CLOSUP Working Paper Series
Number 12

February 2009

The Effect of Grade Retention on High School Completion

Brian A. Jacob
University of Michigan

Lars Lefgren
Brigham Young University

Center for Local, State, and Urban Policy
Gerald R. Ford School of Public Policy
University of Michigan

The Effect of Grade Retention on High School Completion

Brian A. Jacob and Lars Lefgren [*]

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. In this paper, we use plausibly exogenous variation in retention generated by a test-based promotion policy to assess the causal impact of grade retention on high school completion. We find that retention among younger students does not affect the likelihood of high school completion, but that retaining low-achieving eighth grade students in elementary school substantially increases the probability that these students will drop out of high school.

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 the social, emotional and cognitive development of 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 policy. 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 State Notes 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 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 concerns.

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 (Brian A. Jacob and Lars Lefgren 2004).

[1] In a survey of 47 empirical studies with a variety of academic achievement measures, C. Thomas 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 (E.M. Shulz et al. 1986; Russell W. Rumberger 1987; James B. Grissom and Lorrie A. Shepard 1989; Michelle Fine 1991; Melissa Roderick 1994). Several more recent studies have found moderate, positive effects of retention (Nancy L. Karweit 1991; Louisa H. Pierson and James P. Connell 1992; Karl L. Alexander, Doris R. Entwisle, and Susan L. Dauber 1995; A. Gary Dworkin, Jon Lorence, Laurence A. Toenjes, and N. Hill Antwanette 1999; Eric R. Eide and Mark H. Showalter 2001).

[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 (Brian A. 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.

In this paper, we utilize a similar identification strategy to examine the long-run effects of retention on high school completion. We find that retention among sixth grade students does not affect the likelihood of high school completion, but that retaining low-achieving eighth grade students in elementary school substantially increases the probability that these students will drop out of high school. The eighth grade dropout effect we find is concentrated among African-American girls, a group with the highest graduation rates in our analysis sample of low-achieving minority children (although still very low by any conventional measure).

It appears that the different effect we find across grade level is driven by the fact that students retained in earlier grades have more opportunities to catch up with their peers and, conversely, student who narrowly missed retention in earlier grades have more opportunities to "fall back" in subsequent years. For example, we find that roughly two-thirds of the students who are retained in the sixth grade have been able to rejoin their original cohort within two years (i.e., by the Spring of what was their original eighth grade year). Moreover, we find that students who were retained in sixth grade were significantly less likely to be retained in eighth grade, relative to their peers who narrowly avoided retention in sixth grade.

Our findings suggest that one must carefully assess the costs and benefits of grade retention policies for older students. Moreover, the difference in results between sixth and eighth grade students illustrates the importance of understanding how policies interact to affect the long-term impact of a particular intervention.

The remainder of this paper is organized as follows. Section 1 provides background on the Chicago accountability policy. Section 2 describes our data and Section 3 explains our empirical strategy. Section 4 presents the results, and Section 5 concludes.

I. Background

In this section, we briefly describe the Chicago accountability policy and then discuss the potential impacts of 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 likely to vary with 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 equivalents[3], 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.

The structure of the accountability program allowed retained students to catch up to the remainder of their cohort. Students who were retained had an opportunity to re-take the exams the following year. If they passed at this point, they were able to rejoin their original cohort. 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 in that subsequent accountability grade. Also,

[3] See the data section below for a discussion of the grade equivalent metric.
[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.

retained students may have been 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 were able to attend summer school in order to obtain credits at a faster rate. For these reasons, the impact of grade retention may have been mitigated through a combination of optimizing behavior on the part of students and their increased capacity to overcome subsequent hurdles to graduation.

For all third and sixth grade students, and many eighth grade students, retention meant remaining in the same grade in the same elementary school. However, eighth grade students who would have been 15 years old or older in the following school year were placed in special "transition" centers due to a concern that these older students should not remain among the younger elementary students. These centers were located throughout the city, occasionally in the same building as a regular high school. Students in transition centers received targeted instruction, largely in reading and mathematics, intended to help them quickly transition to the ninth grade. Moreover, in many cases students in the transition centers were encouraged to re-take the promotion exam in January of the following year, which if they passed, allowed them to rejoin their class more quickly. Hence, the "treatment" received by these older eighth graders was potentially quite different than that received by younger eighth graders from the same elementary school. For this reason, we will analyze these groups separately.

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 (Elaine M. 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. A recent analysis of grade retention in junior high school in Uruguay that uses an identification strategy similar to Jacob and Lefgren (2004) finds that retention increases dropout rates (Marco Manacorda 2007). The analysis below builds on this earlier work. In particular, we focus 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.

