This manuscript was published in the Journal of Research on Educational Effectiveness, Vol. 11, No. 2, 192–216; https://doi.org/10.1080/19345747.2017.1390024.
Published online December 27, 2017.

The Causal Effects of Grade Retention on Behavioral Outcomes

Paco Martorell
UC Davis, School of Education
pmartorell@ucdavis.edu
One Shields Ave.
Davis, CA 95616
(530) 752-3353

Louis T. Mariano
RAND Corporation
loum@rand.org
1200 S. Hayes St.
Arlington, VA 22315
(703) 413-1100

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305E120006 to the RAND Corporation. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Correspondence concerning this manuscript should be addressed to: Louis T. Mariano, RAND Corporation, 1200 S. Hayes St., Arlington, VA 22202. Lou_Mariano@rand.org

The Causal Effects of Grade Retention on Behavioral Outcomes

Abstract

This study examines the impact of grade retention on behavioral outcomes under a comprehensive assessment-based student promotion policy in New York City. To isolate the causal effect of grade retention, we implement a fuzzy regression discontinuity (RD) design that exploits that grade retention is largely determined by whether a student scores below a cutoff on a standardized test score. We use data on students subject to the policy over a nine year span to examine impacts on attendance and disciplinary event outcomes. We do not find evidence of systematic effects of retention on behavioral outcomes in either direction. We do find sporadic non-sustained significant effects of retention on behavioral outcomes. When present, these isolated non-persistent effects tend to be beneficial when found for retained elementary school students and mixed for retained middle school students.

Keywords: Student Retention; Regression Discontinuity; Behavioral Outcomes; Suspensions; Attendance

The Causal Effects of Grade Retention on Behavioral Outcomes

The widespread availability of standardized assessment scores resulting from the adoption of test-based accountability policies, such as the No Child Left Behind Act of 2001, has made it possible to base grade promotion decisions in large part on standardized test performance. As of 2009, Marsh, Gershwin, Kirby, and Xia (2009) reported six states and 13 large school districts having implemented assessment-based student promotion policies. More recently, Workman (2014) reported that 16 states had a standards-based promotion policy in place for third grade reading alone as of December, 2014.

The theory behind test-based promotion policies is that providing an additional year of instruction in the current grade affords struggling students the opportunity to improve their proficiency in the curriculum of the current grade, which in turn enables students to fully engage in the curriculum of subsequent grades. Given that it extends a student's education timeline by a full year, grade retention is an intensive remedial intervention. One estimate suggests the total cost of retaining students in the United States exceeds $12 billion annually (West, 2012). Grade retention is also very controversial for reasons that go beyond its monetary costs. Critics argue that grade retention is punitive in nature, and that being a year behind one's peers may result in disengagement with school (Jackson, 1975; Roderick, 1994).

Such disengagement could manifest itself through an increase in behavioral problems problems (Byrd, Weitzman, & Auinger, 1997). Developmental psychologists have shown that stressful life events are correlated with behavioral problems (Compas, 1987), and have also found that students see grade retention as one of the most stressful life events that can befall them (Anderson, Jimerson, & Whipple, 2005). Indeed, a large literature summarized in the meta-analyses of Holmes (1989) and Jimerson (2001) has documented a strong association between grade retention and worse academic, socioemotional, and behavioral outcomes.[1] Partly in response to these results on behavioral outcomes and related findings from developmental psychology, professional organizations such as the American Federation of Teachers (1997) and the National Council of Teachers of English (2015) have come out against grade retention policies. Such behavioral problems resulting from grade retention would be worrisome given recent evidence linking behavioral problems in school to significantly lower wages and earnings in adulthood (Segal, 2013).

On the other hand, it is also possible that grade retention could reduce behavioral outcomes. For instance, if retained students compare themselves to their younger peers during the repeated year, their academic self-concept may improve.[2] Along the same lines, if grade retention delivers the academic benefits proponents claim that it does, behavioral problems may improve as a result of being held back.

Motivated by this theoretical ambiguity, this study examines whether grade retention affects behavioral problems using administrative data on students subject to a comprehensive assessment-based student promotion policy instituted by New York City Department of Education (NYCDOE).[3] Specifically, we test the hypothesis that grade retention affects two measures of behavioral problems: absenteeism and suspensions. Absenteeism is a useful measure for this study because it is reflective of student engagement in school (Rohrman, 1993).

---

[1] For the prior literature aggregated effect sizes, a negative effect size is an effect in the favor of the promoted students. Jimerson (2001) reports an aggregated negative meta-analytic effect size of retention on behavior (non-attendance) of -0.11 across 30 outcomes in 11 studies and an aggregated effect-size of -0.65 on attendance outcomes across two studies. Holmes (1989) reports a comparable effect size of -0.13 on behavioral (non-attendance) outcomes, but a much smaller yet still disadvantageous effect size of -0.18 on attendance outcomes.
[2] A theoretical rationale for this perspective comes from social comparison theory (Festinger, 1954; Wu, West, & Hughes, 2010), which posits that students draw conclusions about their ability from environmental cues, such as comparisons to peers. See Martin (2011) for a discussion of other theoretical frameworks with relevance for the link between grade retention and student outcomes, including absenteeism.
[3] Previous studies have examined impacts of this policy on assessment outcomes (Mariano et al., 2009; Mariano & Martorell, 2013).

Suspensions, on the other hand, are a direct measure of a disciplinary response to behavioral infractions, and prior work has noted a correlation between grade retention and suspensions (Raffaele, 1999). Examining suspensions is also interesting given recent concerns about high rates of suspension and racial disparities in suspension rates (Losen & Martinez, 2013).

Estimating the effect of grade retention is difficult because students that are retained likely would have worse outcomes than promoted students because they are retained precisely due to their difficulties in school. As documented in more recent literature, associations reported in the meta-analyses of Holmes (1989) and Jimerson (2001) may not reflect causal effects (Alexander, Entwisle, & Dauber, 2003; Allen, Chen, Willson, & Hughes, 2009; Hong & Raudenbush, 2005; Lorence et al., 2002; Mariano & Martorell, 2013). We address this concern by using an empirical strategy that exploits the fact that grade retention is triggered largely by whether a student scores below a threshold on a standardized test score. This allows us to use a fuzzy regression discontinuity (FRD) research design (Hahn, Todd, & Van der Klaauw, 2001; Imbens & Lemieux, 2008) that centers on comparisons of students who score just above or just below this cutoff. Under plausible and empirically supported assumptions (which we describe in detail below), this approach will generate valid estimates of the causal impact of grade retention for students scoring close to the cutoff score whose grade retention status is determined by falling above or below the cutoff score.

This study builds off of a number of recent studies that use a similar research design to look at the impact of grade retention on student outcomes. This research has generated mixed results, with some studies (Jacob & Lefgren, 2004; Roderick & Nagaoka, 2005) finding small and short-lived benefits of grade retention, and others finding larger and more persistent positive effects (Greene & Winters, 2007; Mariano, Kirby, & Crego, 2009; Mariano & Martorell, 2013;

Schwerdt, West, & Winters, 2017). Other studies examining the impact of grade retention on school completion have found that grade retention, particularly in later grades, increases dropout (Jacob & Lefgren, 2009; Manacorda, 2012).

However, we are only aware of one other study that attempts to estimate the causal effect of grade retention on behavioral outcomes. Ozek (2015) uses a fuzzy RD design to examine the impact of retention under Florida's third grade retention policy. He finds that retention generates a small increase in the likelihood of disciplinary incidents in the first two years post-retention, followed by a small decrease in such incidents in the third year. Our study builds on this analysis in two ways. First, we examine the effects of grade retention in a setting where the "marginal" retained student is at a lower point in the achievement distribution since a smaller proportion of students in New York City meet the criteria for assessment-based retention decisions than in Florida. Thus, our results provide new information about the effect of grade retention for students in need of the greatest academic assistance, which is a population of considerable interest.

A second contribution of the study is that, the NYCDOE policy covered grades 3-8, allowing us to examine how effects vary by grade. These analyses are motivated by developmental theories which posit that stressful life events such as grade retention may have different effects depending on a child's developmental stage (Anderson et al., 2005; Compas, 1987). In particular, the socially disruptive, stigmatizing, and academic effects all could vary by age, and hence by grade. Examining effects by a single grade rather than, for instance, grouping elementary and middle school grades, allows us to assess impacts in transition grades to middle or high school. The findings from earlier studies also provide further rationale for examining effects by grade, with more negative effects on outcomes such as high school completion found for later grades than earlier grades (Jacob & Lefgren, 2009; Manacorda, 2012; Schwerdt et al., 2017). This turns out to be important since we find some evidence that the effects of retention vary across grades.

Our analyses reveal no persistent systematic effects of student grade retention on disciplinary incidents and attendance. In the grade-specific analyses, we do find sporadic non-sustained significant effects of retention on behavioral outcomes. When present, these isolated non-persistent effects tend to be beneficial when found for retained elementary school students and mixed for retained middle school students.

Below we examine the broader NYCDOE promotion policy under study in more detail, including both retention and other treatments, and describe the available data. Next we describe our fuzzy regression discontinuity approach to estimating policy effects on behavioral outcomes. We then review our results and present a concluding discussion.

New York City's Student Promotion Policy

In the 2003-2004 school year, NYCDOE implemented an ambitious reform initiative that included a new assessment-based promotion policy for general education students in grade 3 (Mariano & Martorell, 2013; McCombs, Kirby, Marsh, & DiMartino, 2009). This policy was extended to grade 5 in the fall of 2004, to grade 7 in 2006-2007, to grade 8 in 2008-2009, and finally to grades 4 and 6 in 2009-2010, and it remained in place through the 2012–2013 school year. Students in charter schools, special education students and early English language learners were exempt from the policy.

This policy mandated reliance on standardized test scores for grade promotion decisions. Students who scored in the lowest performance level (Level 1)[4] on either the math or English Language Arts (ELA) annual spring assessments were at risk of being retained in grade under the

[4] There are four levels on the New York assessment reporting scale. Level 3 indicates proficiency.

policy. However, students who scored Level 1 in the spring had multiple additional opportunities to demonstrate eligibility for promotion (Mariano & Martorell, 2013; McCombs, Naftel, Ikemoto, DiMartino, & Gershwin, 2009), starting with a portfolio review at the end of the regular school year in June. Those who did not demonstrate Level 2 on the June portfolio were required to attend the City's summer instructional program that consisted of approximately twenty 4.5 hour instructional days, including at least 1.5 hours in each of ELA and math per day (see Ikemoto, McCombs, DiMartino, and Naftel, (2009), for a full description of summer school implementation).

