

NBER WORKING PAPER SERIES

THE EFFECT OF GRADE RETENTION ON HIGH SCHOOL COMPLETION

Brian Jacob
Lars Lefgren

Working Paper 13514
<http://www.nber.org/papers/w13514>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2007

We would like to thank the Chicago Public Schools for providing the data used in this study. We are particularly grateful to Dan Bugler, Amy Nowell and Andrea Ross for their help in securing the necessary data and to Wei Ha and JD LaRock for excellent research assistance. All remaining errors are our own. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

© 2007 by Brian Jacob and Lars Lefgren. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Effect of Grade Retention on High School Completion
Brian Jacob and Lars Lefgren
NBER Working Paper No. 13514
October 2007
JEL No. I21,I28,J01,J24

ABSTRACT

Low-achieving students in many school districts are retained in a grade in order to allow them to gain the academic or social skills that teachers believe are necessary to succeed academically. This practice is highly controversial, with many researchers claiming that it leads to higher dropout rates although selection issues have complicated previous analyses. In this paper, we use a regression discontinuity design to examine the impact of grade retention on high school completion. We find that grade retention leads to a modest increase in the probability of dropping out for older students, but has no significant effect on younger students.

Brian Jacob
Gerald R. Ford School of Public Policy
University of Michigan
735 South State Street
Ann Arbor, MI 48109
and NBER
bajacob@umich.edu

Lars Lefgren
130 Faculty Office Building
Brigham Young University
Provo, UT 84602-2363
l-lefgren@byu.edu

I. Introduction

In most developing countries and some developed countries, students are promoted from one grade to the next based on their academic performance. This was once the case in the United States. Starting in the mid-1960s, however, educators became concerned that retention (i.e., the practice of requiring students to repeat a grade) adversely impacts social, emotional and cognitive development among children, and leads at-risk students to drop out of school. Since this time, grade retention has gone in and out of vogue among educators in the United States, although it has remained relatively rare since the early 1960s.

The recent push for educational accountability, however, has brought this issue back to the forefront of education debates. In an effort to improve student achievement, many states have recently implemented policies that require elementary school children to meet explicit performance goals in order to be promoted. Indeed, nineteen states explicitly tie student promotion to performance on a state or district assessment (ECS 2000) and three of the largest school districts in the country – New York City, Los Angeles and Chicago – have similar policies. These policies are intended largely to provide incentives to students, teachers and parents, although advocates claim that retaining those students who cannot meet the promotional standards will also benefit the students themselves.

On the other hand, critics argue that retention will harm those low-achieving students most at risk of failure. They point to a vast research literature within education documenting the negative impacts of retention, particularly in terms of reducing high school completion.¹ Because

¹ In a survey of 47 empirical studies with a variety of academic achievement measures, Holmes (1989) found that retained students scored 0.19 to 0.31 standard deviations below comparable students who had not been retained. Moreover, a variety of studies have found that retention is associated with an increased likelihood of dropping out (Shultz, Toles et al. 1986, Rumberger 1987, Grissom and Shepard 1989, Fine 1991, Roderick 1994). Several more recent studies have found moderate, positive effects of retention (Karweit 1991, Pierson and Connell 1992, Alexander, Entwisle et al. 1994, Edie and Showalter 2001, Dworkin, Lorence et al. 1999).

retention decisions are typically made by the teacher or school principal on the basis of a host of unobservable student characteristics such as maturity or parental involvement, however, all of these studies are plagued by serious selection issues.

The introduction of a student accountability program in the Chicago Public Schools (CPS) provides an opportunity to circumvent these selection concerns. In 1996, the CPS instituted an accountability policy that tied promotional decisions to performance on standardized tests, resulting in a highly non-linear relationship between current achievement and the probability of being retained.² In prior work, we used the variation generated by this non-linearity to examine the effect of grade retention on short-run achievement gains. We found no consistent differences in the performance of retained versus promoted students in the short-run (Jacob and Lefgren 2004).

In this paper, we utilize a similar identification strategy to examine the long-run effects of retention on high school completion and cognitive skills. We begin by examining the effect of grade retention in sixth and eighth grade on the likelihood of dropping out of school. We find that sixth grade retention does not have a significant impact on the likelihood that a student will drop out, or the age at which a student will drop out. This suggests that students at risk of retention drop out soon after they are legally able, regardless of whether they are held back in sixth grade. On the other hand, retention in eighth grade increases the likelihood that a student will drop out by roughly 8 percentage points, or 14 percent.

To reconcile these results, it is important to realize that interventions in one period interact with policies and opportunities in subsequent periods. In particular, students who are retained in the sixth grade are significantly *less* likely to be retained in the eighth grades when they face high-

² In prior work, we have examined the potential motivational effects of these requirements and found that the policy increased achievement, particularly among older students (Jacob 2005). In this analysis, we set aside the incentives associated with the policy and instead focus on the direct academic consequences of summer school and grade retention for those students who fail to meet the promotional standards.

stakes testing again. In our sample of students scoring close to the promotional cutoffs in sixth grade, retained students began high school only about one fifth of a year after their promoted peers on average. In contrast, many students retained in the eighth grade were able to take advantage of compensatory policies to rejoin their original cohort, but still began high school over 0.4 years later than promoted students. Because students retained in earlier grades have more opportunities to catch up with their peers, retention in these grades has less severe long-term impacts on school completion than does later grade retention.

Interestingly, among students who ultimately drop out, grade retention has little impact on the age at which the students drop out, but does reduce the number of credits earned in high school. This suggests that students who face the greatest risk of dropout fail to catch up with their peers in the same way that is observed among the broader population of retained students.

The remainder of this paper is organized as follows. Section 2 provides background on the Chicago accountability policy. Section 3 describes our data and Section 4 explains our empirical strategy. Section 5 presents findings the results, and Section 6 concludes.

2. Background

In this section, we briefly describe the Chicago accountability policy and then discuss the potential impacts of the grade retention. Because of the myriad channels through which retention might operate, we argue that the theoretical impact of grade retention on student outcomes is ambiguous and is likely to vary by the age of retention.

In 1996-97, Chicago instituted a policy to end social promotion – the practice of passing students to the next grade regardless of their academic skills or school performance. Under the policy, students in third, sixth, and eighth grade were required to perform at predefined levels in

both reading and mathematics in order to be promoted to the next grade. In Spring 1997, the promotion standards for third, sixth, and eighth grade were respectively 2.8, 5.3, and 7.0 grade equivalents³, which roughly corresponded to the 15-20th percentile in the national achievement distribution.⁴ Students who did not meet the standard in the spring were required to attend a six-week summer school program, after which they could retake the exams. Those who passed the August exams moved on to the next grade. Students who again failed were required to repeat the grade.⁵

One of the most notable aspects of the Chicago accountability policy was the large fraction of students that were affected. In the first three years of the program, roughly 30-40 percent of students in grades three, six and eight failed to meet in the promotional requirements in the spring, and each year approximately 10-20 percent of students in these grades were eventually retained. In comparison, between 1993 and 1995 only one or two percent of the students in these grades were retained.

The structure of the accountability program does allow retained students to catch up to the remainder of their cohort. Students who were retained in eighth grade had an opportunity to retake the exams the following year. If they passed at this point, they were able to rejoin their original cohort in the second semester of ninth grade. Furthermore, insofar as students who were retained in third or sixth grade entered the next accountability grade (i.e., sixth or eighth grade respectively) at a higher level than their promoted counterparts, they were less likely to be retained

³ The grade equivalent scale is a nonlinear but monotonic transformation of the raw score (i.e., the number of items answered correctly), which is created such that a student at the 50th percentile in the nation scores at the eighth month of her current grade. For example, a third grader performing at the national average will score 3.8 grade equivalents. Hence, third grade students in Chicago who scored more than one grade below average were subject to retention. The ITBS is reported in units of one-tenth of a grade equivalent, with 10 potential values reported for each grade which reflect ten months of school per academic year (e.g., 3.0 to 3.9 inclusive).

⁴ The CPS has raised the promotional cutoffs several times since 1997. The eighth grade cutoff was raised to 7.2 in 1998, 7.4 in 1999 and 7.7 in 2000. The sixth grade cutoff was raised to 5.5 in 2000.

in that subsequent accountability grade. Also, retained students may be better prepared to pass their high school courses and thus accumulate high school credits at a faster pace. Finally, once students reach high school, they are able to attend summer school in order to obtain credits at a faster rate. For these reasons, the impact of grade retention may well be mitigated through a combination of optimizing behavior on the part of students and their increased capacity to overcome subsequent hurdles to graduation.

As mentioned above, in prior work we found that the retention induced by the new accountability policy did not affect student achievement in the short-run (Jacob and Lefgren 2004). A subsequent study examined dropout rates after the introduction of the Chicago accountability policy (Allensworth 2005). Focusing strictly on 8th graders, this study examined the interaction between retention per se (which was predicted to increase dropout rates) and rising achievement levels experienced by students after the introduction of the accountability policy (which was predicted to decrease dropout rates). Using data on cohorts before and after the introduction of the policy, and controlling for a series of background characteristics including latent 8th grade achievement, the authors find that grade retention under accountability was positively associated with the likelihood of dropping out, although less so than grade retention that occurred among pre-accountability cohorts.

The analysis below builds on this earlier work. In particular, we focus specifically on the impact of grade retention, and pay particular attention to the differential impact of this “treatment” across different age groups and how these policies (i.e., retention in both 6th and 8th grade) interact with each other. Moreover, following our earlier work, we utilize a regression discontinuity

⁵ Students over the age of fifteen who were retained were placed in special “transition” centers.

approach that should mitigate the potential selection biases associated with OLS comparisons, even those that control for a rich set of observable characteristics.

3. Data

This study utilizes administrative data from the Chicago Public School system. Student records provide individual level information on enrollment, demographics and achievement scores. Unique student identification numbers allow us to follow individual students throughout their tenure in the public school system. School level data provides demographic and school resource information, including the racial and socio-economic composition at the school.

The baseline sample consists of students who were (a) enrolled in the third, sixth or eighth grades and attended summer school during the first three years of the program (1997, 1998 and 1999), (b) took both the math and reading exams in the Spring, (c) were subject to the accountability policy, and (d) enrolled in the CPS the following Fall.⁶ Table 1 presents summary statistics for our sample, with columns 1, 5 and 9 describing the full baseline sample of third, sixth and eighth graders respectively. We see that students in the CPS are drawn from a highly disadvantaged and minority population. More than half of all students are black and nearly one-third of students are Hispanic. Over three quarters of the students qualify for free lunch. Of the remainder, many receive reduced price lunch. Other striking statistics include the substantial

⁶ In creating this baseline sample, we exclude roughly 9 percent of students who did not take the math or reading exam in the Spring, an additional 11 percent who were not subject to the accountability policy due to bilingual or special education placements, and an additional 2 percent who did not return to the Chicago Public Schools in the Fall immediately following the Spring exam.

fraction of students that have attended bilingual programs, are in foster care, or live with a non-parent relative.⁷

In the analysis, we focus on a subset of these students. First, we focus on the students who failed the promotional cutoffs in the Spring, and were therefore at-risk of retention. Second, we limit our analysis to those students who took both the math and reading exams in August, since this was the criteria on which the retention decision was made. Finally, because our analysis is based on a comparison of students who barely passed the promotional criteria with those who just missed passing, we limit our analysis sample to the set of students who scored within a 1.5 grade equivalent range surrounding the cutoff.⁸ As we show below, our results are robust to alternate choices of this range around the cutoff.

Our final analysis sample includes 8,573 and 5,402 sixth and eighth grade students respectively. Summary statistics for this sample is shown in columns 2, 6 and 10. Not surprisingly, the students in our analysis sample are even more disadvantaged than the overall CPS population. Students in our analysis sample are more likely to be black, qualify for free lunch, and to have performed poorly on their May tests. Within this sample, just over forty percent of students are ultimately retained. These retained students are quite similar in observables, however, to the students in our analysis sample who are ultimately promoted.

Our primary outcome measure is whether students have completed high school by Fall 2005. With the exception of the latest sixth grade cohort, we are measuring dropout status at least

⁷ The relatively low percentage of students in special education or bilingual programs is a result of our sample construction which excludes the majority of students in these programs since they are not subject to the promotional policy. The relatively small number of students in these programs who were subject to the policy are included.