II. 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.

Our analysis sample consists of students who were (a) enrolled in the sixth or eighth grades 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, (d) failed the promotional cutoff and attended summer school, (f) took both the math and reading exams at the end of the Summer and (f) enrolled in the CPS the following Fall.[5] We exclude third graders from our analysis because they would not have progressed sufficiently to be at risk of dropping out by the end of our sample period.[6] Finally, we limit our primary analysis sample to students who scored between 1 grade equivalent below and 0.5 grade equivalents above the promotional cutoff at the end of the Summer.[7] Our sample includes 11,777 sixth grade students and 7,509 eighth grade students, roughly 41 percent of whom were retained.

Table 1 presents summary statistics for our primary analysis sample. Not surprisingly, the sample students are highly disadvantaged. 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 fraction of students that have attended bilingual programs, are in foster care, or live with a non-parent relative.[8] Comparing the students in our analysis sample who were retained to those who were promoted, we see that they are nearly identical in terms of observable demographics and prior achievement.

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 five years following the start of high school for all sixth and eighth grade students. Students

---

[5] 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.

[6] The most recent data available to us ends in Fall 2005, at which point the 1997 cohort of third grade students would only be in 11th grade even if they were never retained.

[7] The asymmetry in the ranges above and below the cutoff is due to the fact that there is not a sharp discontinuity in the probability of grade retention at the official cutoff. Instead, the probability of retention drops quickly in a marginal area just below the cutoff. The asymmetric range around the official cutoff ensures that we have adequate data both above the official cutoff and below the marginal area.

[8] 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 few students in these programs who were subject to the policy are included.

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. Based on the experiences of earlier cohorts, however, 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, although excluding this cohort yields qualitatively comparable results.

Dropout status can be a difficult variable to measure in many administrative datasets since it requires tracking individual students over time and determining the status of students who are no longer regularly attending the public school system. As a specification check, therefore, we re-estimate our models using high school graduation (which is measured very well in the district records) as the outcome, and find comparable results (see Table 4).

Finally, given its central role in our analysis, it is worth describing the grade equivalent metric in which Chicago test scores are reported. The grade equivalent (GE) scale is a non-linear 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 Iowa Tests of Basic Skills (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). Interestingly, the test is scored such that not all grade equivalent values are possible in each subject-grade-year, and some grade equivalent values are much less common than others. 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 does not present any problems for the analysis, although as we discuss below, this does make some figures that aggregate across failure groups, grades and years difficult to interpret visually.

III. Empirical Strategy

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

(1) $Y_i = \beta retain_i + \Gamma X_i + u_i + \varepsilon_i$

where Y is the outcome, $\boldsymbol{X}$ 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 taking advantage of Chicago's grade retention policy, which led to a highly non-linear relationship between a student's August test performance and the probability of grade retention. This non-linearity generates plausibly exogenous variation in retention, which we leverage to identify the causal effect of grade retention on the probability of dropout. The intuition underlying our strategy is similar to that behind traditional regression discontinuity designs (RDD). However, in our preferred specification we instrument grade retention with a continuous, though highly non-linear, function of the August test score to most effectively model the relationship between test performance and grade retention. In sensitivity analyses shown in Table 3, we estimate a more standard RDD model that uses variation in grade retention attributable only to surpassing the official cutoff and find virtually identical results.

We continue by outlining how the rules governing student promotion under Chicago's accountability policy facilitate an evaluation of the policy. We then present visual evidence regarding the impact of grade retention on dropout behavior. Finally, we describe specific estimation strategies we employ.

A. Identification

By tying student promotion to performance on standardized 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. Students who scored below predetermined cutoffs in either reading or math at the end of the school year were assigned to attend summer school. Students were re-tested at the end of the summer, and those who again failed to achieve these cutoffs were required, in principle, to repeat the grade.

August test scores, therefore, became the gatekeeper for determining promotion. For our analysis, we define an index, which is the August test score relative to the cutoff in the marginal subject. To begin, we normalize August reading and math scores by subtracting the cutoff score. For example, if a sixth grade student scored 5.1 grade equivalents on the August math score, her normalized score would be 5.1 – 5.3 (the cutoff for this grade), which equals -0.2. The construction of the index score then varies across three groups. The first group includes students who failed reading but passed math during the school year. For these students, the index is their August reading score minus the August reading cutoff. For students who failed math but passed reading, the index equals their August math score minus the August math cutoff. Finally, for students who failed both math and reading, the index is the minimum of their normalized August reading and math scores.

Figures 1a-c present some graphical evidence on the relationship between student index scores and a variety of outcomes for sixth grade students. To simplify the presentation, we aggregate data from the three "failure groups" defined above and the three academic year cohorts (i.e., 1997, 1998 and 1999). In all graphs, the height of the circles reflects the actual average value of the outcome for a particular index score group and the size of the circles reflects the number of students in the group. The solid line represents the predicted value of the outcome based on a spline specification, with knot points at the index value of zero (the official promotional cutoff) and -0.4, which is roughly the beginning of what we will refer to below as the "marginal" area. The two vertical lines are meant merely to draw the reader's attention to these values.

