100)	(1.789)	(10.064)	(8.048)
3 years (n = 45,093)	31.78***	73.82***	109.7***	111.5***
	(2.473)	(2.155)	(11.808)	(9.992)
4 years (n = 34,987)	–116.0***	–76.95***	–17.52	–19.26*
	(2.868)	(2.561)	(12.845)	(10.924)
5 years (n = 23,563)	–77.68***	–48.60***	–27.04**	–25.08**
	(2.800)	(2.473)	(12.148)	(10.344)
6 years (n = 12,905)	–57.20***	–31.37***	–4.812	–8.717
	(3.156)	(2.796)	(12.617)	(10.803)
Performance and demographic covariates	No	Yes	No	Yes

Note: Based on discontinuity sample with 10-point bandwidth. Dependent variables are developmental scale scores in reading and math; reported standard deviations are for grade 3. All estimations control for special education status in grade 3, LEP status in grade 3, a linear function in grade 3 reading scores that allows for different trends at both sides of the cuttoff, and cohort dummies. Performance and demographic covariates include math scores, gender, age, race, and free or reduced-price lunch status in grade 3. Robust standard errors in parentheses.

Table 5: Effect of Grade Retention on Student Achievement (rescaled)

	Specification			
	OLS		IV	
Dependent Variable	(1)	(2)	(3)	(4)
Reading (SD= 370)				
1 year (n = 74,443)	70.58***	90.07***	223.8***	225.8***
	(2.064)	(2.058)	(10.409)	(9.941)
2 years (n = 59,554)	26.13***	44.76***	153.7***	154.6***
	(2.286)	(2.261)	(11.660)	(11.175)
3 years (n = 45,175)	–14.99***	3.861	88.82***	90.71***
	(2.691)	(2.651)	(12.475)	(11.995)
4 years (n = 35,001)	–49.81***	–32.57***	49.82***	51.19***
	(2.970)	(2.934)	(13.522)	(12.963)
5 years (n = 23,568)	–57.34***	–42.45***	.764	4.004
	(3.170)	(3.125)	(13.743)	(13.123)
6 years (n = 12,912)	–73.64***	–56.91***	–21.77	–21.80
	(3.859)	(3.798)	(15.998)	(15.104)
Math (SD= 306)				
1 year (n = 74,327)	36.00***	85.30***	131.8***	132.9***
	(2.097)	(1.729)	(10.213)	(8.028)
2 years (n = 59,354)	–17.50***	24.89***	67.11***	66.77***
	(2.097)	(1.788)	(10.010)	(8.025)
3 years (n = 45,093)	–25.57***	17.44***	58.04***	59.84***
	(2.465)	(2.143)	(11.848)	(9.972)
4 years (n = 34,987)	–83.37***	–45.10***	10.81	9.156
	(2.854)	(2.542)	(12.647)	(10.757)
5 years (n = 23,563)	–71.11***	–42.27***	–21.46*	–19.60*
	(2.772)	(2.447)	(12.037)	(10.246)
6 years (n = 12,905)	–87.89***	–61.14***	–30.88**	–35.09***
	(3.185)	(2.836)	(12.931)	(11.080)
Performance and demographic covariates	No	Yes	No	Yes

Note: Based on discontinuity sample with 10-point bandwidth. Dependent variables are developmental scale scores in reading and math; reported standard deviations are for grade 3. All estimations control for special education status in grade 3, LEP status in grade 3, a linear function in grade 3 reading scores that allows for different trends at both sides of the cutoff, and cohort dummies. Performance and demographic covariates include math scores, gender, age, race, and free or reduced-price lunch status in grade 3. Robust standard errors in parentheses.

Table 6: Effect of Grade Retention in Grade 3 on Future Grade Retention and Grade Level

	Specification			
	OLS		IV	
Dependent Variable	(1)	(2)	(3)	(4)
Retention Probability				
2 years (n = 59,679)	–.0506***	–.0614***	–.110***	–.109***
	(.001)	(.002)	(.010)	(.010)
3 years (n = 44,271)	–.00833***	–.0124***	–.0295***	–.0295***
	(.001)	(.001)	(.007)	(.007)
4 years (n = 33,946)	–.0240***	–.0290***	–.0416***	–.0423***
	(.002)	(.002)	(.011)	(.011)
5 years (n = 22,746)	–.00226	–.00701**	–.0404***	–.0426***
	(.003)	(.003)	(.014)	(.014)
6 years (n = 12,384)	.00821**	.00525	–.00162	–.00205
	(.004)	(.004)	(.014)	(.014)
Grade Level				
2 years (n = 59,679)	–.944***	–.932***	–.878***	–.879***
	(.002)	(.002)	(.011)	(.011)
3 years (n = 44,271)	–.920***	–.902***	–.828***	–.829***
	(.003)	(.003)	(.014)	(.014)
4 years (n = 33,946)	–.885***	–.862***	–.755***	–.758***
	(.004)	(.004)	(.020)	(.020)
5 years (n = 22,746)	–.863***	–.835***	–.679***	–.685***
	(.006)	(.006)	(.030)	(.029)
6 years (n = 12,384)	–.857***	–.826***	–.734***	–.746***
	(.009)	(.009)	(.037)	(.036)
Performance and demographic covariates	No	Yes	No	Yes

Note: Based on discontinuity sample with 10-point bandwidth. Dependent variable is a dummy indicating grade retention in the top panel and the student's grade level in the bottom panel. All estimations control for special education status in grade 3, LEP status in grade 3, a linear function in grade 3 reading scores that allows for different trends at both sides of the cutoff, and cohort dummies. Performance and demographic covariates include math scores, gender, age, race, and free or reduced-price lunch status in grade 3. Robust standard errors in parentheses.