At the conclusion of the summer program, these students took a City summer assessment in the subjects in which they had scored at Level 1 in the spring. Students still scoring Level 1 on the summer assessment were eligible for retention. The summer assessment program was a continuation of NYCDOE's prior spring testing program that predated NCLB state testing requirements. In 2001, when the State of New York spring assessments covered only fourth and eighth grades, NYCDOE contracted with CTB McGraw Hill to create criterion referenced spring assessments for third, fifth, sixth, and seventh grades aligned to the state curriculum, such that the NYCDOE had a contiguous spring testing program for grades three through eight.[5,6] When the promotion policy started in 2004, an alternate form of the spring NYCDOE assessment was used as the end of summer assessment. When the State began testing all grades three through eight in 2006, NYCDOE continued their curriculum aligned assessment program for use as the end of summer assessment.[7]

---

[5] Harcourt Assessment took over the annual ELA assessment development in 2003. By 2010, both tests were eventually contracted to Pearson.

[6] A standard setting exercise to develop cuts for level scores relative to curriculum standards was conducted in 2001; in future years, cut scores were determined through the equating of scale scores within grade across years. A new standard setting exercise was conducted in 2010, which updated the cut scores and the reporting scale.

[7] When the policy was extended to grades four and eight, aligned versions of the City assessments were added.

Students scoring Level 2 on the summer assessment had the opportunity for a second portfolio review in August. Those not demonstrating Level 2 or better performance by the end of this process were retained in grade unless the student's principal and community superintendent granted an exemption.[8] Despite multiple opportunities to meet the promotion criteria, below we show clear evidence that scoring below the summer assessment Level 2 cutoff on either math or ELA sharply increased the likelihood of being retained.

Grade retention was just one of the treatments students might have experienced under this comprehensive promotion policy. The policy complemented the threat of retention by placing considerable emphasis on an early intervention structure for struggling students. Kirby, McCombs, and Mariano (2009) note that the provision of supplemental services for struggling students, along with multiple opportunities to demonstrate Level 2 performance, placed NYCDOE's promotion policy in alignment with tenants for the validity and fairness of test-based promotion decisions articulated by the National Research Council (Heubert & Hauser, 1999).

At the beginning of each school year, schools identified students who were potentially at risk of retention. This group included students retained the prior year and students scoring at Level 1 or in the lower half of Level 2 on a prior year spring assessment. These students received a variety of additional instructional services both within and outside the classroom during the school day, as well as additional services outside of regular school hours (Ikemoto, et al., 2009; McCombs, et al., 2009). As detailed in the Methods section below, our research design focuses on comparisons of students with summer assessment scores at the boundary between Level 1 and Level 2. All students contributing to the identification of the effects of retention, whether

[8] Eighth grade students also had to pass core courses in ELA, math, science, and social studies to be promoted. In practice, it was highly uncommon for eighth grade students to pass a core course while scoring Level 1 on the spring assessment of the same subject.

promoted or retained, had a Level 1 spring assessment score in math, ELA, or both and were eligible for the same supplemental services.[9] The provision of the services, however, is an important contextual factor in evaluating the impact of retention under this policy. The effects of retention under this policy discussed below must be interpreted in that context; in the absence of the broader policy supports, the effects of retention may differ.

Data

Data Sources

NYCDOE has provided relevant administrative data for all cohorts of students from 2003–2004 through 2011–2012 in all grades subject to the promotion policy. Each cohort contains approximately 54,000 to 63,000 general education students subject to the promotion policy. Available baseline data include student characteristics, including, but not limited to, English language learner status, free or reduced lunch status, gender, and race/ethnicity; baseline and prior years' spring and summer ELA and mathematics assessment scores; baseline and prior years' attendance and suspension information; summer attendance record; school attended; and portfolio data for select cohorts. Behavioral outcome data include attendance and suspension data for each year post baseline through 2013–2014. The available suspension data include information on both the length of the suspension and the severity of an incident.

The summer assessment results are critical measures that trigger retention eligibility. As detailed by Mariano and Martorell (2013), the spring assessments determine eligibility for a separate treatment, summer instruction. The summer assessment determines retention, unless avoided through summer portfolio or other exemption. There were a number of changes to the scales used for the summer assessments during the study period, necessitating rescaling to a

[9] Promoted students would have received these services in the next grade; services would have been tailored to their needs. Promoted eighth grade students would not have received the supplemental services in high school.

common metric to enable cross-cohort analyses. Because the summer assessment is only administered to a non-random subset of the reference population, attempting to use the empirical standard deviation of the scores in a z-score transformation to place them on a common scale would overinflate the spacing of the scores. More importantly, with the testing drift present in the state spring test scores between 2007 and 2009, necessitating recalibration of the state spring test in 2010 (New York State Board of Regents, 2010), the subset of the population taking the assessments in those years is not expected to match, causing resultant z-scores to not be exchangeable over that time. As an alternative, we rescale the scores by dividing by the width of Level 2 on the reporting scale for each grade and year and centering so that zero represents the treatment cut point; we reference this as "Level 2 scaling" below. Dividing by the Level 2 width is a desirable transformation because it retains the relative distance between scale scores; the assessment designers used the width of Level 2 to equate scores across years (Pearson Psychometric and Research Services, 2011).[10] To accommodate both ELA and math summer assessment scores determining eligibility, we use the minimum of the two Level 2 scaled scores as the running variable (i.e., the variable determining treatment assignment eligibility) in the RD analyses, as further discussed below.

Outcomes

In this study we examine impacts of grade retention on behavioral problems. Specifically, we estimate effects of retention on measures of suspensions and school attendance. For suspensions, we consider whether a student has any suspensions as well as the number of days suspended. We also consider these measures separately for suspensions for less- and more-severe

[10] Note that in analyses that pool across cohorts and grades, the scaling we use could generate slight differences in the set of students close to the cutoff than an alternative scaling (e.g., z-score transformation). However, the results are similar in models that do and do not include grade-by-cohort effects, suggesting that this issue is unlikely to appreciably affect the results.

disciplinary infractions. Infractions are classified by NYCDOE into 5 categories. We define a suspension as a "less severe" suspension if it was for an infraction deemed as one of the 3 lower-level category infractions ("uncooperative/noncompliant behavior", "disorderly behavior", or "disruptive behavior") and a "more severe" suspension if it was for one of the two highest infraction levels ("aggressive or injurious/harmful behavior", "seriously dangerous or violent behavior"). For attendance, we examined effects on the attendance rate (defined as the number of days a student was in attendance divided by the number of days in the year in which a student was enrolled) as well as an indicator for "chronic absence", which the NYCDOE defines as having an attendance rate of less than 90 percent. The behavioral outcome measures are available for each year post baseline through 2013–2014.

An important issue for an analysis of the impacts of grade retention is the timing of when outcomes are measured. As we are examining non-academic outcomes, we follow the example in several recent studies on the impacts of grade retention (Greene & Winters, 2007; Roderick & Nagaoka, 2005; Schwerdt et al., 2017) and conduct "same-age comparisons." That is, we measure outcomes for promoted and retained students at the same point in time following the summer assessment when they are in different grades.

Appendix Table A1 shows means for the outcomes by grade. Examining these patterns is important because it helps to shed light on whether suspension and attendance are meaningful for younger students. Both suspensions and attendance vary strongly by grade, with higher attendance and lower suspension rates for younger students. This variance across grades in part motivates our analysis of the retention effects by grade. Despite the low suspension rates in elementary school (especially for third graders), suspensions certainly occur in every grade, and therefore we think suspension is a useful indicator of behavioral problems even for the youngest students in our sample. Similarly, even for younger students, the attendance rate is well below 100 percent, and the incidence of chronic absenteeism is over 30 percent for all grades.

Analytic Sample

Our analysis focuses on students who were subject to New York City's promotion policy and who took the summer assessment in the period 2004–2011 after receiving a Level 1 spring score. We further restrict the sample to those who have valid scores for at least one subject since our research design uses summer assessment scores as the running variable in the FRD estimation. There are 92,425 students meeting these criteria.

For all analyses, we further limit the sample to students who are enrolled in a New York City public school in the fall following the summer assessment and for whom we observe the grade in that year. This restriction applies to about 3 percent of the students taking the summer assessment and is necessary so that we can determine which students were retained. The final size of the analytic sample is also determined by the bandwidth chosen to fit the FRD model. For example, using a bandwidth of 1 results in an analytic sample of 76,167 students across all grades and cohorts subject to the policy. For analyses of outcomes two and three years following the summer assessment, we also limit the sample to students who were enrolled in those years.[11] Also note that for all analyses, we pool across all available cohorts subject to the promotion policy in the relevant grade.

Summary Statistics

Table 1 shows descriptive statistics for our analysis sample. Sixty-six percent of students who take the summer assessment are promoted. In the year prior to the summer assessment, about 8 percent of students are suspended at least once and on average spend 1.2 days suspended.

[11] One potential problem with restricting the sample to students who were enrolled in future years is that it may impart a sample selection bias if future enrollment in a New York City school is affected by the outcome of the summer assessment. We discuss this issue below.

The daily attendance rate is 90 percent, which is low considering the NYCDOE's definition of "chronic absence" is less than 90 percent attendance. Forty-nine percent of the students in our sample are black and 39 percent are Hispanic. In contrast, in NYCDOE as a whole in the 2012–2013 school year, only 29 percent of students were black, while 40 percent were Hispanic. About five percent of students are not enrolled in a NYCDOE school 2 years after the summer assessment and 9 percent are not enrolled 3 years after the summer assessment.

When comparing promoted and retained students, we see some notable differences on some characteristics but also a number of similarities. Most notably, promoted students have much better math and ELA scores in the spring prior to the summer assessment. Retained students are also 4 percentage points more likely to be black and 2 percentage points less likely to be Hispanic. There are not large differences in the attendance rates and proportion suspended in the year prior to the summer assessment; however, retained students averaged an additional 0.2 days suspended, 14 percent higher than the promoted students in the analysis sample. Attrition 3 years after baseline is nearly identical for retained and promoted students.

Sample sizes for the first year post-retention are indicated in Table 4. Third grade contains the most cohorts in the sample and, by far, the highest sample size at 25,898 students across all cohorts. Grade four contains the fewest students in the sample, 4,753, which is not surprising given the policy was implemented in grade four six years after the policy started in grade three.

Methods

Research Design

Our identification strategy rests on comparisons of students who were barely above and barely below the Level 2 cutoff on the summer assessment. While students scoring below Level

2 are likely to be different in many ways from those scoring Level 2 or higher, these differences are likely to be much smaller among those students scoring close to the Level 2 cutoff in all dimensions other than the probability of being retained. Therefore, we use a fuzzy regression discontinuity design (Hahn et al., 2001; Imbens & Lemieux, 2008) centering on comparisons between students who score just above and below the Level 1 summer assessment cutoff to identify the effect of grade retention.