Figure 1a shows the relationship between the index score and the likelihood of retention. Three distinct regions are apparent in the graph. Above the cutoff, nearly all the students were promoted with a very small number of retentions for reasons such as poor attendance regardless of academic achievement. More than 0.4 points below the cutoff, roughly 78 percent of students were retained. Had the policy been strictly enforced, all of these students would have been retained. However, the policy allowed for parents and teachers to request and obtain waivers for students and it appears that roughly 22 percent of students in this group did so. Between the cutoff and 0.4 points below, students had a positive but declining probability of being retained. In this range, which we will refer to as the marginal area, the likelihood of retention dropped from 70 percent to 7 percent at the cutoff. In informal discussions, officials indicated that school and district administrators were more likely to grant waivers for students who just missed the cutoff, which explains this pattern.

While there is a clear non-linear relationship between the index score and the likelihood of grade retention, this relationship is not discontinuous as is the case in a traditional regression discontinuity design (RDD). If this were the case, one could 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, all of whom were promoted. Given the imperfect compliance in our case, this approach is inappropriate.

Instead, our preferred estimation strategy exploits the non-linearity between the index score and retention probability to estimate the causal effect of grade retention. As described in more detail below, this amounts to a two-stage least squares procedure in which we instrument grade retention with a non-linear function of the index score, controlling for a smooth function of the index score. Identification arises from differences in the functional form between the first and second stages. In an important specification check, however, we estimate a "standard" RDD model that relies solely on the variation generated by the official cutoff and obtain comparable results.

Figure 1b illustrates the intuition behind our identification. The graph shows the relationship between index score and the probability of dropping out of school. Because of the structure of the math and reading tests used, when we aggregate data across different failure groups and years the relationship between index score and the raw dropout probability is very jagged.[9] Hence, to better illustrate the relationship of interest, we graph average dropout

---

[9] As noted above, the test is scored such that not all grade equivalent values are possible in each subject-grade-year, and some grade equivalent values are much less common than others. Now recall that the index scores we use in our analysis reflect math scores for some students (e.g., students who failed reading in the Spring and thus need to pass the math exam) and reading scores for other students (e.g., students who failed math in the Spring). Given how the exams were scored, certain index values will disproportionately reflect students who failed math only, failed reading only or failed both math and reading. It turns out that students who failed math are more likely to eventually drop out than students who failed reading, and students who failed both math and reading are the most likely to drop out. For this reason, when we aggregate across failure groups and years, the raw dropout probability varies across nearby

residuals against the index score. The dropout residual is obtained by regressing (separately by grade) our dropout variable on dummy variables for failure group x cohort (recall that there are three failure groups: the student failed math only, reading only, or both math and reading), along with student demographics and May math and reading scores, which are themselves also interacted by the failure group x cohort indicators. Note that we do not control for the August test scores in calculating the dropout residual since the index value is defined in terms of these scores. Even in the absence of the retention policy, student test scores are clearly associated with high school completion. Hence, we would expect that the likelihood of dropping out is negatively associated with the index score.

And, indeed, in Figure 1b we see that the index score is negatively associated with our dropout residual measure. If retention were associated with high school completion, we would expect the probability of dropping out to decrease sharply in the marginal area, since the likelihood of dropping out decreases sharply in this range. Instead, the graph in Figure 1b shows a smooth, slightly downward sloping relationship between index score and the likelihood of dropping out, implying that grade retention in the sixth grade had little influence on high school completion.

Interestingly, the comparable figures for young eighth grade students (Figures 2a-c) paint a very different picture. Figure 2a shows that the relationship between index score and retention for these students mirrors the pattern for sixth graders – namely, very high retention rates for students with index values below -0.2 (the beginning of the marginal area for this grade), a sharp decline in the likelihood of retention over index values in the marginal area, and virtually no retention among students scoring above the official cutoff. In contrast to the pattern among sixth

---

index values, which makes the relationship between the index score and the probability of dropout very jagged and graders, Figure 2b shows that the probability of dropping out declines noticeably in the marginal area. The fact that the probability of being retained and of dropping out decline over the exact same range suggests that retention decreases the likelihood of completing high school. Interestingly, an examination of these patterns for older eighth grade students in Figures 3a-c imply that grade retention does not have a causal impact on dropping out for these students.

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. Since our instruments are non-linear 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 with 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 (as we do here), 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 non-linear measures we use as instruments are not correlated with any unobserved student characteristics that could independently influence the outcome of interest. In other words, our strategy would yield biased results if, for reasons *unrelated* to grade retention, there was a sharp change in the slope of the relationship between our index value and grade retention in the vicinity of the cutoff. A specific reason for this concern is that students or

---

difficult to interpret.

teachers may manipulate effort or engage in cheating to move students from below the cutoff to just above.[10]

One way to test this assumption is to plot the relationship between the index score and observable student characteristics that may be correlated with student outcomes. To efficiently summarize a variety of potentially important observable characteristics, we use the predicted probability of dropping out derived from a linear probability model with controls for a full set of failure group x cohort indictors, along with the Spring math and reading scores from the current year, the math and reading scores from prior years, and student demographics, all of which are themselves interacted with the failure group x cohort indicators. Figures 1c, 2c and 3c show the relationship between the index value and the predicted probability of dropping out. While the relationship is somewhat noisy, we see no evidence of any systematic change in relationship between our index and the predicted probability of dropout in our marginal area for either sixth or eighth graders.[11] In an online appendix we formalize this test by estimating models that mirror our basic specification described below, but where the predicted dropout probability is the dependent variable, and find no significant results. Additional analyses, also in the online appendix, show no evidence of students clumping just above the promotional cutoff, which would occur if teachers cheated to push marginal students above the cutoff.