⁸ Specifically, we focus on students who scored from one grade equivalent below the cutoff to one-half grade equivalent above the cutoff. The asymmetry in the ranges above and below the cutoff is due to the fact that the CPS did not perfectly enforce the grade retention policy for students scoring just below the cutoff. For this reason, the probability of retention slopes downward in a small region below the cutoff. Thus while there is a broader range below the cutoff than above, the range around this marginal region is roughly similar.

five years following the start of high school for all sixth and eighth grade students.⁹ Students who were retained as sixth graders in 1998-99, however, will be enrolled in 12th grade in Fall 2005, assuming that they have continued in school and progressed normally following their retention in sixth grade. If a large fraction of students dropout in their senior year or later, including this cohort may understate the effects we estimate. Based on the experiences of earlier cohorts, it appears that the majority of students who leave high school do so prior to their senior year. For this reason, we include the 1999 sixth grade cohort. We later show results excluding this cohort, which yield qualitatively comparable results.

4. Empirical Strategy

Identification

Prior studies have attempted to ascertain the effect of grade retention by estimating some variant of the following basic regression:

$$(1) \quad Y_i = \beta \text{retain}_i + \Gamma X_i + u_i + \varepsilon_i$$

where Y is the outcome, X is a vector of demographic and past performance variables, and retain is a binary variable that takes on a value of one if a student is retained and zero otherwise, u represents unobserved (to the researcher) student ability, ε is an error term. However, if students are selected for retention on the basis of factors that are unobservable to the researcher and also influence educational outcomes, such as maturity or parental involvement, then β is likely to be biased.

We address these selection concerns by using a regression discontinuity design (RDD) to identify the effects of grade retention. By tying student promotion to performance on standardized

⁹ The third graders in our sample have progressed far enough in school to examine dropout rates.

tests, the Chicago accountability policy created a highly non-linear relationship between a student's current achievement and her probability of moving on to the next grade. In particular, students who scored below predetermined cutoffs in either reading or math at the end of the school year were assigned to attend summer school. Those who failed to achieve these cutoffs at the end of summer school were required, in principle, to repeat the grade.

August test scores, therefore, became the gatekeeper for determining promotion. Figure 1 shows the relationship between the August reading score (relative to the cutoff) and the probability of grade retention for students who passed math and failed reading prior to summer school on the initial May exam. The graphs are based on the three cohorts of students included in the analysis – students who attended these grades between 1997 and 1999, who took the Spring math and reading exams, and were included under the accountability policy. The solid line maps the predicted probability of retention from a lowess regression with a smoothing factor of .2. The circles reflect the actual average probability of retention for each August test score group, where the size of the circles reflects the number of students in the group. Had the policy been strictly enforced, all students below the cutoff would have been retained. In practice, however, some students just below the reading and math cutoffs were promoted. Additionally, a few students who scored above the cutoffs were retained for other reasons (e.g., failure to attend classes). Imperfect compliance to the stated state policy led to the non-linear, though not discontinuous, relationship that we observe in the figures.

Three distinct regions are apparent in the graphs. Above the cutoff, nearly all the students were promoted. Between the cutoff and .3 or .4 grade equivalents below, students had a positive but declining probability of grade retention. We will refer to this as the marginal area. Finally, nearly all students scoring worse than .3 or .4 grade equivalents below the cutoff were retained.

Roughly 24 percent of students who scored below the cutoffs at the end of summer school were promoted, while 3 percent of students who scored above the cutoffs were retained. If we limit the sample to those close to the cutoff, roughly 29 percent who scored below the cutoffs at the end of summer school were promoted, while 3 percent of students who scored above the cutoffs were retained.

The imperfect adherence to the retention policy makes it impossible to implement a strict regression discontinuity design (RDD). Under such a strategy, we would estimate the causal impact of grade retention by simply comparing the outcomes of those students just below the cutoff, all of whom were retained, to the outcomes of those just above the cutoffs, who were promoted. In lieu of a strict RDD, however, we can take advantage of the nonlinearity between the August scores and retention probability to employ a “fuzzy” regression discontinuity design. This amounts to a two-stage least squares procedure in which we instrument grade retention with a nonlinear function of the August test scores. In the second stage we control linearly (or with a polynomial) for the same test scores. Identification arises from differences in the functional form between the first and second stages.

This identification strategy relies on two fundamental assumptions. First, this approach assumes that we have appropriately accounted for the underlying relationship between August test scores and our outcomes of interest. Even in the absence of the retention policy, student achievement scores are clearly associated with student outcomes such as high school completion. Since our instruments are nonlinear functions of August test scores, we must control adequately for these same scores in our second stage models. If our choice of functional form is inadequate, our instruments will be correlated to the residual and our strategy will yield biased treatment effect estimates. This assumption is most likely to be satisfied if we focus our analysis on a sample of

individuals close to the cutoff, where the relationship between test-scores and the probability of dropout can be well approximated with a linear or low-order polynomial function.

Second, this strategy assumes that once we have controlled for continuous measures of August achievement, the nonlinear measures we use as instruments are not correlated with any unobserved student characteristics that could independently influence the outcome of interest. This assumption is closely related to the first. In most RD designs, the specific concern here is that participants can strategically influence the measure that determines treatment. McEwan and Urquiola (2005), for example, show that more advantaged schools in Chile were able to intentionally manipulate their grade level enrollments so that they were just below the number at which a new classroom would be required in order to avoid the expenditures associated with an additional teacher. The result was a sharp discontinuity in unobservable school characteristics at exactly the cutoff that one might use to examine the effect of class size on student achievement, which may violate this identifying assumption.

In our case, there is little reason to believe that students had the ability or incentive to precisely manipulate their test scores. Teacher cheating, however, may be a possible avenue of such manipulation (Jacob and Levitt 2003). In order to test for the presence of such manipulation, Figure A1 shows kernel density plots of the August test scores. If such manipulation occurred, we would expect to find a discontinuity in the density function near the cutoff. For example, if teachers edited student answer keys to ensure that the children were not retained, we might expect to find an unusually large fraction of students scoring just above the promotional cutoff. We find

no unusual discontinuity immediately above or below the cutoff, suggesting that no such manipulation occurred.¹⁰

While Figure 1 focuses on the individuals who passed math in May and thus were only required to pass reading in order to be promoted, our data actually offers three separate experiments that can be used for estimating the impact of grade retention on the dropout rate:

1. Individuals who passed math in May and needed only to pass reading in August to be promoted. In this case, the index that determines treatment assignment is the August reading score.
2. Individuals who passed reading in May and needed only to pass math in August to be promoted. In this case, the index that determines treatment assignment is the August math score. Figure 2 shows the relationship between August math scores and the likelihood of grade retention for this sample. The relationship closely mirrors that seen for Group 1, the sample of students who failed reading.

¹⁰ While we are not relying on a sharp RDD that involves the comparison of groups immediately below and above the cutoffs, it is helpful to understand the magnitude of differences in skill that are suggested by students scoring at different ranges on the exams. The reading exam focuses explicitly on comprehension and consists of 36, 44 and 50 questions for third, sixth and eighth grade respectively. In the area relevant for students at risk of retention (i.e., students scoring between the 10th and 20th percentile on the national distribution), there is a one-to-one correspondence between grade equivalents and raw scores but not all GE are possible. On the 1997 exam, for example, eighth grade students answering 16 items correctly scored a 6.7 GE while students answering 17 items correctly scored a 6.9 GE. This means that students who scored just above the reading cutoff correctly answered just 1 or 2 items more than their peers who scored in the marginal area, though this reflects a difference of roughly one-quarter to one-half a year's worth of learning. The math exam is composed of three different subsections, each of which are normed separately and then aggregated so that the total math score is based on between 110-160 items, and there is not always a one-to-one mapping between total raw score and total grade equivalent score. At the same time, this means that scores are the math exam are likely a more precise measure of a student's underlying skill. In the area relevant for students at risk of retention, a change in 0.1 grade equivalents corresponds to two raw points for third graders and roughly 1.5 raw points for sixth and eighth graders. Students scoring just above the cutoff answered approximately 4-8 more items correctly than students scoring in the middle of the marginal area, which also translates to one-quarter to one-half of a year's worth of learning. Unlike some standardized exams, the ITBS grade equivalent (GE) scores are not created using item response techniques (IRT), but are simply a transformation of the total raw score. For this reason, the GE metric does not take into account that a correct response on a more difficult item may reflect greater knowledge than a correct response on a less difficult item. Given information on the underlying item responses for each student, it would be possible to construct an IRT score that more truly reflects a student's underlying ability. Using this measure, one might be able to create more precise RDD estimates by, for example, comparing students with the same IRT score who scored just above and below the GE-based cutoff.

3. Individuals who failed both reading and math in May and thus need to pass both subjects in August for promotion. For this group, both math and reading scores in August determine promotion. Figure 3 shows the relationship between August math and reading scores and the likelihood of retention for this sample. As described below, we use a non-linear function of August scores in both subjects to instrument for grade retention.

Though these groups differ in the particulars of how assignment to grade retention is made, the intuition underlying our identification strategy is identical – we will exploit the sharply nonlinear relationship between August test scores and retention controlling in a flexible manner for the underlying relationship between achievement scores and the outcome of interest.

Estimation

For those students in group 1 who passed math but failed reading in May, we model the likelihood of retention as:

$$(2) \quad \text{retain} = \gamma_1 \text{passed reading} + \gamma_2 \text{marginal reading} + f(\text{reading}, \text{math}) + \Gamma X + \eta,$$

where $f(\text{reading}, \text{math})$ is a smooth function of the student’s August reading and math scores, X is a vector of student demographic characteristics including the student’s May math and reading scores, and η is an error term. The exogenous instruments in the equation are dummy variables that indicate whether a student’s August reading score was at or above the cutoff (“passed reading”) or whether the student’s August score fell just below the cutoff in what was described as the marginal area (“marginal reading”).¹¹

¹¹ We achieve the virtually identical results if we construct instruments that allow the probability of grade retention to fall *within* the marginal area. Doing so, however, greatly multiplies the number of instruments used without providing any measurable improvement in statistical efficiency.

For those students in group 2 who passed reading but failed math in May, we estimate a nearly identical first stage equation where the exogenous instruments are dummy variables for scoring in the passing or marginal area for math in August.

$$(3) \quad retain = \gamma_1 \text{passed math} + \gamma_2 \text{marginal math} + f(\text{reading}, \text{math}) + \Gamma X + \eta ,$$

Finally, for those students in group 3, the likelihood of retention is determined by both reading and math scores in August. While it is possible to specify this non-linearity in multiple ways, a simple set of variables indicating whether the student passed or was marginal in each subject captures the majority of the variation quite well. For this reason, we estimate the following first-stage equation:

$$(4) \quad \begin{aligned} retain = & \gamma_1 \text{passed both} + \gamma_2 \text{passed reading, marginal math} \\ & + \gamma_3 \text{passed math, marginal reading} + \gamma_4 \text{marginal both} , \\ & + f(\text{reading}, \text{math}) + \Gamma X + \eta \end{aligned}$$

The exogenous instruments in this equation correspond to the different areas in Figure 3, which maps each instrument into a location on the plane of August reading and math scores. “Passed both” corresponds to area E, “passed reading, marginal math” to B, “passed math, marginal reading” to C, and “marginal both” corresponds to area D. The omitted category is area A in which students failed both reading and math.

For the purposes of increasing our precision, we stack the data and estimate a single treatment effect using variation from all three groups. We later show the estimates separately by group to explore whether the effect of retention differs. To estimate the full model on the stacked data, we run a fully interacted regression model that allows the coefficients on all of the covariates and on the August test scores to vary by group. We constrain only the coefficient on grade retention to be the same across groups. In the first stage, this approach is equivalent to running completely separate specifications for each group. Our second stage, however, takes the form

$$(5) \quad Y = \beta_1 \text{retain} + f(\text{reading}, \text{math}) + \lambda_1 \text{group} + BX + \Gamma W + \varepsilon,$$

where Y is the academic outcome of interest, $group$ is a set of dummy variables that index the whether the student failed reading, failed math or failed both math and reading in May, X are student demographics and W is a vector that includes the full set of interactions between X or f and $group$.