Specifically, we estimate the following system of equations:

$$Y_i = \theta W_i + f_Y(X_i) + \varepsilon_i \tag{1}$$

$$W_i = \pi T_i + f_W(X_i) + v_i$$

where Y_i represents an outcome for student *i*, W_i denotes grade retention status, $f_Y(X_i)$ and $f_W(X_i)$ are control functions of the summer assessment score, X_i , and ε_i and v_i are residuals. The variable T_i is a dummy variable for scoring below the Level 2 cutoff, and serves as the instrumental variable (IV) for W_i. The parameter θ represents the effect of grade retention.[12] While not shown, we also control for baseline variables.[13] Doing so is not necessary for producing consistent estimates of the impact of retention (provided the fuzzy RD identification assumptions listed below hold); however, it does reduce residual variance and therefore improves precision.

The validity of this approach rests on two key conditions. The first is that barely falling below the Level 2 cutoff affects the probability of retention (i.e., $\pi \neq 0$), which we show below clearly holds in this context. The second is that barely falling above or below the Level 1 cutoff

---

[12] Note that this analytic design makes no assumptions about the mechanism by which grade retention may or may not affect students' behavioral outcomes.

[13] These covariates include grade-by-cohort dummies, English Language Learner status, race, gender, old for grade status, any days suspended, days suspended, attendance rate, missing attendance rate, spring math z-score, spring ELA z-score, missing math z-score, and missing ELA z-score.

only affects Y_i by changing the probability of being retained (i.e., that T_i and ε_i are uncorrelated). This condition requires that whether a student falls above or below the Level 2 cutoff is "as good as" randomly assigned in a narrow region around the cutoff score (Lee, 2008). We discuss evidence consistent with this assumption below. This second condition also requires that the only mechanism through which falling below the Level 2 cutoff affects student outcomes is by changing the likelihood of being retained. This seems plausible; the only purpose of the summer assessment is to determine whether students in the summer school program have mastered enough material to be promoted to the next grade. One important threat to validity is that there is differential attrition from the sample following the summer assessment. We present evidence below on this issue when discussing the validity of the research design.

When these assumptions hold, our research design will capture the "local" average effect for students whose academic ability makes them likely to score close to the Level 2 cutoff and whose retention status is affected by scoring below or above the Level 2 cutoff (Imbens & Angrist, 1994; Lee, 2008).[14] Our estimates will therefore not necessarily be applicable to students with scores very far away from the Level 2 cutoff nor for students who avoid being retained even when scoring below Level 2. Nonetheless, they will still be informative about the effects for a policy relevant subgroup. This is because the students scoring near the Level 2 cutoff are those deemed by NYCDOE to be just at the margin for needing and benefiting from retention. Similarly, a central feature of the policy was that retention be based in part on the outcome of standardized tests, and our estimates are applicable for students whose grade retention status is determined by test score performance.

---

[14] A further issue of interpretation concerns the pooling of students across grades and years since the Level 1/Level 2 cutoff varies by grade and year. As shown by Cattaneo et al. (2016), the pooled estimates in this context can be interpreted as a "double average," or the weighted average across cutoffs of the local average treatment effects for students facing a given cutoff value.

Estimation

As in any RD analysis, misspecification of $f_Y(X_i)$ and $f_W(X_i)$ is a serious concern as it could lead one to incorrectly conclude that there are (or are not) discontinuities at the Level 2 cutoff. We report results from both parametric and nonparametric methods of implementing the fuzzy RD design. The nonparametric estimates are obtained using the "local polynomial" regression estimator proposed by Calonico, Cattaneo, and Titiunik (2014) ("CCT local linear" method below). IV estimates are obtained as the ratio of the local linear estimates of the reduced-form and first-stage effects. Their approach uses a "data driven" procedure to obtain the bandwidth of scores around the Level 1 cutoff to use in the estimation, and local linear regression to estimate the treatment effects.[15] They show that their method generates more accurate confidence intervals than other nonparametric estimators that have been used in studies using regression discontinuity. The parametric model produces IV estimates of θ via 2-stage least squares (2SLS) regression, where the functions f_Y and f_W are modeled as a second degree polynomial ("2nd order polynomial" below).[16] To show robustness to bandwidth choice, we report results for the parametric models with two different bandwidths: 1.0 and 0.5. These bandwidths were chosen since most of the selected CCT bandwidths are between 0.5 and 1.0. Although the summer assessment Level 1 data from most grades and cohorts extend at least four times the width of the Level 2 scores (i.e., extends to at least a bandwidth of 4.0), about 75 percent of the summer assessment data fall within a bandwidth of 1.0.

[15] We actually use a simple extension of the method proposed by CCT to account for baseline covariates. Specifically, we use the residuals from a linear regression of a given outcome on the complete set of baseline covariates as the dependent variable in the CCT local linear procedure. Provided that the baseline covariates are uncorrelated with T_i close to the Level 1 cutoff (which is true in a valid RD design), the point estimates from this modified procedure should reconcile with estimates obtained if we just used the unadjusted outcomes as the dependent variable.

[16] We use linear models despite the likely nonlinearity of the true conditional expectation functions that we are attempting to estimate. This is because our goal is to estimate local average treatment effects at the Level 1/Level 2 cutoff, which can be done successfully using linear IV without having to make further parametric modeling assumptions (Angrist and Pischke, 2009).

An important issue for implementing this approach in this setting is that the summer assessment consists of two scores, whereas the framework developed above assumed a one-dimensional running variable. We account for dual assignment scores by using the minimum of the math and ELA scores as the running variable (where each subject score has been rescaled to be zero at the passing cutoff, after scaling by the width of the Level 2 score range, as discussed above). Since students are assigned to retention if they score below the Level 2 cutoff on either math or ELA, whether a student is assigned to retention is determined by whether or not this minimum score variable is negative.[17]

In addition to the results reported, we further explored the robustness of the results by fitting several additional model specifications. To examine the sensitivity of function form and bandwidth choices, we also fit several additional specifications, including first through third degree polynomial estimates at bandwidths of 0.5 and 1.0, as well as global polynomial estimates (exclusive of the lowest observable scale score). To explore the sensitivity of results to pooling estimates, both across subject areas and school years, we also considered model specifications that fit the estimates using each subject separately, as well as specifications that consider estimates both before and after 2010, when the NYCDOE summer assessment was rescaled. The results of these additional robustness checks are consistent with the general conclusions reported below are available from the authors upon request.

A final consideration for the estimation is how we account for the fact that we examine numerous outcomes separately by grade. To account for the simultaneous estimation of a series

[17] Wong et al. (2013) refer to this method as the "centering approach" and Reardon and Robinson (2012) refer to it as the "binding-score RD". For sensitivity, we also ran each of the models below using a single-assessment running variable; conclusions for both individual assessments as the running variable were similar to those using the minimum score.

of treatment effects, we implement a False Discovery Rate adjustment (Benjamini & Hochberg, 1995) at a rate of 0.05. Given that the outcomes examined are strongly correlated, for example, incidence of suspension and days suspended, the False Discovery Rate adjustment based upon the total number of tests is likely too conservative in compensating for the multiple testing. Therefore, in the tables below, we report statistical significance with and without the adjustment for multiple hypotheses testing. We discuss results based upon the unadjusted estimates.

RD Validity

As noted above, our approach requires that scoring below the Level 2 cutoff require quasi-random assignment local to the Level 2 cutoff. This assumption is plausible in this context. Neither the teachers nor students know the number of correct answers needed to meet the Level 2 standard.[18] Even with no test score manipulation, restricting the sample to students who remained in a NYCDOE school in the years following the summer assessment could undermine the research design if falling below the Level 2 cutoff affects the composition of students remaining in the sample. For instance, if students assigned to retention are more likely to leave the sample, the estimated impacts of retention could be upward biased if the students induced to leave were ones who would tend to have higher rates of behavioral problems than those who remained. On the other hand, if students induced to leave were the ones with more motivated parents and this was associated with lower rates of behavioral problems, there would be downward bias.[19] We investigate sample attrition in Table 2 by modeling the probability of enrollment using a strict RD model with the same minimum scale score running variable used in our primary analyses. Scoring just below Level 2 reduces the probability or being enrolled in

[18] Multiple choice answer sheets were scanned locally at the summer school test site and automatically uploaded to the DOE's administrative database, where a scoring key and raw score to scale score conversion were applied.
[19] The inclusion of baseline control variables in the FRD model should help mitigate potential bias resulting from restricting the sample to students who remained in a NYCDOE school following the summer assessment.

Year 1 (the academic year following the summer assessment) by about 1 percentage point.[20] Although these estimates are statistically significant, the magnitude of the differential attrition is small and the overall Year 1 attrition rate is very low; almost 97 percent of students are observed enrolled in the year after the summer assessment. [21] For Years 2 and 3, the attrition rates are much higher but there is no evidence of differential attrition at the Level 2 cutoff.

We also conduct standard investigations of the validity of the RD design. Since these analyses were done on the main analytic sample – students who enrolled in Year 1 – they would capture any violations in the research design due to either differential attrition or from manipulation of the summer assessment scores at the Level 2 cutoff. First, we examine whether students are equally likely to score above or below the Level 2 cutoff by implementing the procedure proposed in McCrary (2008) to test for the presence of a discontinuity in the density of test scores. The estimated discontinuity in the log of the density is .022 with a standard error of .031 ($p=0.483$),[22] and as shown in Figure 1, there is no visual indication that the distribution of test scores is discontinuous at the Level 2 cutoff. We replicated this test within each grade and for each subject individually across and within grade. No significant indications of a discontinuity were discovered. Second, we tested whether there were sharp differences in baseline covariates at the Level 2 cutoff. Table 3 shows estimated discontinuities in baseline characteristics at the Level 2 cutoff. None of these estimates were found to be statistically significant for the CCT local linear or 2nd order polynomial specifications. Note that this analysis was conducted for the sample of students that remained in the data through year 1, so if

[20] Throughout this article, we refer to the *Xth* year after the summer assessment as "Year X".
[21] Note that for this attrition analysis, we include all students who took the summer assessment (and who met other inclusion criteria), including those who attrit before Year 1. This allows us to test for differential attrition in Year 1. In contrast, in our primary analyses, we restrict the sample to those who enroll in Year 1 so that we can determine whether they were retained (which is necessary to conduct the fuzzy regression discontinuity analysis).
[22] As suggested in McCrary (2008) we compute the basic bandwidth and use half that bandwidth to reduce bias in estimating the discontinuity.

differential attrition were generating differences in observable characteristics at the Level 1/Level 2 cutoff, it would have been detectable in this analysis. Overall we interpret these results as supporting the assumptions underlying the research design, despite some differential attrition in Year 1.