<u>B. Estimation</u>

Following the standard parametric RDD (see Guido W. Imbens and Thomas Lemieux (2008) for an overview), we specify the following estimation equation:

(2) $$y = \gamma_1 retained + f(index) + BX + \varepsilon,$$

---

[10] Patrick J. McEwan and Miguel Urquiola (2005), for example, find evidence that schools in Chile manipulate enrollments to avoid hiring additional teachers.

where y is the academic outcome of interest, $f(index)$ is a polynomial in the index, and $\boldsymbol{X}$ is a vector of covariates. In a sharp RDD, treatment status is determined entirely by whether or not the index score exceeds the cutoff, and one can simply estimate equation (2) via OLS. In the case of a fuzzy RDD, there is a discontinuous change in the probability of treatment at the cutoff, though the probability of treatment does not change by one. In such cases, one can recover γ_1 by estimating equation (2) via instrumental variables where the excluded instrument is a binary indicator for surpassing the cutoff.

In our case, the probability of grade retention does not typically change discontinuously at the official cutoff. While a dummy variable indicating whether a student passed the official cutoff does generate variation in grade retention, it fails to fully characterize the relationship between the index score and the probability of grade retention. Instead, as shown in Figures 1-3, the empirical relationship between our index and the probability of grade retention can be effectively captured by a linear spline. Because it better matches the observed pattern in the data, the use of a spline function (rather than a single indicator variable for passing the official cutoff) should increase our statistical precision.

In results not presented here, we find that the precise nature of the relationship varies across grade, cohort year and failure group. For example, among eighth grade students who attended summer school in 1997 or 1998, the relationship between the index score and the probability of grade retention was discontinuous at the cutoff. However, waivers apparently became more common in 1999, as the relationship between index score and the probability of grade retention exhibited a spline similar to the other grades and cohorts.

---

[11] As mentioned earlier, this noisiness is due to the fact for some index values students who failed math are disproportionately represented. Such students are more likely to ultimately dropout.

Due to differences in the implementation of the accountability policy, we can further increase the statistical precision of our estimates by allowing each of the nine separate experiments corresponding to a year x failure group (e.g. failed math in 1997, failed reading in 1998, failed both math and reading in 1999) to have a different first-stage relationship between index score and retention. In our specification, each of these first-stage equations has the following form:

(3) $$retain = \gamma_1 index + \gamma_2 index\_marginal + \gamma_3 index\_above + \Gamma X + \eta ,$$

where *index* is defined as above, the variable *index_marginal* takes on a value of zero at the bottom of the marginal area and increases one-for-one to the cutoff and maintains a constant value above that point, and the variable *index_above* is zero below the cutoff and equal to *index* above the cutoff. Collectively, these three variables define a linear spline with knot points at the bottom of the marginal area and the official cutoff. This non-linearity between the index score and the probability of grade retention generates our identifying variation. ***X*** is a vector of student demographic characteristics including the student's May math and reading scores and test scores in prior years. We determine the marginal area for each year x failure group by performing a grid search in which we iterate over possible marginal areas and select the marginal area that maximizes the first stage r-squared.[12]

In practice, we estimate a single first-stage equation in which we interact all instruments and covariates with dummy variables for each year x failure group experiment. This is numerically equivalent to estimating the first-stage equations separately but allows us to perform joint inference on our excluded instruments. In this combined first-stage equation, we have

[12] One may be concerned that by choosing the marginal area empirically, we obtain point estimates that are biased and/or less precise than indicated. However, as shown in Table 3, our findings are quite similar, though slightly less

eighteen excluded instruments (i.e., *index_marginal* and *index_above* for each of the nine experiments). In our preferred specification (Table 2, row 4), $f(index)$ is assumed to be linear and all covariates are interacted with indicator variables for each experiment. However, we present an extensive set of sensitivity analyses in which we vary the sample range, the covariates and the functional form.

C. Interpretation

Given the nature of our identification strategy, it is particularly useful to discuss the Local Average Treatment Effect (LATE) we estimate. 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, in the vicinity of the cutoff and marginal area, whose treatment status was determined by their August test performance. The marginal area 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.

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. 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

---

precise, when we include only a passed indicator variable as our instrument, which exploits only variation 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.