Interpretation

Given the nature of our identification strategy, it is particularly useful to discuss the interpretation of the treatment effects we estimate. Imbens and Angrist (1994) emphasize the importance of the Local Average Treatment Effect (LATE), which they define as the average effect of an intervention on those individuals who were induced to participate on account of variation in the instruments. In our analysis, the instruments induced students who scored just below the cutoff to be retained. Thus, our estimates reflect the treatment effect for those individuals who received treatment because they scored in the marginal area of reading and/or math, which corresponded to roughly the 12-14 percentiles in math and 20-22 percentiles in reading. Since grade retention is generally targeted at very low-achieving students, our estimates should be particularly relevant for policymakers.¹²

Given the likelihood of heterogeneous treatment effects, it is unlikely that our estimates capture either the Average Treatment Effect (ATE) or even the effect of Treatment-on-the-Treated

¹² It is also worthwhile pointing out that identification of the retention effect also depends on the fact that the treatment effect does not vary over the range of reading performance below the cutoff that we use in estimation. If this treatment effect varies greatly with reading performance below the cutoff, we will be unable to identify the baseline relationship between current and future performance. This is because the relationship between current and future performance below the cutoff will reflect the effect of reading performance on the efficacy of treatment. In practice, this is likely a fairly weak assumption because the range of data we use to estimate the retention effect is still quite small, generally less than + or - 1 grade equivalent from the cutoff. If this assumption holds, the estimated coefficient corresponds to the LATE for those individuals who received treatment because they scored a given distance below the cutoff.

(TT). Unless one is willing to assume that the treatment effect is constant or does not vary according to prior reading and math ability, one cannot use our estimates to say what would have occurred if all students had been treated.¹³ In the following sections, we examine the heterogeneity of treatment effects across a variety of observable student characteristics such as race, gender, SES and prior achievement.

Several additional features of the Chicago program should also be considered when interpreting the estimates. First, because the cutoffs were binding for a large number of children, the social promotion policy changed the peer ability distribution for retained and promoted students, particularly in schools where large numbers of students were held back. To the extent that improving peer ability levels benefits students who are promoted, one might expect our estimates to overstate the negative effects of retention associated with a more modest program. Although even this is unclear, since students who barely passed the cutoff and were promoted are likely to be substantially behind their peers and may struggle if the teacher is able to move more quickly through class material. Second, the district provided some additional financial resources to help retained students. While these allocations were modest, it is possible that they could have improved student outcomes. Third and perhaps most importantly, under the Chicago accountability system, retained students had multiple mechanisms through which to catch up with their peers. While this policy may well differ from the informal retention policies used in many schools in the past, it is likely quite similar to the large scale accountability programs being implement in cities such as New York and Los Angeles.

¹³ Similarly, if social stigma is a factor in retention, it may have been mitigated in this context by the large number of retentions that occurred. Retaining a small number, even those with roughly the same achievement levels, may have a different impact.

5. Results

In this section, we discuss the effects of retention on school completion and academic achievement. To begin, we examine the impact of sixth and eighth grade retention on the likelihood that a student completes high school. In addition to examining dropout rates, we explore the age at which the student left school and the number of credits she accumulated prior to leaving. Finally, we examine the long-run effects of retention on academic achievement.

The Impact of Retention on High School Completion

To begin, it is useful to examine the simple OLS estimates of the relationship between grade retention and the likelihood of dropping out of high school. To maximize comparability across designs, we limit the OLS analysis to the RDD sample described above. Note that the outcome variable takes on a value of 1 if the student dropped out of the CPS and a value of zero if the student graduated from the CPS. For the relatively small percent of students who left the CPS prior to graduating or dropping out, we have set this outcome to missing.¹⁴ Of course, if retention influences attrition from the CPS, this may bias our estimates. We show that this is not a concern in the section below. We also included dummy variables indicating whether the observation came from the “failed reading,” “failed math,” or “failed both” samples.

The OLS estimates are shown in Table 2. In columns (1) and (4), we present estimates from that include no controls. The resulting point estimates imply that grade retention is associated with a 3 percentage point increase in the probability of dropping out for students retained in the sixth grade and an 8 percentage point increase for students retained in the eighth grade. Given the

¹⁴ All students in our sample should have graduated by the Fall 2005. In practice, there are a handful of students who appear to still be enrolled in the CPS even after 5 or 6 years in high school. We have kept these observations in the data, but excluding them does not change our results.

baseline probability of grade retention for sixth and eighth grade promoted students in the sample of 53 and 57 percent respectively, these estimates suggest that grade retention increases the likelihood of dropping out by 6 to 14 percent.

In columns (2) and (3), we redo the analysis adding linear controls for the students' August reading and math scores. For sixth grade students, the point estimate on grade retention falls to 0.00. For eighth grade students, the point estimate falls by forty percent to 0.05. This suggests that even for the relatively homogeneous group of students we examine in the RDD sample, prior performance is significantly correlated with the likelihood of retention as well as the probability of dropping out. Interestingly, the inclusion of a host of other demographic and prior achievement controls in columns (3) and (6) does not significantly change the point estimates.

The OLS analysis suggests that observable student characteristics are correlated with both the probability of being retained and of dropping out of high school. To the extent that unobserved characteristics are similarly correlated, the OLS estimates are likely biased. We address this concern by estimating the RDD models described above.

Table 3 shows the results from our first stage regression. In the interests of brevity, we show only the coefficients on our instruments, though we control for August and earlier math and reading scores as well as the full set of demographic characteristics and prior achievement measures described earlier. All of the instruments are individually significant. As expected, for students who failed reading and passed math in May, scoring in the marginal area in reading in August is associated with large reductions (23 to 54 percentage points) in the risk of grade retention. For students in this group who score at or above the reading cutoff in August, the reduction in the probability of grade retention is even larger (between 68 and 87 percentage points). Examining the coefficients for the students who failed math or failed both reading and math in

May, we see very similar results. Overall, the table suggests that our instruments have strong power for predicting the probability of grade retention. The F-statistics of our instruments are 138.31 for sixth graders and 367.14 for eighth graders, suggesting that there is little cause to worry that our estimates suffer from bias associated with weak instruments. The larger F-statistics for the eighth grade students is attributable to closer compliance on the part of the school district to the stated policy for this group.

Table 4 shows the second stage RDD estimates. For comparison, column (1) and (4) show our OLS estimates with the full set of covariates. Columns (2) and (5) show the RDD estimate in which we only control for linear measures of the August test scores. Note that we include interactions to allow the effect of August math and reading scores to differ by summer school group. For example, the relationship between August math scores and the future probability of dropping out may be different for the failed reading group relative to the failed math group if the former did not take the August math test seriously. Similarly, one might suspect that the relationship between August scores and the likelihood of dropping out differs for the group that failed both reading and math in May since this group was noticeably lower achieving than either the failed math only or failed reading only groups. The final columns show the RDD estimates that control for past test scores as well student demographics. The comparison between the RDD estimates with and without covariates provides a useful check on the validity of our instruments. If the exclusion restriction is met, once we control for smooth measures of August test scores, the non-linear terms of August test scores that we use as instruments should not be correlated with any unobserved factors that determine our outcome of interest. If our estimates change significantly

with the inclusion of other covariates, this would suggest that our instruments may be picking up some factors that determine the likelihood of dropping out independent of retention.¹⁵

Consider first the results for the sixth graders. The point estimate shown in column (2) is actually negative, though very small in absolute value at -0.02. Adding the covariates in column (3) yields a point estimate of -0.04, which is not significantly different from the prior specification. This provides evidence to support our identifying assumption. The standard error on the estimates in column (3) implies a confidence interval between -.12 to .04. Given the baseline dropout rate of 53 percent, we can reject that grade retention increases the drop out rate by more than 8 percent. It therefore seems plausible that grade retention in the sixth grade has little substantive effect on the probability a student drops out.

Finally, it is perhaps worth noting that while the difference between the estimates in column 1 and 3 are not statistically significant at conventional levels, the direction of the change suggests that the OLS estimates may be biased upward. Given the formulaic nature of the retention policy, this must come about as a result of teachers and administrators either (a) providing waivers to students who have unobservable good attributes or (b) retaining students for reasons other than test scores (recall that a small number of students were retained for poor attendance or grades). Another possible explanation for the difference between the OLS and the RDD estimates involves the heterogeneity in effects. The RDD estimates reflect the impact of grade retention on those students who are retained as a result of scoring just below the marginal area on the August exam, whereas the OLS estimate captures the average effect for all students within the sample.

¹⁵ Of course, the true concern is that our instruments are correlated with unobservable factors. However, to the extent that the instruments are correlated with the observable covariates, one might be concerned that they are also correlated with unobserved determinants of dropping out.

For eighth graders, the picture is somewhat different. The RDD estimates shown in column (6) suggests that grade retention increases the probability of dropping out by 8 percentage points, or roughly 14 percent. Again the specification is largely invariant to the set of covariates that are included. It thus appears that eighth grade retention is more harmful than sixth grade retention, possibly due to the greater social dislocation caused by preventing students from moving to high school with their peers.

As described above, if grade retention influences the propensity to leave the CPS, the dropout estimates above may be biased by differential attrition. Table 5 examines this possibility. We show RDD estimates of the impact of grade retention on the likelihood that a student will move to another public school district or transfer to a private school. In our sample, roughly 10-15 percent of promoted students switch public school districts and 2-3 percent transfer to a private school. The RDD estimates are uniformly small and statistically insignificant, suggesting that differential attrition is not a concern in the analysis. In addition, results presented in Appendix Table A1 demonstrate that our main findings are not sensitive to a variety of alternative specifications, including the addition of second and third order polynomials or changes in the area around the cutoff we use for the estimation.

Table 6 examines the heterogeneity of effects across a variety of subgroups. Several interesting facts emerge. First, it appears that there was little effect of retention for the 1997 cohort, the first group of children to experience the policy. Second, for the eighth graders, it appears that the largest impact of retention occurred among African-American girls who failed both reading and math in May. To the extent that students who failed both reading and math were noticeably lower achieving than those students who failed only one subject, these results suggest that retention is most harmful to the very lowest achieving students in the system. The

disproportionate effect on girls is interesting as well. Girls have substantially lower dropout rates than boys across the ability distribution in general, and in our RDD sample as well. For example, in our sample of low-achieving students, 49 percent of girls who are promoted in eighth grade drop out compared with 66 percent of promoted boys. The point estimates suggest that grade retention has virtually no effect on the likelihood that boys will drop out of school. Hence, the retention effect of 16 percentage points for girls implies that retention is simply moving girls up to the dropout likelihood of boys. Finally, the fact that retention seems to affect African-American students more dramatically than Hispanic students is interesting, although we do not have any good explanation for it at this point.

Understanding the Dynamic Impacts of Grade Retention on Academic Progression

While the analysis above provides a reduced form estimate of the impact of grade retention on high school completion, it is important to consider why the effect sizes may differ for retention in the sixth and eighth grades. In addition, by examining the impact of grade retention on academic progression, we can gain some insight into how retention interacts with the subsequent educational environment to mitigate possible adverse effects. For these reasons, we examine the following additional questions: (1) How does grade retention in the sixth grade affect the probability of retention in the eighth grade? (2) How does grade retention in the sixth and eighth grade affect the age that students begin high school? (3) What is impact of grade retention on the age of high school completion? (4) What is the effect of retention on the age of dropout? (5) What impact does retention have on the number of credits a student has obtained at the time she leaves school, either by graduating or by dropping out?

Table 7 presents some estimates that address these questions. In the first panel, we see that retention in the sixth grade is associated with a 9 percentage point decline in the probability of retention. This suggests that students held back in an early grade are indeed more prepared to be successful when they ultimately advance to subsequent grades. Examining the next column, we see that students retained in the sixth grade begin high school only 0.21 years later than similar students who are promoted. This suggests that over three years, we observe massive convergence of the retained students with their original cohort. This is due both to the existence of multiple accountability grades and to the fact that retained students were offered an opportunity to rejoin their cohort by passing an exam in the following January (one semester into their retained year). Given the rapid rate of convergence, it is perhaps unsurprising that we ultimately observe little impact of sixth grade retention on total credits accumulated or the age that students leave school.

Moving down to the next panel, we can examine how grade retention in the eighth grade affects the progression of children through school. Interestingly, we see that retained students still begin high school only 0.42 years later than promoted students. This suggests that the majority of students who are retained in the eighth grade manage to rejoin their cohort before the next academic year. This is entirely due to the opportunity provided by the January retest. Despite the substantial convergence of retained students, retained students ultimately leave school at roughly the same time as their promoted peers, having earned significantly fewer credits.

The next panel focuses on those students who ultimately complete high school. Here we are essentially splitting the sample on an endogenous (i.e., outcome) variable. Because retention has no impact on high school completion for sixth graders, this does not introduce any complications. When looking at the eighth grade result where we do find a modest impact of retention on dropout rates, however, one should bear in mind that the results may be subject to

selection bias due to the impact of grade retention on the fraction of students who actually complete high school.