Results

First Stage Estimates

Table 4 displays first-stage estimates of the effect of scoring Level 1 on the summer assessment on the probability of being retained for each policy grade and also pooled across all grades. The pooled estimates demonstrate a strong increase in the probability of retention for Level 1 students. Level 1 students around the Level 2 cut-point have a probability of retention approximately 0.65 higher than their Level 2 counterparts at the cut-point, and the F-statistics for the hypothesis that the discontinuity is zero are all greater than 400, well above the thresholds commonly-used to determine whether there are "weak instruments" (Staiger & Stock, 1997; Stock & Yogo, 2005). These pooled first stage results are displayed graphically in Figure 2. As seen in the figure, the probability of retention for the Level 2 students at the cut point is essentially zero, with only the Level 1 students being retained, consistent with expectations under the policy. Examining the probability of retention results by grade in Table 4, the first stage effect for Level 1 eighth grade students at the Level 2 cut point is lower than the pooled average, at approximately 0.50 (note the grade-specific estimates are produced by separate regressions by grade). All other grades have a first stage effect of at least 0.63 or higher, with fourth graders showing a first stage effect of 0.74. These results evidence implementation of the retention portion of the policy and support the FRD design detailed above.

Effects of Retention on Suspensions

Table 5 shows estimates of the impact of retention on suspension outcomes pooled over all grades, and Tables 6a and 6b break these estimates out by grade (Appendix Table 1 shows means of the dependent variables by grade). Incidence of suspension and the total days suspended are both considered, up to three years removed from the year of retention. The pooled estimates shown in Table 4 show no significant effects of retention on suspension incidents or cumulative suspension days. This is corroborated by the visual evidence in Figure 3 which shows that both the incidence of suspension and number of days suspended do not change discontinuously at the Level 1 cutoff score.

However, several individual grade-specific results are present. The Year 1 estimates for fifth graders (Table 6a) suggest retention may reduce suspensions, although the estimates for both the incidence of suspension and number of days are not always statistically significant across specifications. However, these effects are not statistically significant for Years 2 or 3, and we find no robust evidence of retention effects for third or fourth graders. For older grades (Table 6b), retention generates an increase of about 2-3 days suspended in Year 2 for seventh graders. However, there are no other effects found for seventh graders. Similarly, we do no find consistent evidence across specifications of effects for sixth or eighth graders.

Taken together these results suggest that a systematic or persistent effect of retention on suspension outcomes is not present in the aggregate or at any individual grade. We do find some instances where the estimates indicate that retention in a particular grade increases or decreases suspensions, but these effects never extend beyond a single year.

Effects of Retention on Attendance

Table 5 displays estimates of the effects of retention on attendance pooled over all grades and Tables 7a and 7b considers these effects by grade. When pooling across grades, we do not find any evidence that retention affects the incidence of chronic absence. This is consistent with the graphical evidence in Figure 4. For the attendance rate, the Year 1 effect is small and statistically insignificant. The Year 2 estimates suggest retention increases the attendance rate by about 1 percentage point. The Year 3 estimates are of similar magnitude but are less precise and statistically significant only for the parametric model with a bandwidth of 1. The graphical evidence in Figure 4 is not definitive, but offers some indication that the attendance rate in Years 2 and 3 is a little higher for students barely falling below the Level 2 cutoff than for those scoring barely above it.

Turning to the effects by grade, we do not find any evidence of effects on the attendance rate or chronic absenteeism for grades 4 and 6. In contrast, the estimates suggest that grade retention reduces chronic absenteeism for third graders by about 5 percentage points in Years 1 and 3 (the estimated effect for Year 2 is smaller and statistically insignificant).[23] At the same time, we find little evidence that retention in grade 3 increases the overall attendance rate; only the Year 3 estimate from the 2nd order polynomial specification with a bandwidth of 1 is statistically significant. For fifth graders, retention increases the Year 1 attendance rate by 1 percentage point, but the effects for Year 2 and Year 3 were small and statistically insignificant. The estimated effects on chronic absenteeism were all statistically insignificant.

The estimates for seventh graders indicate retention improves the attendance rate in Year 2 by 4-5 percentage points. The estimates for Year 3 are similar in magnitude but not statistically significant for the bandwidth of 0.5 specification. The Year 1 estimates are modest in magnitude (about 1.5 percentage points) and statistically significant only for the parametric model with a bandwidth of 1. The point estimates all suggest that retention may have lowered the incidence of chronic absence, but none are statistically significant.

[23] Estimates beyond year three are also essentially nil.

For eighth grade, we find that retention increases the likelihood of Year 1 chronic absenteeism by 12-15 percentage points. This worsening of chronic attendance dissipates beyond Year 1; the effects are not statistically significant for Years 2 and 3. While the chronic attendance point estimates drop from Year 1 to Year 2, they remain sizable in Year 2 before dropping further in Year 3. We also find negative effects of retention on the Year 1 attendance rate, although this is only statistically significant for the CCT estimator. For Year 2 and 3, however, we find non-significant estimates of retention on the overall attendance rate that are much smaller and sometimes positive. Thus any effect after Year 1 on chronic absenteeism appears to be driven by small changes in the attendance rate of students near the threshold for chronic absenteeism rather than a wholesale change in the attendance patterns following retention.

To summarize, the significant effects on increased attendance rates in the pooled grade estimates in the second year post-retention appear to be driven primarily by effects among students retained in seventh grade, as the Year 2 and 3 effects on attendance rates for third through sixth graders are either small or negative. The results for seventh graders and the pooled results also stand in contrast to the results for eighth graders, where we find evidence that grade retention leads to a temporary but sizable increase in chronic absenteeism.

Discussion

The results discussed above do not provide evidence of systematic, persistent effects of retention on attendance or suspension outcomes. These findings are in the context of the broader promotion policy instituted by NYCDOE over the cohort years examined, which offered multiple supplemental services in the proximal and following cohort years to both retained and marginally promoted students. When we pool across grades, the estimates tend to be small and statistically insignificant. An exception is that retention appears to have a modest positive effect on Year 2 attendance, which is driven mainly by effects for seventh graders. Otherwise, we do not find strong evidence of pooled effects of retention on behavioral outcomes.

In the grade-specific analyses, we find some "one off" cases where retention appears to affect suspensions or attendance, but these effects do not persist for multiple years post-retention and do not consistently indicate positive or negative effects of retention on behavioral problems. For elementary school grades, some of the results suggest that retention reduces suspensions in the year that retained students were in fifth grade and their promoted counterparts were in sixth grade. However, any such effects do not persist once the retained students would be in sixth grade. A similar pattern appears for isolated improvements in attendance and chronic absenteeism. For middle school grades, we find some conflicting evidence of the effect of retention on attendance, with retention in seventh grade potentially improving attendance while retention in eighth grade increasing the incidence of chronic absenteeism in the following year. Otherwise, any statistically significant effects we find did not last for more than one year.

One interesting pattern that emerges from the analysis of grade-specific estimates is that most of the instances of statistically significant estimates coincide with the transitions into and out of middle school (in New York City, middle schools most commonly cover grades 6 through 8). Thus, to the extent that retention affects suspensions or attendance in our sample, it may be driven by differences in prevalence of suspensions and absenteeism in elementary, middle, and high school. Moreover, since suspensions and truancy are quite rare in elementary grades, these may be a coarse indicator of disciplinary problems in those grades. However, it is not always clear from our results how retention interacts with middle school transitions. For instance, we find that retention of seventh graders improves attendance in the year that promoted students would be in ninth grade, but that retention of eighth graders worsens chronic absence in the year that promoted students would be in ninth grade. This suggests retention during middle school may provoke different short-term reactions depending on the proximity of the retention to ascension to high school.

There are two important caveats to consider with respect to our findings. The effects estimated in this analysis apply to students near the treatment cut-point between Level 1 and Level 2 scoring on the summer assessment. In addition, the FRD design produces treatment effect estimates for the students whose retention status is determined by scoring Level 1 on the summer assessment. The “fuzziness” in the assignment process reflects the use of non-test score factors (e.g., portfolios) to make retention decisions, which aligns with recommendations concerning assessment-based policies (Heubert & Hauser, 1999). However, an implication for our study is that our findings might not be applicable for students whose test scores fell below the grade promotion standard but were nonetheless promoted.

Attendance and suspensions, while being important metrics in understanding the unintended consequences of retention, are relatively broad measures that, of course, do not fully describe student behavior. This may be particularly true in the elementary grades where truancy is less an option of autonomous choice and if suspension is less likely to be utilized in response to inappropriate behavior. Research examining additional facets of student behavior, such as teacher and peer interactions, classroom citizenship, and non-academic classroom function, as well as broader socio-emotional outcomes would help inform a more complete understanding of the impact of retention. Le, Mariano, and Crego (2009) found that retained elementary students in New York City had comparable self-confidence in reading and math and a higher sense of school belonging that their promoted peers, and also emphasized the need to further explore non-academic outcomes of retention.

Our results on attendance and suspension outcomes stand in sharp contrast to those of an earlier literature that finds that grade retention is positively correlated with such behavioral problems. We think this divergence is most likely attributable to our use of a research design that delivers causal estimates of the impact of retention. These results broaden our understanding of the causal impact of retention on future behavioral outcomes in two ways. First, we expand upon Ozek's (2015) prior causal examination of retention based upon third grade ELA performance by examining retention decisions for multiple grades, three through eight, and based upon proficiency in multiple subjects, ELA and math. Second, our effect estimates correspond to a different point on the distribution of proficiency in New York City that that of Ozek's (2015) prior examination of similar outcomes in Florida. The two treatment cut-points are not directly comparable; unlike NYCDOE's summer assessment administered to the spring Level 1 population, Florida's spring assessments, administered to all students, were used to determine retention assignment. However, sixteen percent of Florida third grade students were eligible for summer school, and 8 percent were retained, as opposed to seven percent of third graders required to attend summer school and three percent retained in New York City.[24] In contrast to our results, Ozek (2015) does find that retention generated a short-term increase in behavioral problems, followed by a significant decrease in the third post-retention year. The presence of these differences is a reminder that retention policies in various jurisdictions each have local characteristics. Such characteristics need to be considered in interpreting our and other results and highlight the need to continue to expand the examination of the effects of retention on behavioral outcomes across broader policy circumstances using causal designs.

[24] Across all grades and cohorts in the years studied, six percent of eligible New York City students were required to attend summer school and two and a half percent were retained.