IV. Results

In this section, we discuss the effects of retention on educational attainment, including high school completion. We examine the impact of sixth and eighth grade retention on the likelihood that a student completes high school. In addition, we explore the age at which the student left school and the number of credits she accumulated prior to leaving.

A. The Impact of Retention on High School Dropout

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.[13] Of course, if retention influences attrition from the CPS, this may bias our estimates. The results presented in Table 4 indicate that this is not a concern.

The OLS estimates are shown in Table 2. In the first row, we present estimates that only include indicators for each of the 9 failure group x cohort cells. The resulting point estimates imply that grade retention is associated with a 4 percentage point increase in the probability of dropping out for students retained in the sixth grade and roughly a 7-8 percentage point increase

---

attributable to the formal retention policy and thus does not suffer any such concern.

for students retained in the eighth grade. Given the baseline probability of dropping out for promoted students in these groups (i.e., the control mean of the dependent variable shown at the bottom of the table), these estimates suggest that grade retention increases the likelihood of dropping out by 8 to 15 percent. It is interesting to note that the dropout rate among the older eighth graders is substantially higher than among either of the two other groups. This is due to the fact that older eighth graders were more likely to have been retained in the past, and thus age within grade is a proxy for prior difficulty in school that is associated with future outcomes.

In row 2, we show OLS estimates that include a full set of demographic and prior achievement variables, all of which are allowed to vary by the 9 failure group-cohort cells. Covariates include a third order polynomial in May math and reading scores, second order polynomials in test scores from one and two years prior, as well as controls for age, ethnicity, gender, gender*ethnicity interactions, free or reduced price lunch status, current and past limited English proficiency, special education enrollment, non-parent caretaker, social status, and poverty concentration in student's neighborhood. For each group, the point estimates drop substantially, suggesting that observable measures of prior achievement and student demographics are associated with both grade retention and the likelihood of dropping out. Conditional on this extensive set of controls, we find that students retained in sixth grade are 2.7 percentage points (5 percent) more likely to drop out of school and the younger eighth graders who are retained are 4.5 percentage points (9 percent) more likely to drop out. In this specification, retention is not associated with dropout among older eighth graders.

[13] With the exception of the sixth grade cohort in 1999, 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.

Rows 3 and 4 in Table 2 present these IV results. Following David S. Lee and David E. Card (2008), we use Huber-White standard errors that are clustered by index value x failure group x cohort.[14] Row 3 presents results in which we only control for a linear measure of the index score. Row 4 shows results that control for the full set of student demographics and prior achievement variables described above, all of which are allowed to vary across failure group x cohort. The comparison between rows 3 and 4 provides a useful check on the validity of our instruments. If our estimates change significantly with the inclusion of other covariates, our instruments must be correlated with observable factors that determine the likelihood of dropping out independent of retention.[15]

First stage F-statistics indicate that our instruments are quite strong. The lowest first stage F-statistic is 144, in our sixth grader model that includes covariates. In the eighth grade specifications, the F-statistics are uniformly above 500. In all cases, the instruments are easily strong enough to preclude concerns regarding weak instruments bias.

Consider first the results for the sixth graders. The point estimate shown in row 3 is actually negative, though not statistically significant. The addition of covariates in row 4 does not change the estimate appreciably, providing evidence to support our identifying assumption. The standard error on the estimates in column (3) implies a confidence interval roughly between -0.108 and 0.044. 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

[14] Interestingly, we find that conventional (and White heteroskedasticity robust) standard errors are actually slightly larger than the cluster-correct standard errors we present here, though the differences are quite small so the use of different standard error calculations never changes the statistical significance of the point estimates. These results are available in the online appendix.

[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.

grade retention in the sixth grade has little substantive effect on the probability a student drops out.

For the set of older eighth graders (i.e., those who moved to a transition center in the event of retention), the story is similar. The IV estimates in rows 3 and 4 are small and statistically indistinguishable from zero. Moreover, there is no statistical difference (and little substantive difference) between the estimates with and without covariates (rows 3 and 4 respectively), providing support for our exclusion restriction.

In contrast, the IV estimates for the younger eighth graders suggest that retention has a large, positive impact on dropping out. The preferred estimate in row 4 indicates that retention increases the likelihood of dropping out by 11.2 percentage points, or roughly 21 percent. Again the specification is largely invariant to the set of covariates that are included. It thus appears that retention of younger eighth graders is substantially more harmful for the retained students than is the case for either sixth graders or older eighth graders. We discuss potential reasons for these differences below.

For the younger eighth grade students only, tests of over-identifying restrictions are rejected at the 1 percent level. This implies that the estimated treatment effect varies significantly based on the particular set of instruments used. Thus, the impact of grade retention varies over time and/or the subject that was failed. We explore these differences in greater detail when we explore the heterogeneity of effects in Table 5.