Table 7: Effect of Grade Retention on Student Absence and Special Education Placement

	Specification			
	OLS		IV	
Dependent Variable	(1)	(2)	(3)	(4)
Days absent				
1 year (n = 74,599)	.499***	.304***	–.309	–.374
	(.081)	(.081)	(.391)	(.384)
2 years (n = 59,597)	.499***	.326***	–.0630	–.147
	(.093)	(.092)	(.435)	(.426)
3 years (n = 45,267)	–.333***	–.485***	–1.267**	–1.422***
	(.110)	(.110)	(.497)	(.487)
4 years (n = 35,101)	.268*	.0313	–.654	–.785
	(.149)	(.148)	(.689)	(.673)
5 years (n = 23,659)	1.011***	.831***	1.331	.906
	(.207)	(.207)	(.939)	(.917)
6 years (n = 12,985)	1.735***	1.406***	–.608	–.867
	(.303)	(.302)	(1.167)	(1.140)
Special Ed Placement				
1 year (n = 74,674)	.0129***	.0122***	.0174	.0168
	(.003)	(.003)	(.012)	(.012)
2 years (n = 59,684)	.0197***	.0172***	.0179	.0169
	(.003)	(.004)	(.015)	(.015)
3 years (n = 45,299)	.0152***	.0114***	.00911	.00754
	(.004)	(.004)	(.017)	(.017)
4 years (n = 35,126)	.00608	.00139	.0119	.01000
	(.005)	(.005)	(.020)	(.020)
5 years (n = 23,681)	–.000179	–.00649	.0140	.00620
	(.006)	(.006)	(.024)	(.024)
6 years (n = 13,000)	.000190	–.00751	.0251	.0134
	(.007)	(.007)	(.028)	(.027)
Performance and demographic covariates	No	Yes	No	Yes

Note: Based on discontinuity sample with 10-point bandwidth. Dependent variable is the number of days absent in the school year in the top panel and a dummy indicating special education placement in the bottom panel. Performance and demographic covariates include math scores, gender, age, race, free or reduced-price lunch status in grade 3. Robust standard errors in parentheses.

Table 8: Robustness Checks

	1st Stage	Reading		Math		Retention	
Years		1-3	4-6	1-3	4-6	2-3	4-6
Robustness Check	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Baseline	.310***	165.2***	16.25*	88.91***	–13.97**	–.0739***	–.0338***
	(.006)	(6.581)	(8.450)	(5.268)	(7.007)	(.007)	(.007)
Bandwidth 25	.314***	187.2***	29.17***	103.2***	–.291	–.0642***	–.0309***
	(.004)	(4.468)	(5.680)	(3.529)	(4.722)	(.004)	(.005)
Bandwidth 20	.311***	183.4***	32.67***	98.69***	–3.568	–.0683***	–.0315***
	(.005)	(4.904)	(6.311)	(3.877)	(5.220)	(.004)	(.005)
Bandwidth 15	.314***	177.6***	24.98***	92.36***	–12.75**	–.0713***	–.0313***
	(.005)	(5.478)	(7.028)	(4.355)	(5.825)	(.005)	(.006)
Bandwidth 5	.302***	160.3***	11.95	88.20***	–20.64**	–.0750***	–.0254**
	(.008)	(9.529)	(12.073)	(7.628)	(9.917)	(.010)	(.010)
w/o cutoff ± 1	.321***	171.1***	11.56	81.14***	–14.18	–.0679***	–.0389***
	(.007)	(8.507)	(11.073)	(6.868)	(9.132)	(.009)	(.010)
Quadratic	.313***	162.0***	15.47*	86.68***	–15.28**	–.0730***	–.0334***
	(.006)	(6.410)	(8.193)	(5.167)	(6.766)	(.007)	(.007)

Note: Top row indicates dependent variable. Second row indicates years after potential grade 3 retention. Column 1 shows first stage estimates. Columns 2-7 report IV estimates with performance and demographic covariates. Estimated effects on achievement are based on rescaled developmental scales scores. Robust standard errors in parentheses.