Our results demonstrate that retaining students in elementary and middle school does not necessarily increase behavioral problems as measured by absences and suspensions. Contrary to the prior observational literature, this paper provides an example of a formal test-based promotion policy that did not generate systematic negative effects on attendance and suspension, at least for students scoring close to the test score threshold used for determining grade promotion and in the context of the NYCDOE promotion policy. The instances where we do find negative effects of retention for middle schoolers only last for a single year rather than being persistent long-term effects. Although impacts on behavioral outcomes are only one consideration for evaluating the pros and cons of a retention policy, the experiences of the policy in place in New York City over the period examined imply that the consideration of implementing such policies do not necessarily start at a deficit with respect to unintended behavioral consequences.

References

Alexander, K. L., Entwisle, D. R., & Dauber, S. L. (2003). *On the Success of Failure: A Reassessment of the Effects of Retention in the Primary School Grades* (2 ed.). Cambridge, UK: Cambridge University Press.

Allen, C., Chen, Q., Willson, V., & Hughes, J. (2009). Quality of Design Moderates Effects of Grade Retention on Achivement: A Meta-Analytic, Multi-Level Analysis. *Educational Evaluation and Policy Analysis, 31*(4), 480-499. doi:10.3102/0162373709352239

American Federation of Teachers. (1997). *Passing on Failure: District Promotion Policies and Practices*. Retrieved from https://eric.ed.gov/?id=ED421560

Angrist JD, Pischke JS. 2008. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton, NJ: Princeton Univ. Press

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

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society, Series B, 57*(1), 289-300.

Byrd, R. S., Weitzman, M., & Auinger, P. (1997). Increased Behavior Problems Associated With Delayed School Entry and Delayed School Progress. *Pediatrics, 100*(4), 654–661.

Cattaneo, M.D., Keele, L., Titiunik, R., & Vazquez-Bare, G. (2016). "Interpreting Regression Discontinuity Designs with Multiple Cutoffs." *The Journal of Politics*, 78(4), 1229-1248.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica, 82*(6), 2295–2326.

Compas, B. E. (1987). Stress and life events during childhood and adolescence. *Clinical Psychology Review, 7*(3), 275-302.

Festinger, L. (1954). A theory of social comparison processes. *Human relations, 7*(2), 117-140.

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

Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica, 69*(1), 201-209.

Heubert, J. P., & Hauser, R. M. (1999). *High Stakes: Testing for Tracking, Promotion, and Graduation*. Washington, D.C.: National Academies Press.

Holmes, C. T. (1989). Grade level retention effects: A meta-analysis of research studies. In L. Shepard & M. L. Smith (Eds.), *Flunking grades: Research and policies on retention* (pp. 16-33). London: The Falmer Press.

Hong, G., & Raudenbush, S. W. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis, 27*(3), 205-224.

Ikemoto, G. S., McCombs, J. S., DiMartino, C., & Naftel, S. (2009). Implementation of the Policy: Saturday and Summer Schools. In J. S. McCombs, S. N. Kirby, & L. T. Mariano (Eds.), *Ending social promotion without leaving children behind: The case of New York City* (pp. 81-100). Santa Monica, CA: RAND Corporation.

Imbens, G., & Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica,, 62*(2), 467–475.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of econometrics, 142*(2), 615-635.

Jackson, G. B. (1975). The research evidence on the effects of grade retention. *Review of educational research, 45*(4), 613-635.

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

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

Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School psychology review, 30*(3), 420.

Katz, L. F., Kling, J. R., & Liebman, J. B. (2001). Moving to opportunity in Boston: Early results of a randomized mobility experiment. *The Quarterly Journal of Economics, 116*(2), 607-654.

Le, V., Mariano, L.T., and Crego, A. (2009). The Impact of New York City's Promotion Policy on Students' Socio-emotional Status. In J. S. McCombs, S. N. Kirby, & L. T. Mariano (Eds.), *Ending Social Promotion Without Leaving Children Behind: The Case of New York City* (pp. 165-182). Santa Monica, CA: RAND Corporation.

Lee, D. S. (2008). Randomized experiments from non-random selection in US House elections. *Journal of econometrics, 142*(2), 675-697.

Lorence, J., Dworkin, A. G., Toenjes, L. A., Hill, A. N., Rotherham, A., & Shepard, L. A. (2002). Grade retention and social promotion in Texas, 1994-99: Academic achievement among elementary school students. *Brookings papers on education policy*, 2002, 13-67.

Losen, D. J., & Martinez, T. E. (2013). *Out of School and Off Track: The Overuse of Suspensions in American Middle and High Schools*. Los Angeles, CA: Center for Civil Rights Remedies at the UCLA Civil Rights Project

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

Mariano, L. T., Kirby, S. N., & Crego, A. (2009). Future Outcomes of Students at Risk of Retention. In J. S. McCombs, S. N. Kirby, & L. T. Mariano (Eds.), *Ending Social Promotion Without Leaving Children Behind: The Case of New York City* (pp. 143-163). Santa Monica, CA: RAND Corporation.

Mariano, L. T., & Martorell, P. (2013). The academic effects of summer instruction and retention in New York City. *Educational Evaluation and Policy Analysis*, 35(1), 96-117.

Marsh, J., Gershwin, D., Kirby, S., & Xia, N. (2009). *Retaining students in grade: Lessons learned regarding policy design and implementation*. Santa Monica, CA: RAND Corporation

Martin, A. J. (2011). Holding back and holding behind: grade retention and students' non-academic and academic outcomes. *British Educational Research Journal, 37*(5), 739-763.

McCombs, J., Kirby, S., Marsh, J., & DiMartino, C. (2009). Context and conceptual framework for understanding New York City's promotion policy. In J. S. McCombs, S. N. Kirby, & L. T. Mariano (Eds.), *Ending social promotion without leaving children behind: The case of New York City* (pp. 17-32). Santa Monica, CA: RAND Corporation.

McCombs, J. S., Naftel, S., Ikemoto, G. S., DiMartino, C., & Gershwin, D. (2009). School-Provided Support for Students: Academic Intervention Services. In J. S. McCombs, S. N.

Kirby, & L. T. Mariano (Eds.), *Ending Social Promotion Without Leaving Children Behind: The Case of New York City* (pp. 17-32). Santa Monica, CA: RAND Corporation.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics, 142*(2), 698-714.

National Council of Teachers of English. (2015). Resolution on Mandatory Grade Retention and High-Stakes Testing [Press release]. Retrieved from http://www.ncte.org/positions/statements/grade-retention

New York State Board of Regents. (2010). Regents Approve Scoring Changes to Grade 3-8 Math and English Tests [Press release]. Retrieved from http://www.nysed.gov/news/2015/regents-approve-scoring-changes-grade-3-8-math-and-english-tests

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

Pearson Psychometric and Research Services. (2011). *New York City English Language Arts and Mathematics Summer Assessment Program 2011 Technical Report*. San Antonio, Texas: Pearson Psychometric and Research Services.

Raffaele L.M. (1999). *An analysis of out-of-school suspensions in Hillsborough County*. Tampa, FL: The children's Board of Hillsborough County.

Roderick, M. (1994). Grade retention and school dropout: Investigating the association. *American Educational Research Journal, 31*(4), 729-759.

Roderick, M., & Nagaoka, J. (2005). Retention under Chicago's high-stakes testing program: helpful, harmful, or harmless? *Educational Evaluation and Policy Analysis, 27*(4), 309-340.

Rohrman, D. (1993). Combating Truancy in Our Schools--A Community Effort. *NASSP Bulletin, 76*(549), 40-45.

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(1), 154–169.

Segal, C. (2013). Misbehavior, education, and labor market outcomes. *Journal of the European Economic Association, 11*(4), 743-779.

Staiger, D. O., & Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica* 65(3), 557-586.

Stock, J. H., & Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. In J. H. Stock & D. W. K. Andrews (Eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg* (pp. 80-108). New York, NY: Cambridge University Press.

West, M. R. (2012). Is retaining students in the early grades self-defeating. *CCF Brief, 49*. Retrieved from http://www.brookings.edu/~/media/research/files/papers/2012/8/16-student-retention-west/16-student-retention-west.pdf

Wong, V. C., Steiner, P. M., & Cook, T. D. (2013). Analyzing regression-discontinuity designs with multiple assignment variables a comparative study of four estimation methods. *Journal of Educational and Behavioral Statistics, 38*(2), 107-141.

Workman, E. (2014). Third-grade reading policies. Denver, CO: Education Commission of the States. Retrieved from https://www.ecs.org/clearinghouse/01/16/44/11644.pdf

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

Table 1: Sample Summary Statistics

	All tested students	Analysis Sample (bandwidth of 1 and conditional on observing grade in year following summer assessment)		
		All	Promoted	Retained
Outcomes from year prior to summer assessment				
Any suspension	0.082	0.084	0.083	0.084
Days suspended	1.188	1.177	1.121	1.286
Attendance rate	0.899	0.901	0.904	0.894
Attendance rate missing	0.000	0.000	0.000	0.000
Spring ELA z-score	-1.000	-0.993	-0.929	-1.120
Spring Math z-score	-1.129	-1.104	-1.015	-1.278
Spring Math missing	0.011	0.008	0.009	0.007
Spring ELA missing	0.036	0.029	0.031	0.025
Student Demographics				
Old for grade	0.369	0.367	0.372	0.357
Ever ELL	0.144	0.148	0.156	0.134
Male	0.520	0.520	0.517	0.526
Hispanic	0.387	0.394	0.403	0.376
Black	0.499	0.494	0.479	0.522
White	0.030	0.029	0.033	0.022
Other race	0.084	0.083	0.085	0.081
Enrolled in future years				
Enrolled 1 yr after	0.968	1.000	1.000	1.000
Enrolled 2 yrs after	0.927	0.952	0.953	0.949
Enrolled 3 yrs after	0.891	0.914	0.913	0.914
Sample size	92425	76167	50413	25754

Table 2: Fraction Enrolled in NYCDOE Schools by Year Since Summer Assessment

	Estimated Effect		
Year 1	-0.012** (0.004)	-0.012** (0.004)	-0.010** (0.003)
Year 2	-0.003 (0.006)	-0.003 (0.006)	-0.005 (0.005)
Year 3	-0.009 (0.006)	-0.007 (0.009)	-0.010 (0.006)
Bandwidth	0.5	1	Varies
Specification	2^{nd} order Polynomial	2^{nd} order Polynomial	CCT Local Linear

Note: Entries are estimated discontinuities in the probability of being enrolled at the Level 1/Level 2 cutoff. The estimated standard errors are in parentheses (these are adjusted for clustering at the running variable level in the parametric models). Results in the first two columns use a quadratic function of the running variable (the minimum of the rescaled math and ELA summer assessment scores), where the function is allowed to have different coefficients on either side of the Level 2 cutoff. Results in the final column are based on the local linear estimator proposed by Calonico et al. (2014).
* Significant at $\alpha = 0.05$; ** Significant at $\alpha = 0.01$