Comparing the OLS and IV estimates, we see small differences though there is no consistent pattern to these differences nor are they ever statistically significant. For the sixth graders and the older eighth graders, the IV estimates are smaller than the OLS estimates, but the difference is only meaningful for sixth graders. The differences we observe may be due to unobserved differences between students who obtain waivers and those who do not. Alternatively, the LATE captured by our IV strategy may reflect a different population than the estimate identified by OLS. In any event, the differences are sufficiently small and unsystematic that we hesitate to place much emphasis on their interpretation.

Table 3 presents the results from a variety of sensitivity analyses, all of which indicate that the findings described above are quite robust. Row 1 replicates our preferred baseline specification from row 4 of Table 2. Rows 2 and 3 expand our analysis sample to include students who scored in a larger range around the cutoff. The point estimates for young eighth graders drop slightly, but remain statistically significant and substantively important. The specifications shown in rows 4 and 5 use the larger sample from row 3, but include polynomials in the index score rather than a simple linear trend. The inclusion of polynomials in the index score, particularly those that vary above and below the cutoff shown (row 5), yield point estimates very similar to the baseline estimates. This suggests that the linear approximation for index score is not sufficient to capture the relationship between August test scores and dropout in a larger sample. The similarity of results in rows 1 and 5 thus supports the choice of our baseline sample and specification.

In our baseline specification, the majority of the identifying variation is coming from differences in retention and dropout probabilities among students in or near the marginal area, although we do implicitly use variation from individuals at the extremes (within our limited window) of the index value. Since the range of index values we use is quite narrow, this is unlikely to be problematic. But in order to test this, row 6 shows the results from a specification that exclusively uses variation associated with reductions in the dropout probability in the marginal area. To do so, we allow in our second stage for the relationship between the index value and dropout probability to differ above and below the cutoff, and limit our excluded instruments to the variable capturing the slope in the marginal area. The resulting estimates are again similar to our baseline.

As noted earlier, our baseline specification deviates from a traditional RDD in that it uses a spline function rather than a binary indicator to capture the non-linearity in the first-stage relationship. We are thus using, in part, variation generated by the informal waiver behavior of schools to identify the impact of retention. The specifications in rows 7-8 follow the traditional RDD approach, modeling the first-stage relationship between the index score and grade retention with a binary indicator for passing the official cutoff. The specification shown in row 8 replicates what has become one of the standard approaches in RDD analysis – namely, we include a cubic in the index score that is allowed to vary above and below the cutoff, along with a binary variable for passing the cutoff. All of these terms, along with the other control variables, are allowed to vary across failure group x cohort. The results shown in row 8 are virtually identical to those in our baseline specification.

In our baseline specification, we use student-level data but cluster-correct the standard errors at the level of the failure group x cohort x index value since this is the level at which the identifying variation exists. An alternative, and perhaps more conservative approach, would be to run the analysis using data aggregated to the level of failure group x cohort x index value. In results available from the authors, we show that this approach yields both point estimates and standard errors comparable to our baseline specification.

Table 4 explores the sensitivity of the results along a different dimension. As described above, if grade retention influences the propensity to leave the CPS, the dropout estimates above (in which the dropout variable is set to missing for students who transfer to another public school district or a private school) may be biased by differential attrition. Table 4 shows IV 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 5-15 percent of promoted students switch public school districts and 4-15 percent transfer to a private school. The estimates are uniformly small and statistically insignificant, suggesting that differential attrition is not a concern in the analysis. The final column in Table 4 shows IV estimates of the effect of grade retention on the likelihood of graduating from the Chicago Public Schools. Given the relatively high quality of graduation data, these estimates should be less sensitive to measurement error than the dropout models. Consistent with our early results, we find that retention has no impact on graduation for sixth graders and older eighth graders, but reduces the likelihood of graduating by 9.9 percentage points (22 percent) among younger eighth graders.

Table 5 examines the heterogeneity of effects across a variety of subgroups. It does not appear that the zero effects for sixth graders and older eighth graders are masking any important positive or negative effects among the subgroups we examine, but in some cases the standard errors do not allow us to identify moderate differences. Most interestingly, it appears that the increase in dropout rates for younger eighth graders is driven by very large impacts 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 whose probability of dropping out is substantially higher than their peers even among promoted students. 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 sample as well. For example, in our sample of young eighth graders, 45 percent of girls who are promoted in eighth grade drop out compared with 60 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 18 percentage points for girls implies that retention is roughly 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 this result.

In summary, our analysis suggests that grade retention among low-achieving students in Chicago has no impact on high school completion for sixth grade students or older eighth grade students, but substantially increases the likelihood of dropping out among younger eighth grade students. What might explain these results? To shed further light on the mechanisms underlying our results, Table 6 explores the impact of retention on several measures of academic progress. Note that the sample here is the same as that used in Tables 2 and 3 – namely, it excludes students who transferred to a private school or a different school district. We have shown that retention did not influence the likelihood that a student left the CPS in this way, which mitigates any concern about bias due to differential attrition.