The impact of grade retention on the age when such students begin high school is similar to that of the overall sample. For students retained in the sixth grade, those who ultimately graduate do so approximately one-fifth of a year later than their promoted counterparts. This corresponds closely to the delay they experience entering high school, suggesting that virtually no convergence occurs after they actually begin high school. For students retained in the eighth grade who ultimately graduate, they do appear to make up some time in high school—perhaps via summer school classes. They graduate, however, with about five percent fewer credits than their promoted counterparts.

The final panels of the table examine the academic progression of students retained in the sixth and eighth grades who ultimately drop out from high school. Examining those students retained in the sixth grade, we see that their high school experience is delayed by about the same amount as the overall sample. Furthermore, we see that there is little impact either on the total credits accumulated or on the age when students drop out. These results are similar for students retained in the eighth grade—grade retention appears to have little impact on either on the number of credits accumulated or upon the age of dropout.

The Impact of Retention on Cognitive Skill

Given that the effects of grade retention are only modest on school completion and years attended, it is natural to examine whether grade retention affects the rate of learning. When considering how grade retention influences academic achievement, it is critical to consider the counterfactual. It is possible to compare retained and promoted students at the same grade when

retained students will be one year older, at the same age when retained students will be in a lower grade, or at the time of when students complete or withdraw from school. The proper comparison depends, of course, on the inference one hopes to draw. Most educators focus on the same grade comparisons, consistent with the view that retention is intended to provide children extra time to catch up with their peers. Such comparisons, however, do not tell us necessarily whether grade retention increases the *rate* of learning or the ultimate amount of knowledge a student acquires at the time she leaves (through graduation or dropout) high school. For our purposes, same age comparisons are more useful as they indicate the impact of grade retention on the rate of learning. Performance on a high school exit exam if universally administered would also be helpful as they have the potential to measure the impact of grade retention on the total amount of knowledge accumulated.

For practical reasons, however, we are limited to same grade comparisons. In particular, the incentives faced by test-takers in eighth grade (with high-stakes testing) are much greater than is true in the seventh grade. Thus any same-age comparisons made for elementary school students are likely to be biased. In high school, same age comparisons are impossible because standardized tests are administered only in the eleventh grade. For these reasons we perform same grade comparisons, bearing in mind that retained students will likely be older at the time they take the exam.

The first panel of Table 8 examines students who faced grade retention in the third grade and were subsequently tested in the eighth grade. Examining the panel, we see that grade retention had no impact on the probability that students actually took the exam. The fact that only 70 percent of students who faced grade retention in the third grade took the eighth grade exam reflects the mobility of the student population rather than dropout behavior. While grade retention has no

impact on the probability that students are tested, it does increase the age at which students take the eighth grade exam. In particular, students retained in the third grade are just over a quarter of a year older when they take the eighth grade exam. If retained students had no opportunity to catch up with their peers and promoted students had no opportunity to fall behind at a later date, we would expect retained students to be a full year older when taking the exam. The fact that retained students are only a quarter of a year older indicates that, over the course of five years, most of the retained students had the opportunity to catch up with their peers. Examining the achievement effects, we see that retained students are performing at the same or higher level when they reach the eighth grade than are promoted students.

The bottom two panels present estimates of sixth and eighth grade retention on 11th grade test scores. For both grades, we see that retained and promoted students were equally likely to take the exam, but that retained students were slightly older at the time they were tested. As with the third graders, the fact that the coefficients are considerably less than one suggests that retained students were largely able to catch up with their peers. Perhaps not surprisingly, therefore, the differences in achievement are quite small, and generally not significantly different than zero.

6. Conclusions

Summarizing our findings, grade retention in the sixth grade appears to have few negative long term consequences. In particular, retained students are no more likely to drop out than are promoted students. Additionally, those students who ultimately drop out do so at the same time regardless of whether or not they were retained. Retained students who complete high school attend school for about 0.2 years longer than their promoted peers.

In contrast, grade retention in the eighth grade reduces the probability that a student completes high school by 8 percentage points or about 14 percent. This suggests that grade retention late in a student's academic career can have a significant impact on academic attainment. Examining the impacts of grade retention separately on those students who ultimately drop out, we find no evidence that grade retention affects the timing of dropout. For students who ultimately graduate, grade retention significantly increases the total amount of time spent in elementary and secondary school. While interesting, these last two findings should be viewed as merely suggestive for the 8th graders due to concerns regarding selection bias.

To reconcile the difference between the sixth and eighth grade retention effects, we examine how retention in these grades affects subsequent academic progression. In the Chicago Public Schools, there exist multiple mechanisms for retained students to catch up with their original cohort. Because sixth graders have more opportunities to converge with their cohort, they are able to effectively mitigate the long-term impacts of grade retention.¹⁶ For example, by being retained in the sixth grade, they are less likely to be retained in the eighth. Retained sixth graders start high school only one-fifth of a year after their promoted colleagues. Students retained in the eighth grade, on the other hand, have fewer chances to catch up with their promoted colleagues, begin high school later, suffer more in terms of high school completion.

Our findings underscore the need of understanding educational interventions in the broader context of subsequent interventions and optimizing behavior on the part of students, teachers, and schools. An intervention in one period affects eligibility for subsequent interventions and may change student incentives in ways that may attenuate any negative long term consequences. The importance of institutional framework in understanding the long term effects of grade retention

¹⁶ This assumes that the social stigma associated with retention is not the primary negative effect in this context.

suggest that our findings may not generalize to all settings. They are likely useful, however, for understanding the impact of grade retention in other large districts with similar accountability policies, including New York and Los Angeles.

References

- Allensworth, Elaine M. (2005). "Dropout Rates After High-Stakes Testing in Elementary School: A Study of the Contradictory Effects of Chicago's Efforts to End Social Promotion." *Educational Evaluation and Policy Analysis* 27(4): 341-364.
- Alexander, K. L., D. R. Entwisle, et al. (1994). *On the Success of Failure: A Reassessment of the Effects of Retention in the Primary Grades*. New York, University of Cambridge Press.
- Dworkin, A. G., J. Lorence, et al. (1999). *Elementary School Retention and Social Promotion in Texas: An Assessment of Students who Failed the Reading Section of the TAAS*. Houston, University of Houston.
- ECS (2000). *ECS State Notes*, Education Commission of the States (www.ecs.org).
- Eide, E. R. and M. H. Showalter (2001). "The Effect of Grade Retention on Educational and Labor Market Outcomes." *Economics of Education Review* 20: 563-576.
- Fine, M. (1991). *Framing Dropouts: Notes on the Politics of an Urban Public High School*. Albany, N.Y., SUNY Press.
- Grissom, J. B. and L. A. Shepard (1989). *Repeating and Dropping Out of School. Flunking Grades: Research and Policies on Retention*. L. A. Shepard and M. L. Smith. New York, The Falmer Press: 34-63.
- Hahn, J., P. Todd, et al. (1999). *Evaluating the Effect of an Anti-Discrimination Law Using a Regression-Discontinuity Design*. NBER Working Paper #7131. Cambridge, MA.
- Holmes, C. T. (1989). *Grade Level Retention Effects: A Meta-Analysis of Research Studies. Flunking Grades: Research and Policies on Retention*. L. A. Shepard and M. L. Smith. New York, The Falmer Press: 16-33.
- Hoxby, C. M. (2000). "The Effects of Class Size on Student Achievement: New Evidence from Population Variation." *Quarterly Journal of Economics* 115 (4): 1239-1285.
- Imbens, G. W. and J. D. Angrist (1994). "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2): 467-475.
- Jacob, B. and Levitt, S. (2003). "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *Quarterly Journal of Economics*. 118(3): 843-877.
- Jacob, B. and Lefgren, L. (2004). "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis." *Review of Economics and Statistics*. LXXXVI (1): 226-244.

Jacob, B. (2005). "Accountability, Incentives and Behavior: Evidence from School Reform in Chicago." *Journal of Public Economics*. 89(5-6): 761-796.

Karweit, N. L. (1991). *Repeating a Grade: Time to Grow or Denial of Opportunity*. Baltimore, Maryland, Center for Research on Effective Schooling for Disadvantaged Students.

Katz, L. F. and K. M. Murphy (1992). "Changes in Relative Wages 1963-1987: Supply and Demand Factors." *Quarterly Journal of Economics* 107 (1): 35-78.

Krueger, A. B. (1999). "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics* 114:497-532.

McEwan, Patrick J. and Miguel Urquiola (2005). "Non-random sorting around breaks in the regression-discontinuity design: Evidence from class size reduction." Working paper.

Murnane, R. J., J. B. Willet, and F. Levy (1995). "The Growing Importance of Cognitive Skills in Wage Determination." *The Review of Economics and Statistics* 77 (2):251-266.

Pierson, L. H. and J. P. Connell (1992). "Effect of Grade Retention on Self-System Processes, School Engagement and Academic Performance." *Journal of Educational Psychology* 84: 300-307.

Roderick, M. (1994). "Grade Retention and School Dropout: Investigating the Association." *American Educational Research Journal* 31(4): 729-759.

Rumberger, R. W. (1987). "High School Dropouts: A Review of Issues and Evidence." *Review of Educational Research* 57: 101-121.

Schulz, E. M., R. Toles, et al. (1986). *Association of Dropout Rates with Student Attributes*. American Educational Research Association, San Francisco, CA.

Shepard, L. A. and M. L. Smith (1989). *Flunking Grades: Research and Policies on Retention*. New York, N.Y., The Falmer Press.

Figure 1: The relationship between August test scores and retention among students who failed the reading exam and passed the math exam in May

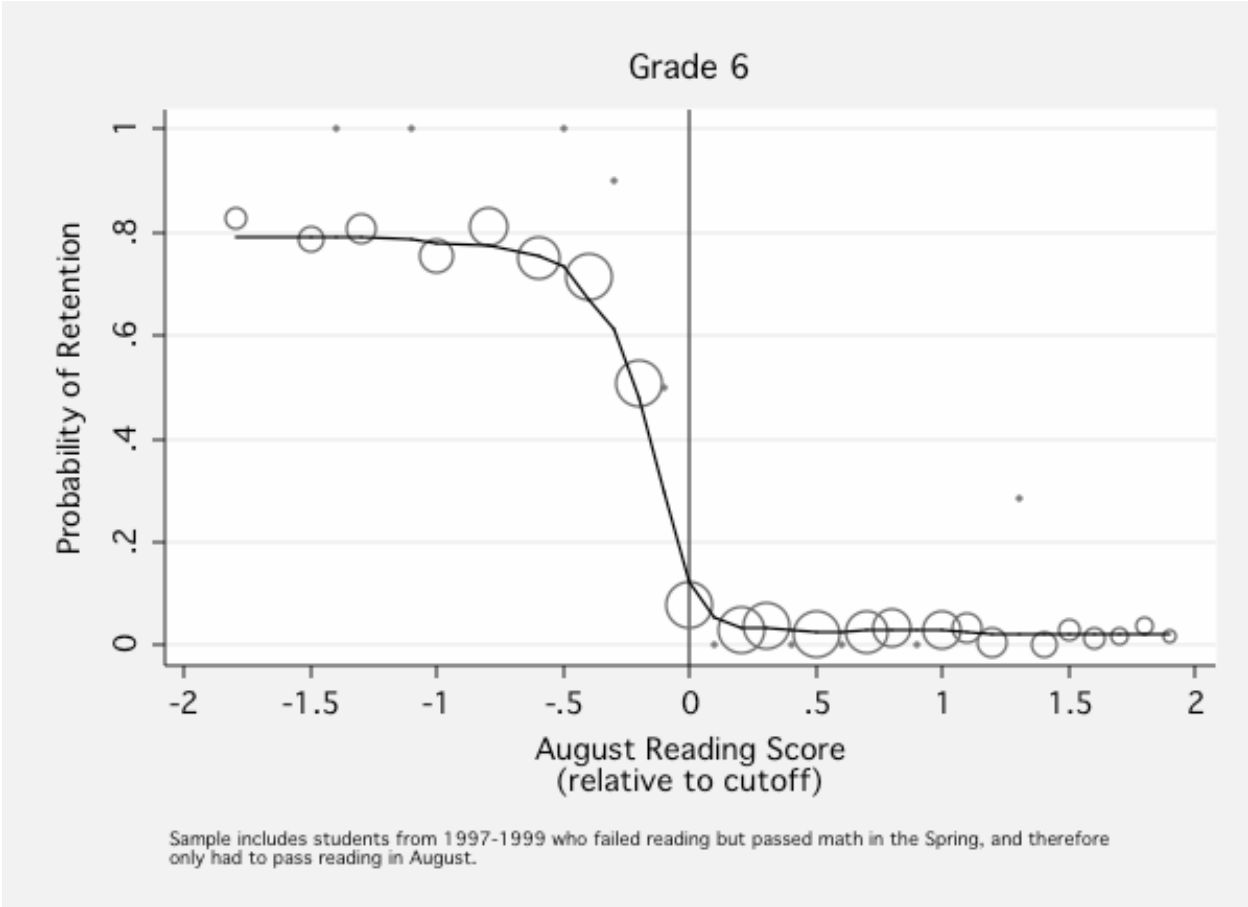


Figure 1 (continued): The relationship between August test scores and retention among students who failed the reading exam and passed the math exam in May