Table 3: Estimated Discontinuity in Baseline Covariates (Coefficient on "Scoring Level 1")

Covariate	Estimated Discontinuity (with s.e.)		
Any days suspended (w/o 2004 and 2005 cohorts)	0.010 (0.008)	0.006 (0.006)	-0.008 (0.007)
Days suspended (w/o 2004 and 2005 cohorts)	-0.099 (0.194)	-0.136 (0.148)	0.077 (0.151)
Any days suspended	0.008 (0.007)	0.005 (0.005)	-0.007 (0.006)
Days suspended	-0.105 (0.167)	-0.113 (0.124)	0.092 (0.121)
Attendance rate	-0.002 (0.002)	-0.002 (0.002)	0.002 (0.002)
Missing attendance rate	0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)
Spring ELA z-score	-0.023 (0.014)	-0.003 (0.011)	-0.004 (0.013)
Spring Math z-score	0.006 (0.016)	0.008 (0.012)	0.011 (0.013)
Missing spring math	-0.001 (0.002)	-0.001 (0.002)	-0.004 (0.004)
Missing spring ELA	0.002 (0.004)	0.003 (0.003)	0.001 (0.002)
Grade	0.011 (0.045)	-0.026 (0.034)	-0.016 (0.048)
Cohort	0.027 (0.065)	0.008 (0.049)	0.027 (0.062)
Old for Grade	-0.013 (0.012)	-0.009 (0.009)	0.013 (0.009)
Ever ELL	0.004 (0.009)	0.001 (0.007)	-0.001 (0.008)
Male	-0.009 (0.012)	0.005 (0.009)	0.000 (0.010)
Hispanic	0.000 (0.012)	-0.002 (0.009)	0.001 (0.009)
Black	-0.001 (0.012)	0.004 (0.009)	-0.003 (0.010)
White	-0.006 (0.004)	-0.005 (0.003)	0.006 (0.003)
Bandwidth	0.5	1	Varies
Specification	2^{nd} order polynomial	2^{nd} order polynomial	CCT Local Linear

Note: Cell entries are estimated discontinuities in baseline variables at the Level 1/Level 2 cutoff, and the estimated standard errors are in parentheses (adjusted for clustering at the running variable level in the parametric models). Results in the first two columns use a quadratic function of the running variable (the minimum of the rescaled math and ELA summer assessment scores), where the function is allowed to have different coefficients on either side of the Level 2 cutoff. Results in the final column are based on the local linear regression estimator proposed by Calonico et al. (2014).
* Significant at $\alpha = 0.05$; ** Significance at $\alpha = 0.01$

Table 4: Estimated Effect of Being Level 1 on Summer Assessment on Probability of Retention

Grade	Estimated First-Stage Effect		
All (N=76,167)	0.642**,f	0.648**,f	0.646**,f
	(0.009)	(0.006)	(0.008)
3 (N=25,898)	0.676**,f	0.686**,f	0.682**,f
	(0.015)	(0.011)	(0.012)
4 (N=4,753)	0.768**,f	0.736**,f	0.739**,f
	(0.033)	(0.023)	(0.024)
5 (N=14,186)	0.621**,f	0.632**,f	0.639**,f
	(0.020)	(0.015)	(0.016)
6 (N=10,475)	0.686**,f	0.682**,f	0.681**,f
	(0.021)	(0.016)	(0.021)
7 (N=11,684)	0.653**,f	0.654**,f	0.653**,f
	(0.021)	(0.016)	(0.016)
8 (N=9,171)	0.489**,f	0.501**,f	0.501**,f
	(0.025)	(0.019)	(0.021)
Controls?	Y	Y	Y
Bandwidth	0.5	1	Varies
Specification	2nd Order Polynomial	2nd Order Polynomial	CCT Local Linear

Notes: Sample sizes refer to the sample enrolled in the year after the summer assessment and that fall within a bandwidth of one from the Level 1/Level 2 cutoff. Cell entries are estimated discontinuities in the probability of being retained at the Level 1/ Level 2 cutoff, and the estimated standard errors are in parentheses (adjusted for clustering at the running variable level in the parametric models). Results in the first two columns use a quadratic function of the running variable (the minimum of the rescaled math and ELA summer assessment scores), where the function is allowed to have different coefficients on either side of the Level 2 cutoff. Results in the final column are based on the local linear regression estimator proposed by Calonico et al. (2014).
* Significant at $\alpha = 0.05$; ** Significant at $\alpha = 0.01$; [f] Significant after a False Discovery Rate correction, $\alpha = 0.05$

Table 5: Effect of Retention on Suspension and Attendance, Pooled Grades 3-8

Outcome	Estimated Treatment Effect (with s.e.)		
Any Suspension, 1 yr after (CCM=0.125)	-0.010 (0.012)	-0.014 (0.009)	-0.013 (0.009)
Any Suspension, 2 yrs after (CCM=0.109)	-0.001 (0.012)	0.004 (0.009)	-0.001 (0.009)
Any Suspension, 3 yrs after (CCM=0.114)	0.017 (0.015)	0.012 (0.011)	0.010 (0.011)
Days suspended, 1 yr after (CCM=1.519)	0.322 (0.302)	0.221 (0.238)	0.146 (0.228)
Days suspended, 2 yrs. after (CCM=1.426)	0.263 (0.387)	0.402 (0.282)	0.363 (0.294)
Days suspended, 3 yrs. after (CCM=2.115)	-0.007 (0.477)	-0.078 (0.342)	-0.136 (0.303)
Attendance rate, 1 yr after (CCM=0.882)	0.003 (0.004)	0.005 (0.003)	0.003 (0.003)
Attendance rate, 2 yrs after (CCM=0.868)	0.012* (0.005)	0.010**,f (0.004)	0.010** (0.004)
Attendance rate, 3 yrs after (CCM=0.846)	0.010 (0.007)	0.011* (0.005)	0.010 (0.005)
Chronic absence, 1 yr after (CCM=0.414)	-0.010 (0.016)	-0.018 (0.012)	-0.020 (0.012)
Chronic absence, 2 yrs after (CCM=0.418)	-0.004 (0.017)	-0.014 (0.013)	-0.012 (0.012)
Chronic absence, 3 yrs after (CCM=0.461)	-0.020 (0.021)	-0.017 (0.015)	-0.021 (0.015)
Controls?	Y	Y	Y
Bandwidth	0.5	1	Varies
Specification	2nd Order Polynomial	2nd Order Polynomial	CCT Local Linear

Notes: Entries are IV estimates of the effect of retention on a given outcome with standard errors in parentheses (adjusted for clustering at the running variable level in the parametric models). Results in the first two columns are 2SLS estimates from a parametric model that uses a quadratic function of the running variable. Results in the final column are IV estimates generated by the local linear regression estimator proposed by Calonico et al. (2014). Sample means are "control complier means" (Katz, Kling, & Liebman, 2001) that are calculated as the difference between the mean outcome for retained students who score just below the Level 1/2 cutoff (i.e., "treated complier mean") and the IV estimate.

* Significant at $\alpha = 0.05$; ** Significant at $\alpha = 0.01$; [f] Significant after a False Discovery Rate correction, $\alpha = 0.05$

Table 6a: Effect of Retention on Suspension, by Grade (Grades 3 – 5)

Grade	Years since summer test	Any Suspensions			Days Suspended		
3	1	-0.001	-0.008	-0.008	0.167	0.003	0.022
		(0.013)	(0.009)	(0.009)	(0.203)	(0.136)	(0.139)
	2	0.016	0.013	0.011	-0.178	0.069	0.001
		(0.015)	(0.011)	(0.010)	(0.210)	(0.160)	(0.168)
	3	0.002	-0.007	0.001	-0.333	-0.303	-0.281
		(0.022)	(0.016)	(0.016)	(0.477)	(0.339)	(0.343)
4	1	-0.011	0.014	0.000	0.059	0.227	0.120
		(0.030)	(0.022)	(0.023)	(0.242)	(0.218)	(0.214)
	2	-0.090*	-0.024	-0.052	-0.730	-0.250	-0.246
		(0.037)	(0.028)	(0.030)	(0.700)	(0.533)	(0.529)
	3	-0.020	-0.041	-0.062	-2.530	-0.900	-2.320
		(0.051)	(0.039)	(0.040)	(1.650)	(1.231)	(1.366)
5	1	-0.039	-0.045*	-0.044*	-0.142	-0.766*	-0.710
		(0.025)	(0.019)	(0.018)	(0.503)	(0.380)	(0.369)
	2	-0.050	-0.005	-0.012	-0.873	-0.377	-0.543
		(0.031)	(0.022)	(0.019)	(1.022)	(0.711)	(0.695)
	3	-0.008	0.032	0.020	-1.307	-0.612	-1.040
		(0.038)	(0.029)	(0.029)	(1.442)	(1.029)	(0.995)
Controls		Y	Y	Y	Y	Y	Y
Bandwidth		0.5	1	Varies	0.5	1	Varies
Specification		2^{nd} order polynomial	2^{nd} order polynomial	CCT Local Linear	2^{nd} order polynomial	2^{nd} order polynomial	CCT Local Linear

Notes: Entries are IV estimates of the effect of retention on a given outcome with standard errors in parentheses (adjusted for clustering at the running variable level in the parametric models). Results in columns 1, 2, 4 and 5 are 2SLS estimates from a parametric model that uses a quadratic function of the running variable, where the function is allowed to have different coefficients on either side of the Level 2 cutoff. Results in columns 3 and 6 are IV estimates generated by the local linear regression estimator proposed by Calonico et al. (2014). See Appendix Table 1 for outcome means by grade.
* Significant at α = 0.05; ** Significant at α = 0.01; [f] Significant after a False Discovery Rate correction, α = 0.05

Table 6b: Effect of Retention on Suspension, by Grade (Grades 6 – 8)