One reason why grade retention among sixth graders may not have had a large impact on high school completion is that the accountability policy in Chicago created opportunities for retained (promoted) sixth graders to catch up with (fall back towards) their peers. All sixth grade students, both retained and promoted, faced another promotional gate once they reached the eighth grade. If the students who were promoted in sixth grade were more likely to be retained in eighth grade (relative to their peers who were retained in sixth grade), then retained and promoted sixth graders may have entered high school at the same time. In addition to this "fall back" scenario, it is possible that the remedial support and re-test opportunities provided to sixth graders who were retained may have allowed them to "catch up" to their promoted peers prior to the eighth grade.

The top panel of Table 6 explores the impact of retention on the academic progression of sixth graders. Column 1 shows IV estimates of the impact of grade retention among sixth graders on the likelihood that they were enrolled in the CPS two years later. The insignificant point estimate indicates that retention in sixth grade did not cause students to leave the CPS in the short-run. Column 2 shows the effect of retention on the likelihood that a student was enrolled in the eighth grade (or higher) two years later. Given that retention among seventh graders was rare during this period, if the retained sixth graders could not "catch up" through re-testing, we would expect the coefficient on the retention variable to be -1. Instead, the point estimate is -0.347, indicating that sixth grade retention reduced the likelihood of being in the eighth grade two years later by only 34.7 percentage points. This implies that roughly two-thirds of students who were retained as a result of "just missing" the cutoff score were able to rejoin their original cohort by the eighth grade. That is, many retained sixth graders caught up with their peers. Column 3 shows that students retained in the sixth grade were roughly 9 percentage points (64 percent) less likely to be retained in eighth grade, relative to their peers who were promoted in sixth grade.[16] Given the rapid convergence between retained and promoted students, it is perhaps unsurprising that we observe little impact of sixth grade retention on total credits accumulated (column 7) or the age that students leave school (column 6).

[16] The sample size for column 3 is considerably smaller than for column 2 because the sample only includes students who were enrolled in the CPS in eighth grade, eligible to be retained (i.e., excluding students in special education or bilingual programs), had non-missing test scores in the Spring and enrolled in the CPS in the following Fall. We find no effect of sixth grade retention on these intermediate outcomes, so we are not worried about sample selection bias in this specification. The sample size for column 6 is smaller than the other columns because (as discussed earlier) a substantial fraction of sixth graders from the 1999 cohort were still enrolled in school at the time of our last data collection. The small differences in the number of observations between column 4 and columns 5-7 is due to missing birthday or course credit information.

Moving down to the next panel, we can examine how grade retention among younger eighth graders affects their progression through school. Recall that our earlier results indicate that retention among this group increases the likelihood of dropping out by roughly 11 percentage points or 22 percent. In column 4, we see that much of this dropout effect occurred during the transition from elementary to secondary school. More specifically, students retained in the eighth grade were 8 percentage points less likely to enter high school in the CPS (column 4). Because grade retention had no impact on the fraction transferring to private schools or different school districts, this 8 percentage points reflects students who dropped out of school prior to attending high school anywhere. The retained students who did enter high school were roughly 0.87 years older than their peers who were promoted due to "just passing" the cutoff (column 5). Interestingly, the results in column 6 suggest that retention did not change the age at which these eighth graders left school, but did reduce the number of credits students accumulated before leaving school (column 7).

The bottom panel shows results for the older eighth graders, a group for whom retention did *not* impact high school completion. In column 4, we see that retention did substantially reduce the likelihood that students in this sample entered high school in the CPS. However, the results in columns 6 and 7 indicate that retention was not associated with credits accumulated or age leaving school. This may be due, in part, to the opportunity for retained students to earn some high school credit in the transition centers. It also suggests that the students in this sample who avoided retention due to "just passing" the cutoff in eighth grade must have been relatively unproductive in terms of earning high school credit despite being able to move directly from eighth grade to high school.

V. Conclusions

In this paper, we exploit the plausibly exogenous variation generated by a test-based promotion policy in the Chicago Public Schools to examine the causal effect of grade retention on high school completion. We find that retention among sixth grade students does not affect the likelihood of high school completion, but that retaining low-achieving eighth grade students in elementary school substantially increases the probability that these students will drop out of high school. The eighth grade dropout effect we find is concentrated among African-American girls, a group with the highest graduation rates in our analysis sample of low-achieving minority children (although still very low by any conventional measure).

It appears that the null finding for sixth graders is driven, at least in part, by the opportunities that the retained students have to "catch up" with their peers as well as the opportunities their promoted peers have to "fall back" in subsequent years. For example, we find that roughly two-thirds of the students who are retained in the sixth grade were able to rejoin their original cohort within two years (i.e., by the Spring of what was their original eighth grade year). Moreover, we find that students who were retained in sixth grade were significantly less likely to be retained in eighth grade, relative to their peers who narrowly avoided retention in sixth grade.