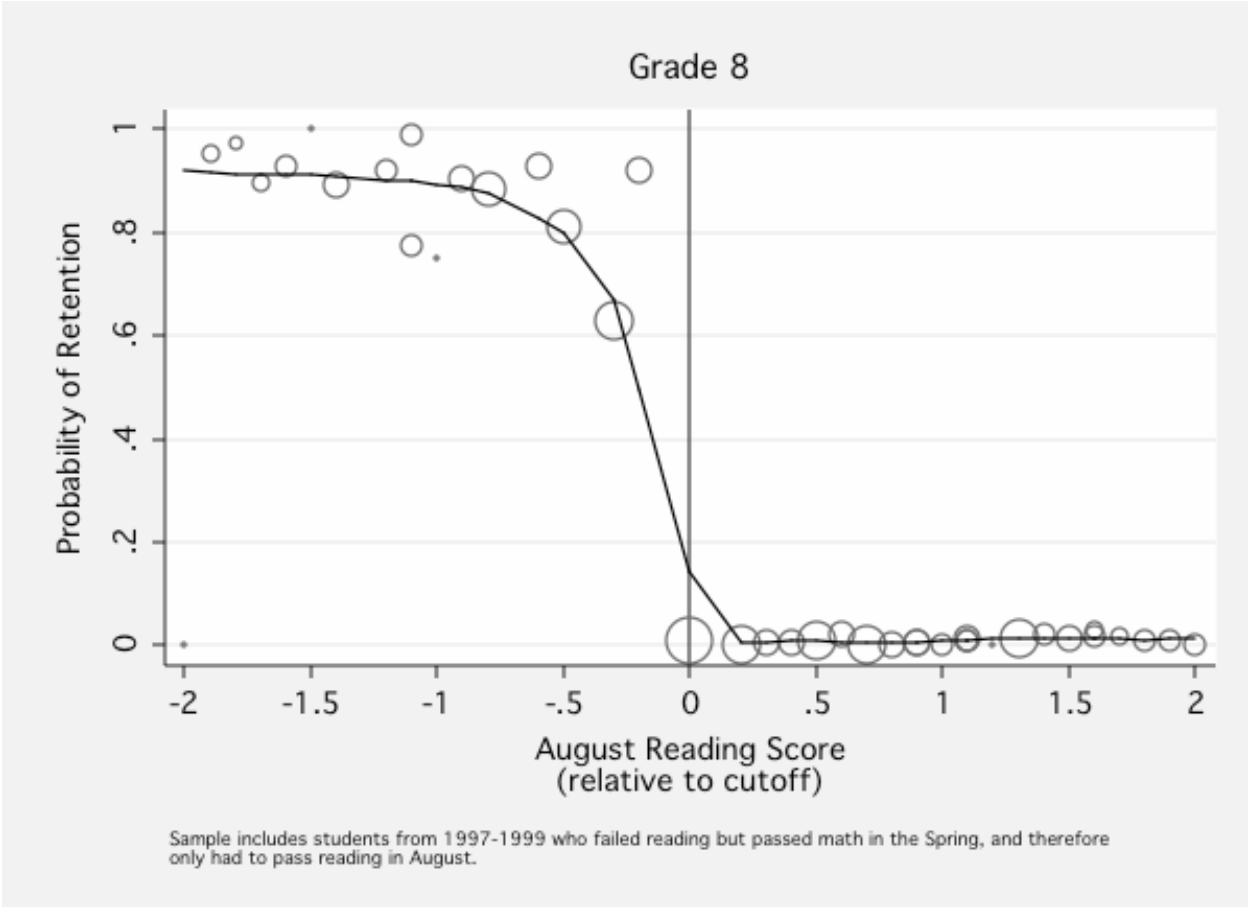


Figure 2: The relationship between August test scores and retention among students who failed the math exam and passed the reading exam in May

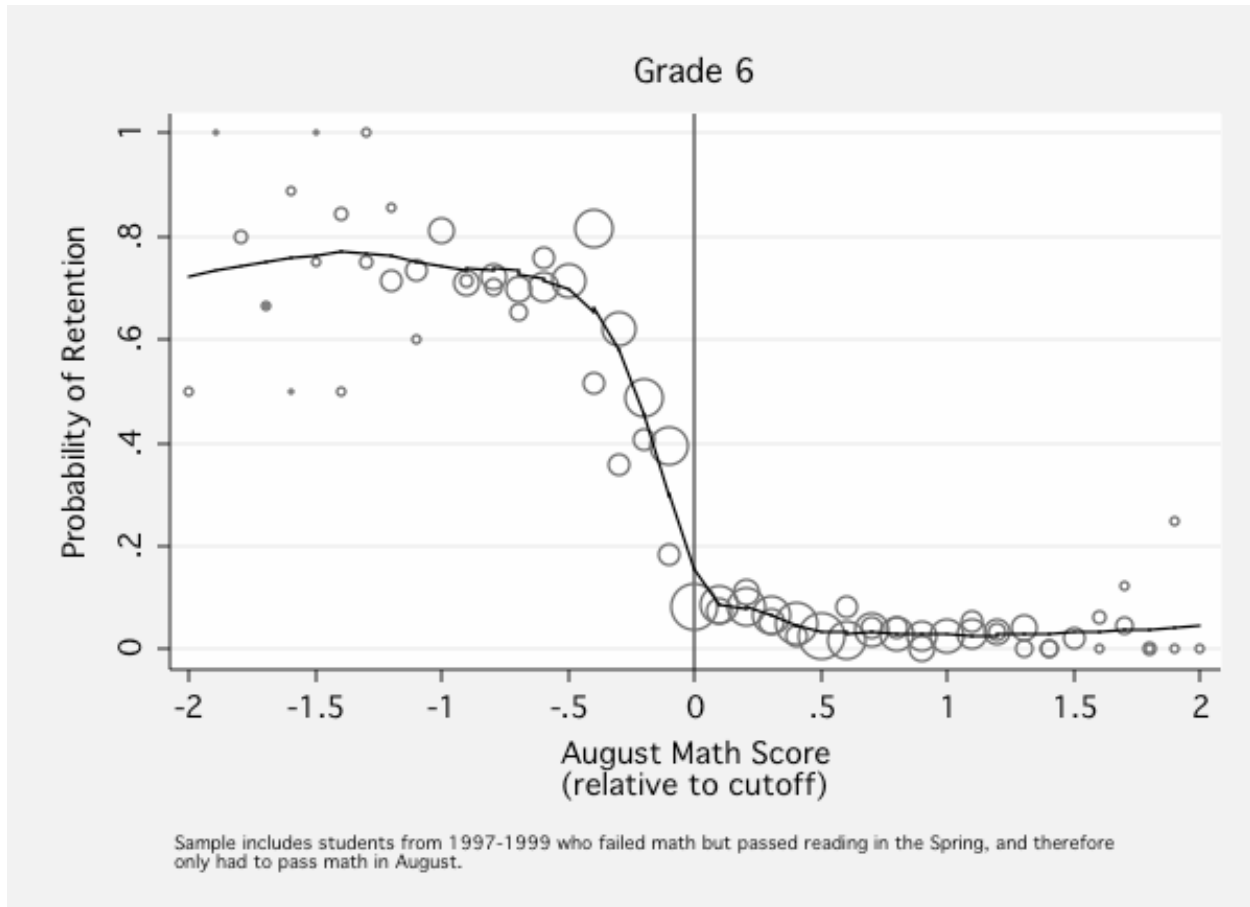


Figure 2 (continued): The relationship between August test scores and retention among students who failed the math exam and passed the reading exam in May

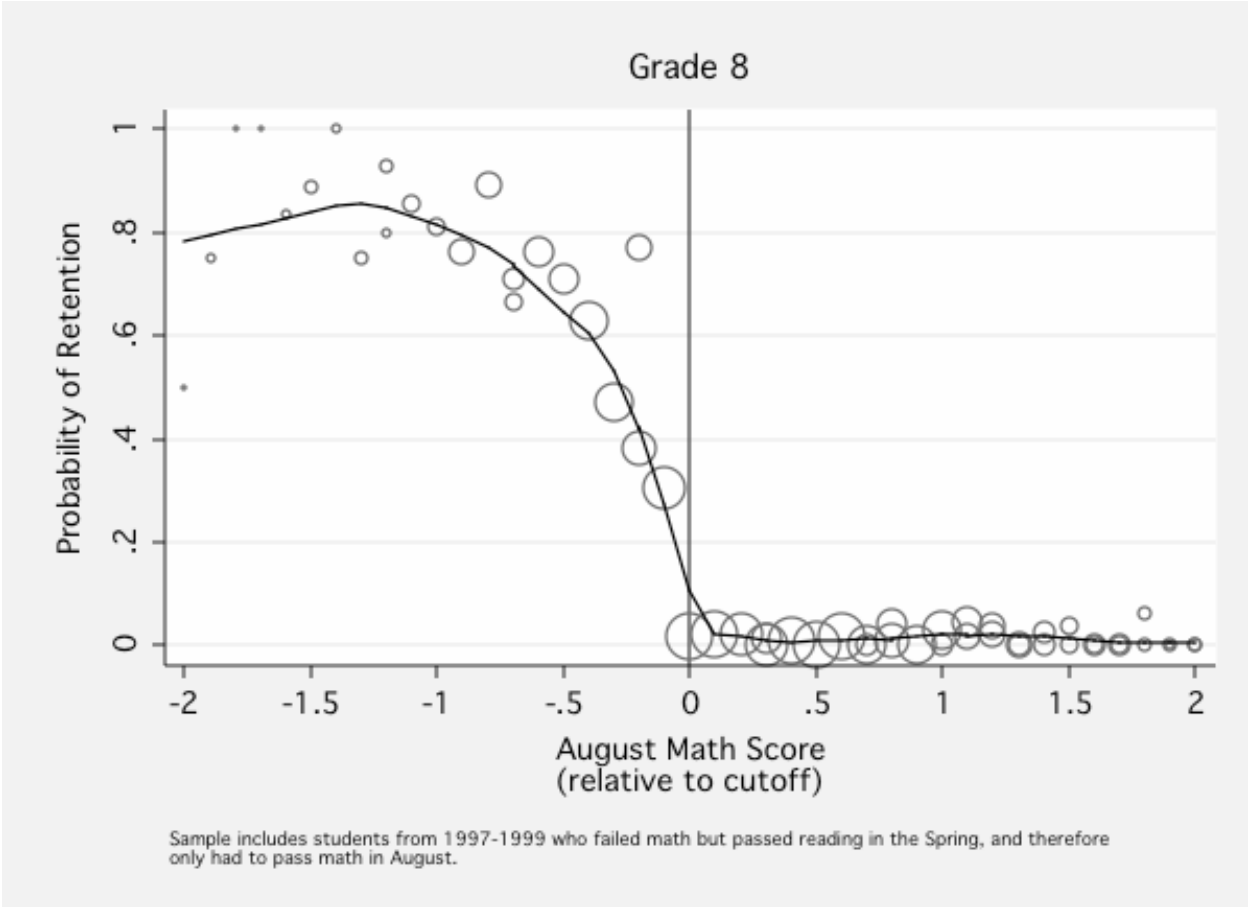


Figure 3: The relationship between August test scores and retention among students who failed both the reading and math exam in May