Table 9: Subgroup Results

	1st Stage	Reading		Math		Retention	
Years		1-3	4-6	1-3	4-6	2-3	4-6
Subgroup	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Baseline	.310***	165.2***	16.25*	88.91***	–13.97**	–.0739***	–.0338***
	(.006)	(6.581)	(8.450)	(5.268)	(7.007)	(.007)	(.007)
Girls	.295***	166.6***	15.27	88.45***	–16.08	–.0807***	–.0281***
	(.008)	(9.382)	(11.987)	(7.838)	(9.860)	(.009)	(.008)
Boys	.325***	163.9***	18.60	88.95***	–13.80	–.0686***	–.0391***
	(.008)	(9.139)	(11.811)	(7.052)	(9.835)	(.009)	(.011)
White	.289***	197.7***	51.78***	104.0***	9.142	–.0845***	–.0465***
	(.011)	(13.394)	(16.827)	(10.168)	(13.395)	(.013)	(.014)
Black	.328***	139.8***	–22.18*	79.82***	–34.72***	–.0831***	–.0258**
	(.009)	(9.969)	(13.065)	(8.420)	(11.575)	(.011)	(.013)
Hispanic	.308***	165.7***	18.25	79.88***	–11.37	–.0427***	–.0288**
	(.013)	(12.145)	(15.320)	(9.583)	(12.075)	(.010)	(.011)
Age $\geq$ 9	.306***	169.0***	21.49**	94.23***	–5.402	–.0631***	–.0382***
	(.007)	(7.718)	(10.020)	(6.153)	(8.355)	(.008)	(.009)
Age $\leq$ 8	.319***	148.2***	5.438	78.33***	–32.53***	–.0984***	–.0234**
	(.012)	(11.304)	(14.116)	(8.837)	(10.779)	(.012)	(.012)
Math Level 1	.423***	144.6***	5.507	80.44***	–29.46**	–.0964***	–.0200*
	(.012)	(9.743)	(12.803)	(8.849)	(11.851)	(.012)	(.012)
Math Level 2	.331***	167.9***	20.05	99.85***	–14.45	–.0711***	–.0426***
	(.010)	(10.306)	(13.114)	(8.019)	(10.700)	(.011)	(.012)
Math Level $\geq$3	.208***	187.7***	14.16	73.50***	.348	–.0495***	–.0377***
	(.009)	(15.894)	(19.699)	(11.327)	(14.469)	(.012)	(.014)

Note: Based on discontinuity sample with 10-point bandwidth. Top row indicates dependent variable. Second row indicates years after potential grade 3 retention. Column 1 shows first stage estimates. Columns 2-7 report IV estimates with performance and demographic covariates. Estimated effects on achievement are based on rescaled developmental scales scores. Robust standard errors in parentheses.

Table A-1: Achievement Results by Cohort

Cohort	2003	2004	2005	2006	2007	2008
Reading						
1 year	58.989***	88.784***	114.329***	111.501***	172.555***	35.159
	(15.411)	(26.192)	(21.839)	(25.299)	(36.475)	(30.796)
2 years	221.823***	164.137***	119.808***	175.696***	206.972***	
	(17.783)	(28.983)	(25.748)	(23.915)	(34.191)	
3 years	81.107***	152.072***	84.745***	99.902***		
	(17.787)	(31.339)	(26.263)	(24.562)		
4 years	35.894**	55.503*	59.001**			
	(17.926)	(29.131)	(23.692)			
5 years	–26.040*	29.369				
	(15.470)	(24.150)				
6 years	15.386					
	(14.916)					
Math						
1 year	68.408***	116.292***	77.516***	112.458***	127.031***	83.242***
	(13.049)	(20.257)	(18.683)	(19.474)	(27.970)	(26.478)
2 years	1.024	21.355	22.700	42.381**	81.110***	
	(12.465)	(20.163)	(17.821)	(18.883)	(25.841)	
3 years	101.271***	108.067***	117.875***	137.378***		
	(15.006)	(27.577)	(20.503)	(20.207)		
4 years	–34.955**	–32.760	13.539			
	(16.416)	(22.038)	(19.578)			
5 years	–30.335**	–15.405				
	(12.270)	(18.708)				
6 years	–8.717					
	(10.803)					
Students	15,687	12,040	12,435	9,981	12,995	11,536

Note: Based on discontinuity sample with 10-point bandwidth. Dependent variables are developmental scale scores in reading and math. The table displays IV estimates with performance and demographic covariates by cohort of students. A cohort is defined by the school year students attended third grade for the first time. The last row indicates the number of students by cohort in the first stage regression for outcomes after 1 year.

Table A-2: Retention Results by Cohort

Cohort	2003	2004	2005	2006	2007
2 years	–.096***	–.185***	–.089***	–.085***	–.099***
	(.017)	(.031)	(.023)	(.023)	(.028)
3 years	–.027**	–.049**	–.031**	–.015	
	(.012)	(.020)	(.014)	(.011)	
4 years	–.049***	–.076***	–.002		
	(.015)	(.024)	(.018)		
5 years	–.038**	–.050**			
	(.016)	(.025)			
6 years	–.002				
	(.014)				
Students	15,687	12,040	12,435	9,981	12,995

Note: Based on discontinuity sample with 10-point bandwidth. Dependent variable is a dummy indicating grade retention. The table displays IV estimates with performance and demographic covariates by cohort of students. A cohort is defined by the school year students attended third grade for the first time. The first row shows first stage estimates. The last row indicates the number of students by cohort in the first stage regression for outcomes after 1 year.