Grade	Years since summer test	Any Suspensions			Days Suspended		
6	1	0.009	0.003	0.005	0.549	0.336	0.384
		(0.031)	(0.024)	(0.028)	(0.970)	(0.813)	(0.908)
	2	0.030	0.018	0.019	1.097	0.142	0.752
		(0.031)	(0.024)	(0.027)	(1.264)	(1.004)	(1.116)
	3	0.080*	0.040	0.060	3.845*	1.513	2.437
		(0.037)	(0.030)	(0.032)	(1.531)	(1.179)	(1.327)
7	1	-0.027	-0.015	-0.017	0.321	0.626	0.500
		(0.034)	(0.026)	(0.027)	(0.998)	(0.826)	(0.701)
	2	0.005	0.019	0.024	2.345*	3.411**,f	3.432**
		(0.034)	(0.026)	(0.025)	(1.151)	(0.865)	(0.903)
	3	-0.001	0.012	0.004	-1.448	-0.356	-0.484
		(0.042)	(0.031)	(0.031)	(1.058)	(0.784)	(0.707)
8	1	0.012	-0.021	0.005	0.978	1.479	1.355
		(0.050)	(0.038)	(0.044)	(1.124)	(0.953)	(0.953)
	2	0.025	-0.029	-0.006	-0.296	-0.771	-0.580
		(0.049)	(0.037)	(0.039)	(1.373)	(1.054)	(1.107)
	3	0.052	0.030	0.049	1.198	0.023	0.951
		(0.049)	(0.038)	(0.044)	(1.450)	(1.047)	(1.349)
Controls		Y	Y	Y	Y	Y	Y
Bandwidth		0.5	1	Varies	0.5	1	Varies
Specification		2nd order polynomial	2nd order polynomial	CCT Local Linear	2nd order polynomial	2nd order polynomial	CCT Local Linear

Notes: Entries are IV estimates of the effect of retention on a given outcome with standard errors in parentheses (adjusted for clustering at the running variable level in the parametric models). Results in columns 1, 2, 4 and 5 are 2SLS estimates from a parametric model that uses a quadratic function of the running variable, where the function is allowed to have different coefficients on either side of the Level 2 cutoff. Results in columns 3 and 6 are IV estimates generated by the local linear regression estimator proposed by Calonico et al. (2014). See Appendix Table 1 for outcome means by grade.
* Significant at α = 0.05; ** Significant at α = 0.01; [f] Significant after a False Discovery Rate correction, α = 0.05

Table 7a: Effects of Grade Retention on Attendance, by Grade (Grades 3 – 5)

Grade	Years since summer test	Attendance Rate			Chronic Absences		
3	1	0.006	0.005	0.005	-0.046	-0.056**,f	-0.051**
		(0.004)	(0.003)	(0.003)	(0.026)	(0.019)	(0.019)
	2	0.004	0.003	0.002	-0.003	-0.010	-0.002
		(0.004)	(0.003)	(0.003)	(0.028)	(0.020)	(0.020)
	3	0.009	0.010*	0.008	-0.070*	-0.054*	-0.056*
		(0.006)	(0.004)	(0.004)	(0.034)	(0.025)	(0.025)
4	1	-0.000	-0.002	-0.001	0.005	0.029	0.024
		(0.007)	(0.006)	(0.006)	(0.049)	(0.038)	(0.037)
	2	-0.000	-0.001	-0.002	-0.025	-0.011	-0.029
		(0.009)	(0.007)	(0.007)	(0.056)	(0.042)	(0.046)
	3	0.016	0.001	0.006	-0.048	0.029	0.007
		(0.015)	(0.012)	(0.011)	(0.076)	(0.058)	(0.060)
5	1	0.015*	0.010*	0.010*	-0.033	-0.032	-0.039
		(0.006)	(0.005)	(0.004)	(0.037)	(0.027)	(0.027)
	2	0.005	0.002	0.003	0.010	-0.014	-0.019
		(0.008)	(0.006)	(0.006)	(0.040)	(0.030)	(0.028)
	3	-0.000	-0.004	0.000	0.014	0.034	0.019
		(0.011)	(0.008)	(0.008)	(0.049)	(0.036)	(0.036)
Controls		Y	Y	Y	Y	Y	Y
Bandwidth		0.5	1	Varies	0.5	1	Varies
Specification		2^{nd} order polynomial	2^{nd} order polynomial	CCT Local Linear	2^{nd} order polynomial	2^{nd} order polynomial	CCT Local Linear

Notes: Entries are IV estimates of the effect of retention on a given outcome with standard errors in parentheses (adjusted for clustering at the running variable level in the parametric models). Results in columns 1, 2, 4 and 5 are 2SLS estimates from a parametric model that uses a quadratic function of the running variable, where the function is allowed to have different coefficients on either side of the Level 2 cutoff. Results in columns 3 and 6 are IV estimates generated by the local linear regression estimator proposed by Calonico et al. (2014). See Appendix Table 1 for outcome means by grade.

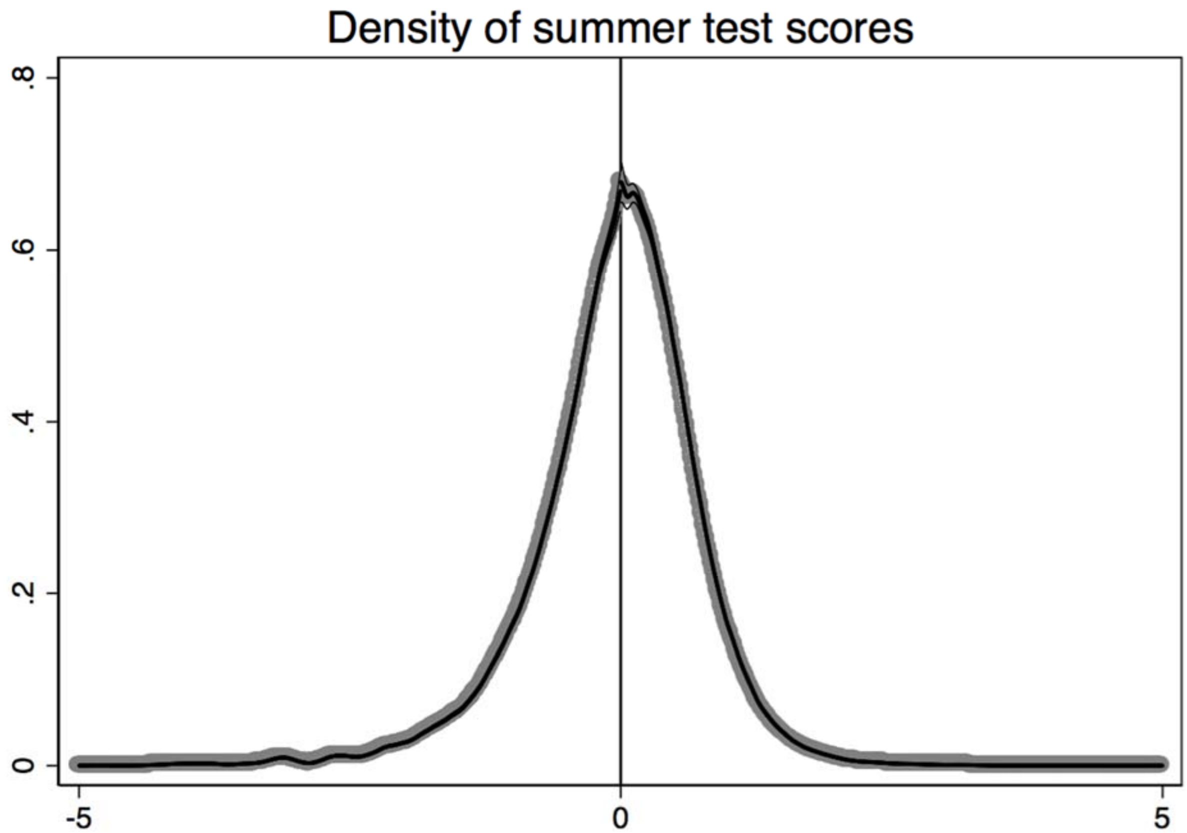
* Significant at α = 0.05; ** Significant at α = 0.01; [f] Significant after a False Discovery Rate correction, α = 0.05

Table 7b: Effects of Grade Retention on Attendance, by Grade (Grades 6 – 8)

Grade	Years since summer test	Attendance Rate			Chronic Absences		
6	1	0.006 (0.007)	0.006 (0.005)	0.009 (0.007)	-0.053 (0.035)	-0.029 (0.027)	-0.052 (0.035)
	2	-0.012 (0.009)	-0.003 (0.007)	-0.009 (0.010)	-0.072 (0.038)	-0.056 (0.030)	-0.069 (0.037)
	3	-0.019 (0.017)	-0.004 (0.013)	-0.009 (0.015)	0.022 (0.049)	0.002 (0.039)	0.013 (0.046)
7	1	0.011 (0.009)	0.017* (0.007)	0.013 (0.007)	-0.018 (0.039)	-0.048 (0.030)	-0.043 (0.032)
	2	0.047** (0.015)	0.039**,f (0.011)	0.046** (0.013)	-0.030 (0.043)	-0.026 (0.033)	-0.026 (0.031)
	3	0.032 (0.025)	0.056**,f (0.019)	0.046* (0.020)	-0.042 (0.056)	-0.069 (0.042)	-0.063 (0.043)
8	1	-0.033 (0.020)	-0.019 (0.015)	-0.036* (0.017)	0.146** (0.056)	0.126**,f (0.042)	0.130** (0.045)
	2	0.033 (0.029)	0.021 (0.022)	0.025 (0.024)	0.089 (0.061)	0.054 (0.046)	0.084 (0.054)
	3	0.017 (0.037)	-0.008 (0.028)	0.008 (0.031)	0.025 (0.066)	0.021 (0.051)	0.024 (0.055)
Controls		Y	Y	Y	Y	Y	Y
Bandwidth		0.5	1	Varies	0.5	1	Varies
Specification		2nd order polynomial	2nd order polynomial	CCT Local Linear	2nd order polynomial	2nd order polynomial	CCT Local Linear