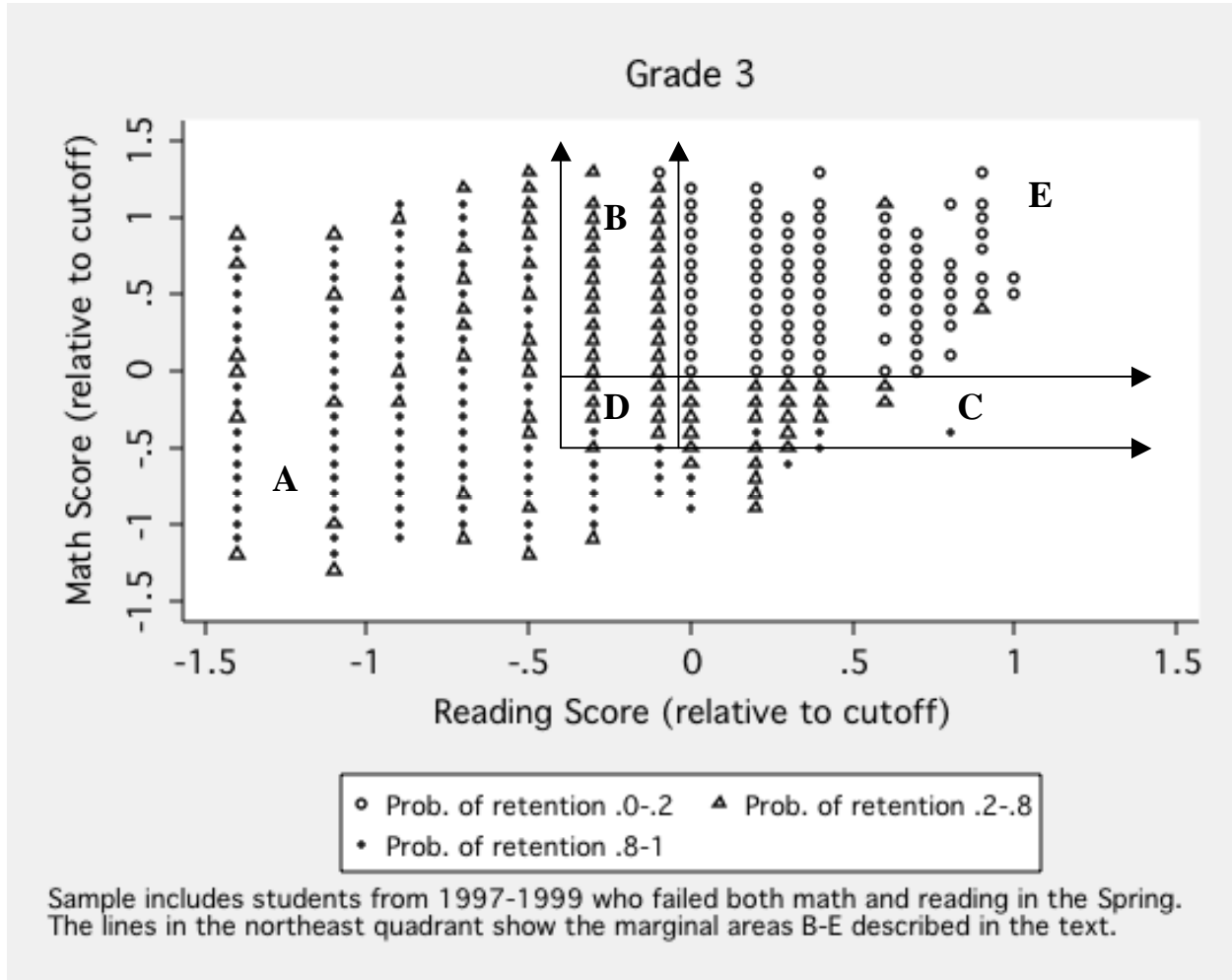


Figure 3 (continued): The relationship between August test scores and retention among students who failed both the reading and math exam in May

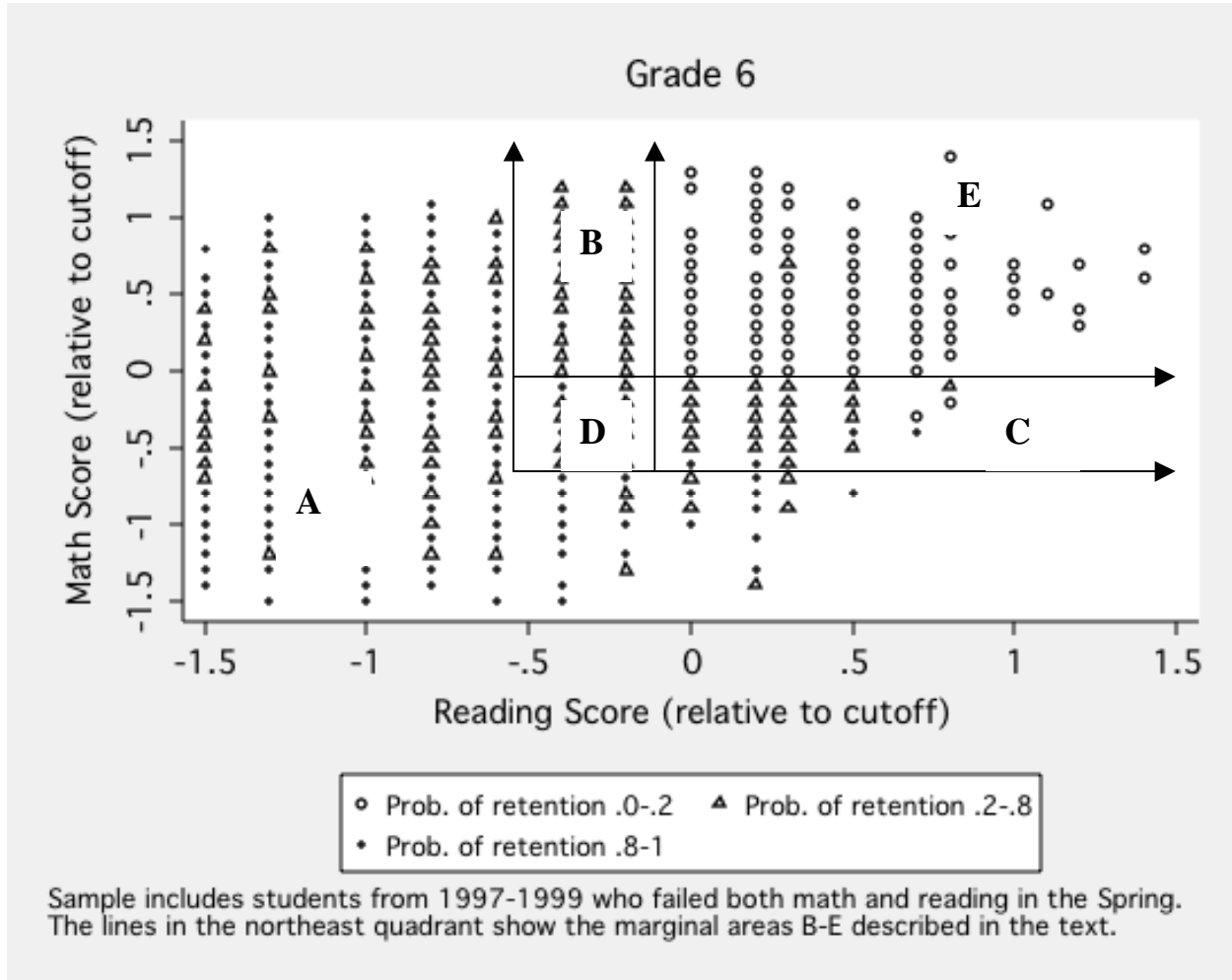


Figure 3 (continued): The relationship between August test scores and retention among students who failed both the reading and math exam in May

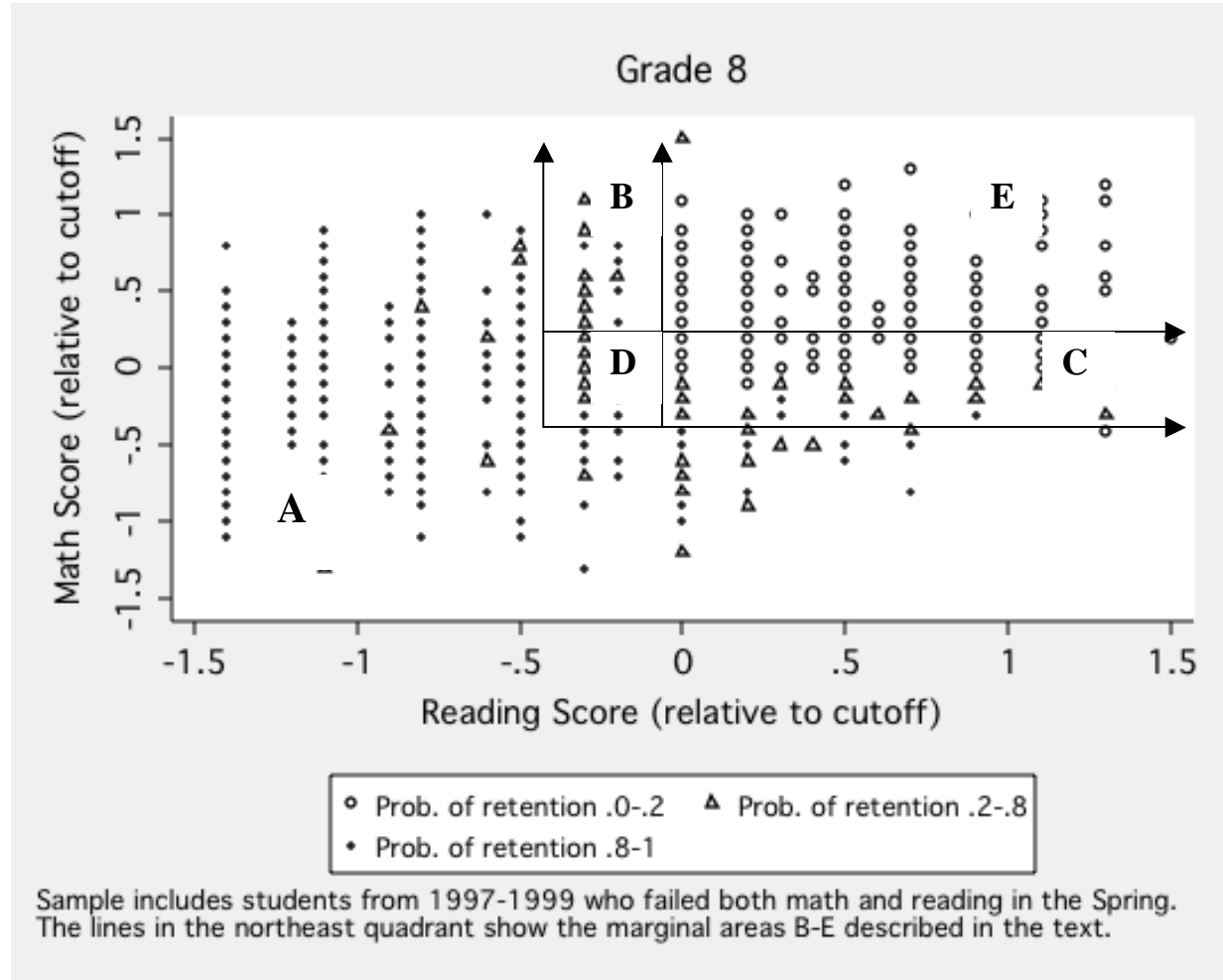


Table 1: Summary Statistics

	3 rd Grade				6 th Grade				8 th Grade			
	Full Sample	RD Sample	Retained	Promoted	Full Sample	RD Sample	Retained	Promoted	Full Sample	RD Sample	Retained	Promoted
<i>Background Characteristics</i>												
Black	0.714	0.822	0.834	0.812	0.553	0.666	0.677	0.657	0.563	0.655	0.660	0.652
Hispanic	0.173	0.137	0.131	0.142	0.319	0.299	0.293	0.303	0.319	0.312	0.313	0.312
Male	0.490	0.526	0.541	0.512	0.480	0.510	0.510	0.510	0.468	0.483	0.499	0.472
Eligible for free lunch	0.808	0.912	0.925	0.899	0.776	0.902	0.910	0.896	0.753	0.854	0.869	0.843
Eligible for reduced-price lunch	0.075	0.044	0.036	0.052	0.088	0.047	0.042	0.050	0.090	0.051	0.040	0.059
Currently in bilingual program	0.023	0.022	0.022	0.022	0.113	0.201	0.209	0.194	0.090	0.174	0.206	0.152
In bilingual program in the past	0.136	0.088	0.082	0.092	0.228	0.099	0.086	0.109	0.269	0.141	0.105	0.166
In special education program	0.033	0.036	0.034	0.037	0.019	0.021	0.024	0.019	0.010	0.013	0.012	0.013
Lives in foster care	0.067	0.089	0.090	0.087	0.045	0.064	0.068	0.062	0.041	0.055	0.051	0.057
Lives with relative other than parent	0.112	0.084	0.078	0.090	0.121	0.102	0.099	0.104	0.092	0.077	0.072	0.080
Social status of neighborhood (composite)	-0.270	-0.390	-0.410	-0.372	-0.279	-0.441	-0.449	-0.436	-0.287	-0.420	-0.468	-0.386
Neighborhood poverty index (composite)	0.354	0.530	0.569	0.494	0.223	0.431	0.459	0.411	0.231	0.381	0.426	0.349
Age	9.380	9.427	9.384	9.466	12.354	12.460	12.420	12.489	14.333	14.437	14.462	14.420
1997 Cohort	0.306	0.326	0.264	0.382	0.325	0.351	0.319	0.374	0.324	0.320	0.349	0.300
1998 Cohort	0.351	0.352	0.393	0.314	0.345	0.325	0.339	0.315	0.349	0.357	0.402	0.325
1999 Cohort	0.342	0.323	0.343	0.305	0.330	0.324	0.342	0.311	0.326	0.323	0.249	0.375
Failed reading only in May	0.200	0.544	0.453	0.626	0.147	0.521	0.445	0.577	0.127	0.442	0.405	0.469
Failed math only in May	0.044	0.099	0.052	0.142	0.060	0.216	0.172	0.248	0.072	0.295	0.187	0.371
Failed both math and reading in May	0.185	0.357	0.495	0.232	0.104	0.263	0.382	0.175	0.110	0.263	0.409	0.160
Retained	0.210	0.477	1.000	0.000	0.127	0.423	1.000	0.000	0.122	0.414	1.000	0.000