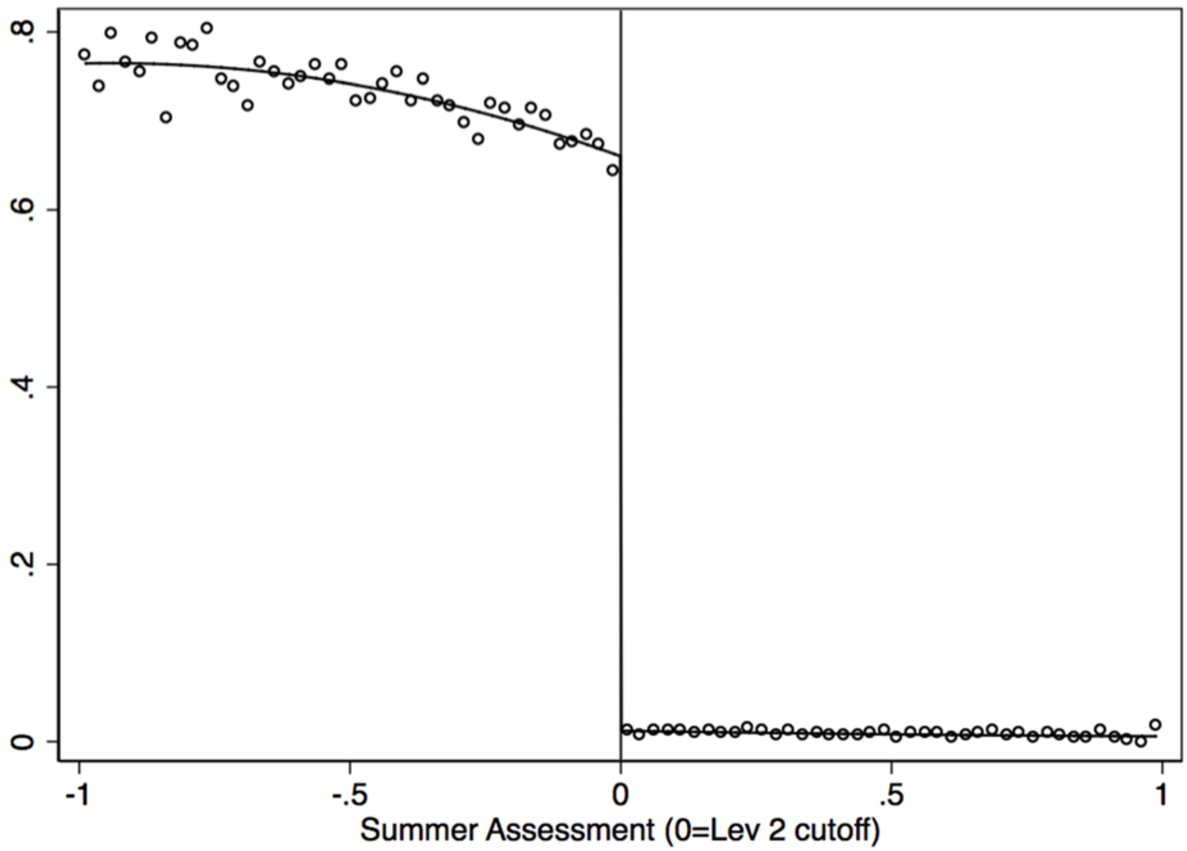
Notes: Entries are IV estimates of the effect of retention on a given outcome with standard errors in parentheses (adjusted for clustering at the running variable level in the parametric models). Results in columns 1, 2, 4 and 5 are 2SLS estimates from a parametric model that uses a quadratic function of the running variable, where the function is allowed to have different coefficients on either side of the Level 2 cutoff. Results in columns 3 and 6 are IV estimates generated by the local linear regression estimator proposed by Calonico et al. (2014). See Appendix Table 1 for outcome means by grade.
* Significant at $\alpha = 0.05$; ** Significant at $\alpha = 0.01$; [f] Significant after a False Discovery Rate correction, $\alpha = 0.05$

Figure 1: Estimated Density of the Minimum Level 2-Scaled Summer Assessment Scores



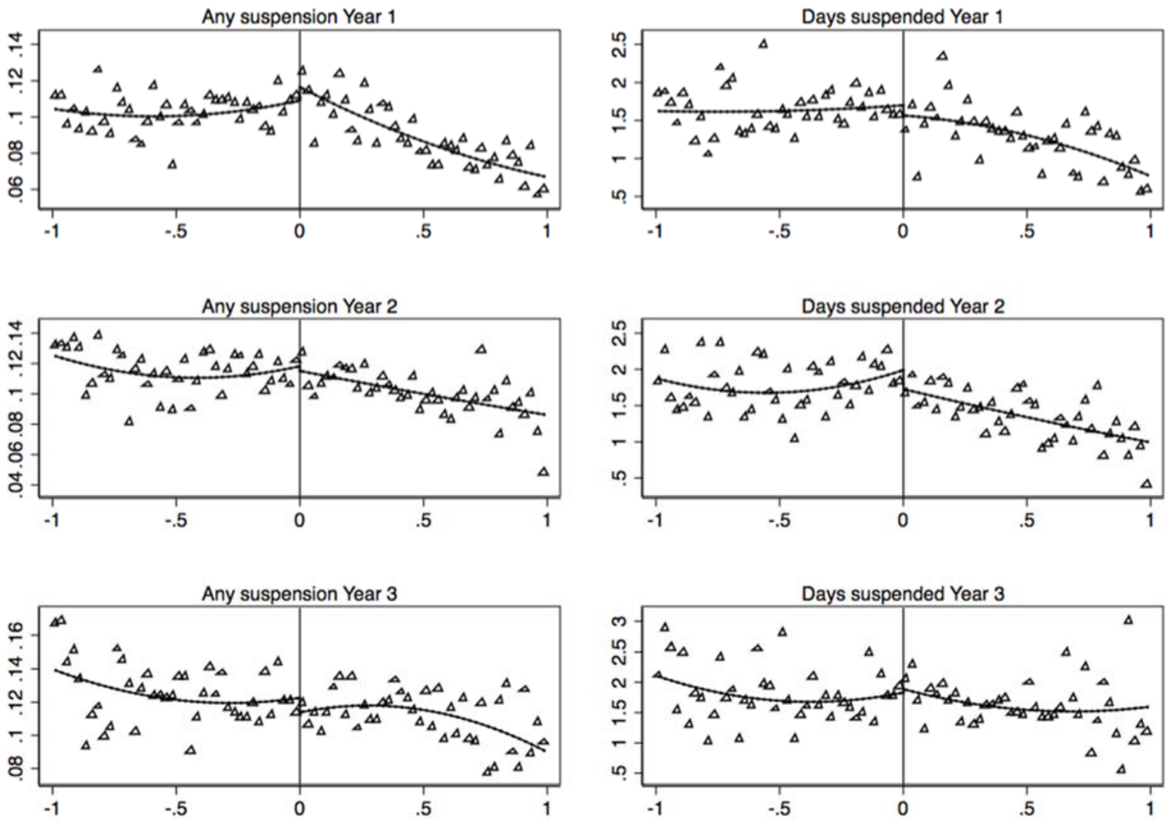
Note: Estimated density obtained using the McCrary (2008) procedure.

Figure 2: Fraction Retained in Grade by Score on Summer Assessment



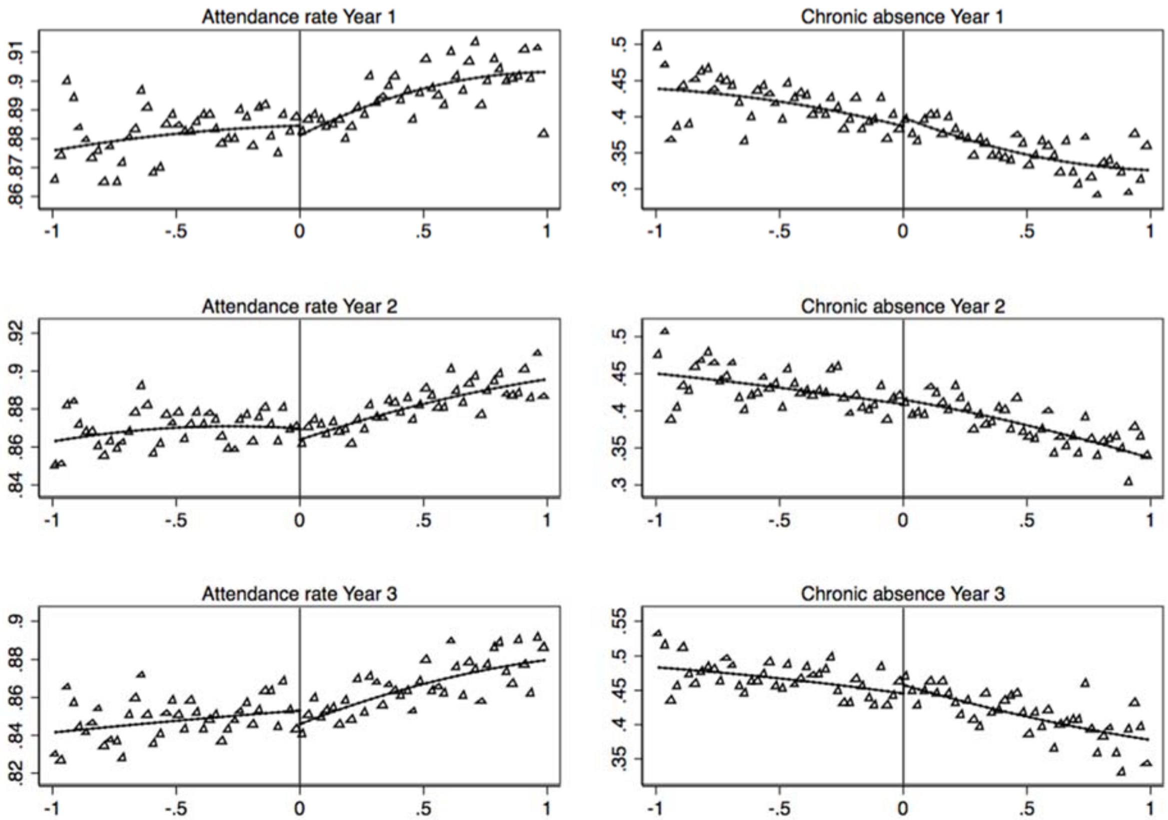
Note: Estimated for all grades and cohorts subject to the policy. Fitted line is estimated from a 2nd degree polynomial specification with bandwidth of 1.

Figure 3: Effects of Grade Retention on Suspensions, Pooled Over All Grades



Note: Estimated for all grades and cohorts subject to the policy. Fitted line is estimated from a 2^{nd} degree polynomial specification with bandwidth of 1.

Figure 4: Effects of Grade Retention on Attendance, Pooled Over All Grades



Note: Estimated for all grades and cohorts subject to the policy. Fitted line is estimated from a 2nd degree polynomial specification with bandwidth of 1.

Appendix

Table A1: Outcome Means by Grade and Year Post-Retention

Grade	Year	Any Suspension	Days Suspended	Attendance Rate	Chronic Absence
3	1	0.031	0.250	0.916	0.317
	2	0.050	0.421	0.915	0.316
	3	0.079	0.845	0.910	0.335
4	1	0.057	0.398	0.920	0.291
	2	0.080	0.759	0.915	0.309
	3	0.109	1.882	0.901	0.362
5	1	0.082	0.987	0.905	0.358
	2	0.125	1.887	0.895	0.386
	3	0.153	2.674	0.875	0.444
6	1	0.148	2.664	0.901	0.341
	2	0.147	2.949	0.884	0.402
	3	0.142	2.417	0.854	0.459
7	1	0.178	3.422	0.865	0.470
	2	0.156	2.477	0.827	0.516
	3	0.147	1.823	0.780	0.575
8	1	0.175	2.521	0.801	0.560
	2	0.157	2.182	0.752	0.603
	3	0.125	1.717	0.715	0.642