Promoted	0.790	0.523	0.000	1.000	0.873	0.577	0.000	1.000	0.878	0.586	0.000	1.000
May Reading Score	3.114	2.218	2.069	2.353	6.121	4.649	4.500	4.759	8.136	6.572	6.264	6.791
May Math Score	3.527	2.876	2.732	3.006	6.503	5.379	5.187	5.520	8.310	7.135	6.961	7.257
August reading score	2.566	2.556	2.312	2.778	5.257	5.042	4.747	5.259	7.213	6.965	6.604	7.221
August math score	3.315	3.227	2.990	3.442	5.938	5.737	5.466	5.937	7.670	7.423	7.181	7.595
<i>Educational attainment as of Fall 2005</i>												
Graduated from CPS	--	--	--	--	0.538	0.356	0.292	0.404	0.576	0.379	0.320	0.421
Dropped out	--	--	--	--	0.404	0.547	0.573	0.528	0.421	0.616	0.676	0.574
Transferred to private school	--	--	--	--	0.052	0.035	0.036	0.035	0.023	0.025	0.027	0.023
Moved to another public school district	--	--	--	--	0.162	0.157	0.159	0.155	0.103	0.115	0.119	0.113
Still enrolled in the CPS	--	--	--	--	0.058	0.097	0.135	0.069	0.003	0.005	0.004	0.006
Age when dropped out	--	--	--	--	16.924	17.039	17.038	17.040	17.310	17.318	17.326	17.311
Number of credits completed in high school before dropping out	--	--	--	--	86.912	81.110	74.663	86.217	101.884	91.292	75.874	103.997
Observations	73,134	15,424	7,355	8,069	73,015	10,611	4,492	6,119	60,279	6,281	2,603	3,678

Notes: The sample in columns 1 and 5 contains all students who were enrolled in the appropriate grades in the Spring of 1997, 1998 and 1999, were included under the accountability policy, and took the May math and reading exam. The sample in columns 2 and 6 is limited to those students who failed the math and/or reading exam(s) in May, attended summer school, took the August math and reading exams, and were either retained or promoted the following year. The samples in columns 3, 4, 7 and 8 are subsets of the samples in columns 2 and 6 respectively.

Table 2: OLS Estimates of Effect of Grade Retention on Probability of Dropping Out

<i>Specification</i>	Sixth Grade			Eighth Grade		
	(1)	(2)	(3)	(4)	(5)	(6)
Retained in Grade	0.03** (0.01)	0.00 (0.01)	0.02 (0.01)	0.08** (0.01)	0.05** (0.02)	0.04** (0.02)
August Reading and Math Controls	No	Yes	Yes	No	Yes	Yes
Other Covariates	No	No	Yes	No	No	Yes
Control group mean	0.53	0.53	0.53	0.57	0.57	0.57
R-Squared	0.03	0.04	0.10	0.03	0.05	0.13
Observations	8,573	8,573	8,573	5,402	5,402	5,402

Notes for Table 2: Sample includes students in the 1997-1999 cohorts who were included under the accountability policy, failed the math, reading or math and reading exams in May, attended summer school and took the August math and reading exams and were either retained or promoted the following year. All models include controls indicating which group the student was in based on her May exam scores: failed math only, failed reading only, or failed both math and reading. Huber-white standard errors are shown in parenthesis. *=significant at 10% level. **=significant at 5% level.

Table 3: First Stage Regressions of Grade Retention on Instruments.

	Specification	
	Sixth Grade	Eighth Grade
<i>Failed Reading in May</i>		
Marginal Reading	-0.23** (0.02)	-0.54** (0.03)
Passed Reading	-0.68** (0.03)	-0.87** (0.02)
<i>Failed Math in May</i>		
Marginal Math	-0.26** (0.03)	-0.51** (0.03)
Passed Math	-0.59** (0.05)	-0.60** (0.03)
<i>Failed Both in May</i>		
Marginal Reading/ Marginal Math	-0.21** (0.04)	-0.31** (0.06)
Marginal Reading/ Passed Math	-0.28** (0.03)	-0.66** (0.05)
Passed Reading/ Marginal Math	-0.32** (0.03)	-0.66** (0.04)
Passed Reading/ Passed Math	-0.71** (0.03)	-0.78** (0.02)
Other Control Variables	Yes	Yes
F-Statistic of Instruments	138.31	367.14
[p-value]	[0.00]	[0.00]
R-Squared	0.47	0.71
Observations	8,573	5,402

Notes: Sample includes all students in the RDD analysis sample – that is, students in the 1997-1999 cohorts who were included under the accountability policy, failed the math, reading or math and reading exams in May, attended summer school and took the August math and reading exams and were either retained or promoted the following year. Other control variables included in the models are: indicators for the summer school group (i.e., failed reading only, failed math only, failed reading and math), linear measures of August and May math and reading scores, prior achievement measures, student demographics and a set of interactions that allow the effect of all of the covariates mentioned above to vary by summer school group. Huber-white standard errors are shown in parenthesis.

*=significant at 10% level. **=significant at 5% level.

Table 4: IV Estimates of Effect of Grade Retention on Probability of Dropping Out

<i>Specification</i>	Sixth Grade			Eighth Grade		
	(1)	(2)	(3)	(4)	(5)	(6)
Retained in Grade	0.02 (0.01)	-0.02 (0.04)	-0.04 (0.04)	0.04** (0.02)	0.09** (0.02)	0.08** (0.03)
August Reading and Math Controls	Yes	Yes	Yes	Yes	Yes	Yes
Other Covariates	Yes	No	Yes	Yes	No	Yes
Estimation Method	OLS	IV	IV	OLS	IV	IV
Control group mean	0.53	0.53	0.53	0.57	0.57	0.57
R-Squared	0.10	0.04	0.10	0.13	0.03	0.05
Observations	8,573	8,573	8,573	5,402	5,402	5,402

Notes: Sample includes all students in the RDD analysis sample – that is, students in the 1997-1999 cohorts who were included under the accountability policy, failed the math, reading or math and reading exams in May, attended summer school and took the August math and reading exams and were either retained or promoted the following year. Columns 1 and 4 replicate the results from Table 3 columns 3 and 6 respectively. The specifications shown in columns 2 and 5 also include controls indicating the summer school group (i.e., failed math only, failed reaing only, or failed both math and reading) as well as a set of interactions that allow the impact of August math and reading scores and year fixed effects to differ by summer school group. The specifications shown in columns 3 and 6 include all of the covariates contained in columns 2 and 5, but also include a host of prior achievement and student demographic characteristics described in the text.. Huber-white standard errors are shown in parenthesis. *=significant at 10% level. **=significant at 5% level.

Table 5: IV Estimates of Grade Retention on Changing Public School Districts or Attending Private School

Variable	Specification					
	Sixth Grade			Eighth Grade		
<i>Dependent Variable</i>	Changed public school district	Transferred to private school	Changed public school district or transferred to private school	Changed public school district	Transferred to private school	Changed public school district or transferred to private school
Retained in Grade	-0.03 (0.03)	-0.00 (0.01)	-0.03 (0.03)	-0.02 (0.02)	0.00 (0.01)	-0.02 (0.02)
August Scores and Other Controls	Yes	Yes	Yes	Yes	Yes	Yes
Estimation Method	IV	IV	IV	IV	IV	IV
Control group mean	0.15	0.03	0.19	0.11	0.02	0.14
R-Squared	0.03	0.01	0.03	0.05	0.01	0.04
Observations	10,611	10,611	10,611	6,325	6,325	6,325

Notes: Sample includes students in the RDD analysis sample. All models include main effects indicating the student's summer school group (i.e., failed math only, failed reading only, or failed both math and reading) along with controls for August and May test scores in math and reading, prior student achievement, and student demographics as well as interactions that allow the impact of all of these covariates to differ based on the summer school group. As described in the text, the key instruments are a set of nonlinear measures of August test scores. Huber-white standard errors are shown in parenthesis. *=significant at 10% level. **=significant at 5% level.

Table 6: The Heterogeneity of Grade Retention Effects on Dropout Rates

	6th Grade	8th Grade
Baseline estimates	-0.04 (0.04) [0.53]	0.08** (0.03) [0.57]
<i>Year</i>		
1997 Cohort	0.02 (0.09) [0.56]	-0.04 (0.06) [0.64]
1998 Cohort	-0.10* (0.06) [0.55]	0.14** (0.05) [0.57]
1999 Cohort	0.03 (0.06) [0.46]	0.17** (0.09) [0.52]
F-Statistic (Equal Coefficients)	2.56 [0.28]	8.38 [0.02]
<i>May Performance</i>		
Failed Reading	-0.02 (0.06) [0.48]	0.03 (0.05) [0.52]
Failed Math	-0.13 (0.11) [0.58]	0.05 (0.08) [0.61]
Failed Reading and Math	-0.03 (0.06) [0.62]	0.17** (0.05) [0.65]
F-Statistic (Equal Coefficients)	0.91 [0.64]	4.28 [0.12]
<i>School Quality</i>		
School Is in Top Half of Reading Distribution	-0.05 (0.06) [0.49]	0.12** (0.04) [0.56]
School Is in Bottom Half of Reading Distribution	-0.02* (0.05) [0.55]	0.04 (0.05) [0.59]
F-Statistic (Equal Coefficients)	0.19 [0.67]	1.37 [0.24]
<i>Race</i>		
Black	0.00 (0.05) [0.57]	0.11** (0.04) [0.60]
Hispanic	-0.13* (0.07) [0.44]	0.01 (0.06) [0.51]
F-Statistic (Equal Coefficients)	1.93	2.38

	[0.17]	[0.12]
<i>Gender</i>		
Male	-0.07 (0.05) [0.58]	0.01 (0.04) [0.66]
Female	-0.01 (0.06) [0.47]	0.16** (0.05) [0.49]
F-Statistic (Equal Coefficients)	0.63 [0.43]	5.84 [0.02]
<i>Family Income</i>		
Free Lunch	-0.06 (0.04) [0.54]	0.08** (0.03) [0.58]
No Free Lunch	0.21 (0.13) [0.45]	0.09 (0.09) [0.52]
F-Statistic (Equal Coefficients)	3.90 [0.05]	0.02 [0.88]

Notes: Sample includes students in the RDD analysis sample. All models include main effects indicating the student's summer school group (i.e., failed math only, failed reading only, or failed both math and reading) along with controls for August and May test scores in math and reading, prior student achievement, and student demographics as well as interactions that allow the impact of all of these covariates to differ based on the summer school group. As described in the text, the key instruments are a set of nonlinear measures of August test scores. Huber-white standard errors are shown in parenthesis. Mean dropout rates of promoted students in each sample are shown in square brackets. *=significant at 10% level. **=significant at 5% level.

Table 7: IV Estimates of the Impact of Grade Retention on Academic Progress of Students

All Students in RDD Sample					
Sixth Grade Cohorts (n=8,573)					
	Dropped out by Fall 2005	Retained in Eighth Grade	Age Began High School	Total credits upon leaving school (uncond.)	Age when left school
Retained in Grade	-0.04 (0.04)	-0.090** (0.029)	0.21** (0.05)	-0.76 (6.67)	0.02 (0.12)
Control group mean (s.d.)	0.53	0.140	14.91 (0.65)	113.55 (82.35)	17.71 (1.37)
R-Squared	0.10	0.080	0.332	0.095	0.047
Observations	8,573	6,984	7,758	8,256	7,744
Eighth Grade Cohorts (n=5,402)					
	Dropped out by Fall 2005	Retained in Eighth Grade	Age Began High School	Total credits upon leaving school (uncond.)	Age when left school
Retained in Grade	0.08** (0.03)	--	0.42** (0.02)	-17.15** (5.20)	0.11 (0.09)
Control group mean (s.d.)	0.57	--	14.73 (0.55)	115.55 (80.58)	17.84 (1.26)
R-Squared	0.05	--	0.625		0.052
Observations	5,402	--	5,134	5,248	5,376

Notes: The number of observations varies across specifications due to students dropping out. In addition, there are fewer observations included in the 8th grade retention specifications because only students who were enrolled in 8th grade in the CPS for a full year were eligible for retention and some students who did not drop out skipped 8th grade or left the system for this grade and later returned. Those with missing credit data are excluded from specifications in columns 2-6. The number of observations in the last column is lower because students who are still enrolled in the CPS at the end of our time period are not included in the estimation.

Table 7 (continued): IV Estimates of the Impact of Grade Retention on Academic Progress of Students

High School Graduates						
	Sixth Grade Cohorts (n=3,056)			Eighth Grade Cohorts (n=2,049)		
	Age Began High School	Total credits earned upon leaving school	Age when left school	Age Began High School	Total credits earned upon leaving school	Age when left school
Retained in Grade	0.155** (0.071)	1.31 (4.02)	0.177** (0.060)	0.503** (0.039)	-9.12** (2.83)	0.317** (0.050)
Control group mean (s.d.)	14.77 (0.60)	196.62 (28.02)	18.59 (0.56)	14.65 (0.55)	195.49 (25.74)	18.55 (0.60)
R-Squared	0.415	0.054	0.538	0.653	0.098	0.554
Observations	3,025	3,025	3,056	2,034	2,034	2,049

Notes: Sample includes students in the RDD analysis sample who graduated high school by Fall 2005. The number of observations in the samples is slightly higher than the number of observations included in some specifications because a small number of students are missing credit information. All models include main effects indicating the student's summer school group (i.e., failed math only, failed reading only, or failed both math and reading) along with controls for August and May test scores in math and reading, prior student achievement, and student demographics as well as interactions that allow the impact of all of these covariates to differ based on the summer school group. As described in the text, the key instruments are a set of nonlinear measures of August test scores. Huber-white standard errors are shown in parenthesis. *=significant at 10% level. **=significant at 5% level

Table 7 (continued): IV Estimates of the Impact of Grade Retention on Academic Progress of Students

High School Dropouts				
Sixth Grade Cohorts (n=4,688)				
	Retained in Eighth Grade	Age Began High School	Total Credits	Age when left school
Retained in Grade	-0.072* (0.043)	0.22** (0.07)	-4.70 (5.48)	-0.12 (0.16)
Control group mean (s.d.)	0.156	15.02 (0.65)	48.11 (48.16)	17.04 (1.44)
R-Squared	0.097	0.295	0.053	0.052
Observations	3,490	3,928	4,402	4,688
Eighth Grade Cohorts (n=3,327)				
		Age Began High School	Total Credits	Age when left school
Retained in Grade		0.36** (0.03)	-3.87 (4.09)	0.144 (0.116)
Control group mean (s.d.)		14.80 (0.55)	54.59 (48.70)	17.31 (1.30)
R-Squared		0.603	0.083	0.042
Observations		3,076	3,188	3,327

Notes: Sample includes students in the RDD analysis sample who dropped out of high school by Fall 2005. The number of observations varies across specifications due to students dropping out. In addition, there are fewer observations included in the 8th grade retention specifications because only students who were enrolled in 8th grade in the CPS for a full year were eligible for retention and some students who did not drop out skipped 8th grade or left the system for this grade and later returned. Those with missing credit data are excluded from specifications in columns 2-6. The number of observations in the last column is lower because students who are still enrolled in the CPS at the end of our time period are not included in the estimation. All models include main effects indicating the student's summer school group (i.e., failed math only, failed reading only, or failed both math and reading) along with controls for August and May test scores in math and reading, prior student achievement, and student demographics as well as interactions that allow the impact of all of these covariates to differ based on the summer school group. As described in the text, the key instruments are a set of nonlinear measures of August test scores. Huber-white standard errors are shown in parenthesis. *=significant at 10% level. **=significant at 5% level

Table 8: IV estimates of long-run effects of retention on cognitive ability

Dependent Variable	Third Grade					
	Full RD Sample	Students in RD sample who took 8 th grade test				
	Took 8 th grade exam	Age when tested	Math Score	Reading Score	Math Score	Reading Score
Retained	0.011 (0.027)	0.28** (0.03)	1.37 (1.34)	2.96* (1.69)	0.74 (1.40)	1.85 (1.76)
Controlling for age when tested	--	--	N	N	Y	Y
Control group mean (s.d.)	0.698	14.51 (0.55)	237.6 (22.6)	229.0 (25.1)	237.6 (22.6)	229.0 (25.1)
S.D. among all CPS students in 2004	--	--	29.21	33.78	29.21	33.78
R-Squared	0.083	0.405	0.347	0.162	0.349	0.167
Observations	15,424	10,207	10,146	10,177	10,146	10,177
Dependent Variable	Sixth Grade					
	Full RD Sample	Students in RD sample who took 11 th grade exam				
	Took 11 th grade exam	Age when tested	Math Score	Reading Score	Science Score	
Retained	-0.017 (0.034)	0.11** (0.05)	0.86 (0.99)	1.46 (1.10)	1.20 (0.94)	
Control group mean (s.d.)	0.375	17.48 (0.57)	137.47 (9.11)	139.67 (8.54)	135.89 (7.53)	
S.D. among all CPS students in 2004	--	--	14.38	14.26	14.71	
R-Squared	0.063	0.564	0.433	0.202	0.231	
Observations	10,611	3,569	3,294	3,290	3,291	
Dependent Variable	Eighth Grade					
	Full RD Sample	Students in RD sample who took 11 th grade exam				
	Took 11 th grade exam	Age when tested	Math Score	Reading Score	Science Score	
Retained	-0.033 (0.027)	0.23** (0.05)	2.19** (0.84)	0.60 (0.99)	1.18 (0.82)	
Control group mean (s.d.)	0.300	17.46 (0.61)	136.83 (8.48)	139.91 (8.24)	134.39 (6.98)	
S.D. among all CPS	--	--	14.38	14.26	14.71	

students in 2004

R-Squared	0.154	0.625	0.471	0.214	0.198
Observations	6,281	1,765	1,680	1,679	1,681

Notes: Sample includes all students in the RDD analysis sample. The outcome measure for the third grade cohort is the ITBS test scores the child received the first time that she took the 8th grade exam. The test score metric is a developmental scale score. The outcome measure for the sixth and eighth grade cohorts is the 11th grade PSAT test score that the child received the first time that she took the exam. The test score metric is a scale score. Within our sixth and eighth grade RD analysis samples, roughly 80 percent of students who take the 11th grade test go on to graduate high school. All models include main effects indicating the student's summer school group (i.e., failed math only, failed reading only, or failed both math and reading) along with controls for August and May test scores in math and reading, prior student achievement, and student demographics as well as interactions that allow the impact of all of these covariates to differ based on the summer school group. As described in the text, the key instruments are a set of nonlinear measures of August test scores. Huber-white standard errors are shown in parenthesis. *=significant at 10% level. **=significant at 5% level

Figure A1: Kernel Density Plots of August Test Scores

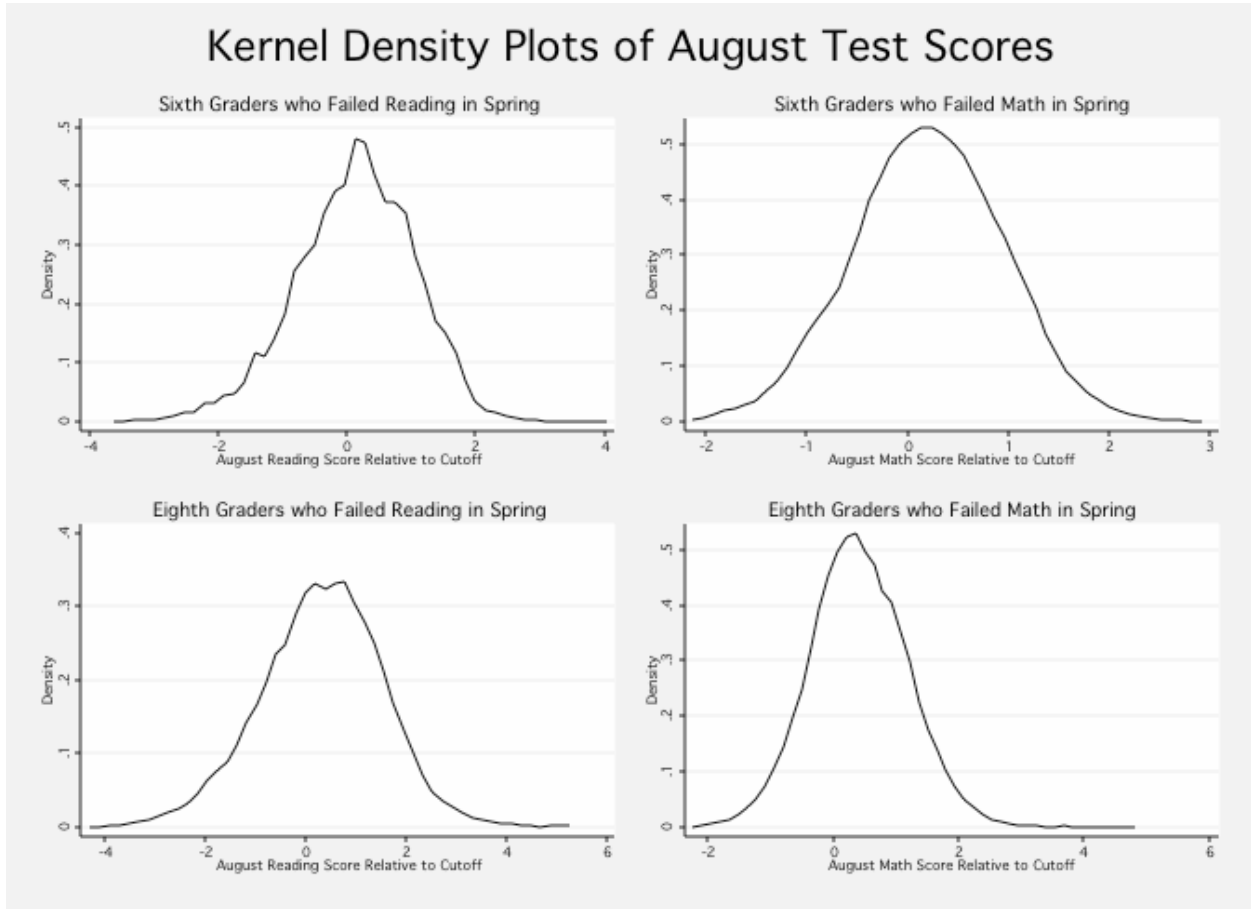


Table A1: The Robustness of Relationship between Grade Retention and Dropout Rate

Specification	Sixth Grade	Eighth Grade
Baseline	-0.04 (0.04)	0.08** (0.03)
Including 2 nd and 3 rd order polynomials in the August test scores	-0.03 (0.04)	0.07** (0.03)
Including 2 nd and 3 rd order polynomials in both August and May test scores	-0.03 (0.04)	0.08** (0.03)
Using a wider range around the August cutoff	-0.02 (0.02)	0.07** (0.02)
Using a wider range around the August cutoff and including 2 nd and 3 rd order polynomials in the August and May test scores	-0.05 (0.04)	0.07** (0.03)
Using a narrower range around the August cutoff	-0.05 (0.05)	0.11** (0.04)

Notes: Sample includes students in the RDD analysis sample. All models include main effects indicating the student's summer school group (i.e., failed math only, failed reading only, or failed both math and reading) along with controls for August and May test scores in math and reading, prior student achievement, and student demographics as well as interactions that allow the impact of all of these covariates to differ based on the summer school group. As described in the text, the key instruments are a set of nonlinear measures of August test scores. Each cell contains an estimate from a separate 2SLS regression that controls for all measures of the past performance and demographic characteristics listed earlier. Huber-white standard errors are shown in parenthesis. * = significant at 10% level. ** = significant at 5% level.