Interestingly, we found that grade retention has no impact on the probability of dropout among older eighth graders who were sent to transition centers instead of remaining in elementary school. This may be because the nature of the retention experience was substantially less demoralizing than that for the younger eighth graders, or because transition centers offered more opportunities for students to catch up to their peers. Alternatively, the smaller impact may simply reflect that these older students were a lower achieving group with a higher baseline probability of dropout.

Our findings have two important policy implications. First, the results presented above suggest that one must carefully assess the costs and benefits of grade retention policies for eighth graders. In doing so, however, one should consider not only the potential cost to students who are retained, but also the potential benefit of the accountability policy on all students. One of the objectives of the policy was to encourage elementary students and their teachers to strive for higher achievement, with summer school and grade retention serving as both a "stick" to motivate as well as an intervention to provide remedial instruction. In prior analysis, Jacob (2005) has shown that Chicago's accountability policy increased student performance substantially, particularly among eighth graders, although that analysis was not able to separate the impact of the social promotion policy from the school-based accountability policy that the district implemented at the same time.

Second, our analysis underscores the need to understand 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 attenuate or exacerbate the long-term impact of the initial intervention. The importance of institutional context in understanding the long-term effects of grade retention suggests that our findings may not generalize to all settings. However, the experience of Chicago's social promotion policy will likely be useful in helping to understand the potential effects of similar accountability policies in other large, urban districts such as New York City and Los Angeles.

* Jacob: Gerald R. Ford School of Public Policy, University of Michigan, 735 S. State Street, Ann Arbor, MI 48109 (e-mail: bajacob@umich.edu); Lefgren: Department of Economics, Brigham Young University, 130 Faculty Office Building, Provo, UT 84602-2363 (e-mail: l-lefgren@byu.edu). 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 Elias Walsh for excellent research assistance. All remaining errors are our own.

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, Karl L., Doris R. Entwisle, and Susan L. Dauber. 1995. *On the Success of Failure: A Reassessment of the Effects of Retention in the Primary Grades.* New York: University of Cambridge Press.

Dworkin, A. Gary, Jon Lorence, Laurence A. Toenjes, and Antwanette N. Hill. 1999. Elementary School Retention and Social Promotion in Texas: An Assessment of Students who Failed the Reading Section of the TAAS. Houston, TX: University of Houston Press.

ECS State Notes. 2000. Education Commission of the States. www.ecs.org (accessed July 23, 2008).

Eide, Eric R., and Mark H. Showalter. 2001. "The Effect of Grade Retention on Educational and Labor Market Outcomes." *Economics of Education Review*, 20: 563-576.

Fine, Michelle. 1991. *Framing Dropouts: Notes on the Politics of an Urban Public High School.* Albany, N.Y.: SUNY Press.

Grissom, James B., and Lorrie A. Shepard. 1989. "Repeating and Dropping Out of School." In *Flunking Grades: Research and Policies on Retention,* ed. Lorrie A. Shepard and Mary L. Smith, 34-63. New York: The Falmer Press.

Holmes, C. Thomas. 1989. "Grade Level Retention Effects: A Meta-Analysis of Research Studies." In *Flunking Grades: Research and Policies on Retention*, ed. Lorrie A. Shepard and Mary L. Smith, 16-33. New York: The Falmer Press.

Imbens, Guido W., and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics*, 142(2): 615-635.

Jacob, Brian A., and Lars Lefgren. 2004. "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis." *Review of Economics and Statistics*, LXXXVI (1): 226-244.

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

Karweit, Nancy L. 1991. "Repeating a Grade: Time to Grow or Denial of Opportunity." Baltimore, MD: Johns Hopkins University Center for Research on Effective Schooling for Disadvantaged Students Report No. 16.

Lee, David S., and David E. Card. 2008. "Regression Discontinuity Inference with Specification Error." *Journal of Econometrics*, 142 (2): 655-674.

Manacorda, Marco. 2008. "The Cost of Grade Retention." London School of Economics Center for Economic Performance Discussion Paper.

McEwan, Patrick J., and Miguel Urquiola. 2005. "Precise Sorting Around Cutoffs in the Regression-Discontinuity Design: Evidence from Class Size Reduction." http://www.u-bourgogne.fr/colloque-iredu/posterscom/communications/mcEwanUrquiola.pdf.

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

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

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

Schulz, E. M., R. Toles, W.K. Rice, I. Brauer, and J. Harvey. 1986. *Association of Dropout Rates with Student Attributes.* San Francisco, CA: American Educational Research Association.

Table 1: Summary Statistics

	6th Grade		8th Graders, Younger than 14.4 years old in May		8th Graders, At least 14.4 years old in May	
	Retained	Promoted	Retained	Promoted	Retained	Promoted
Fraction Retained	0.417		0.382		0.427	
Background Characteristics						
Black	0.675	0.662	0.640	0.657	0.682	0.660
Hispanic	0.295	0.299	0.332	0.309	0.293	0.305
Male	0.509	0.509	0.471	0.448	0.510	0.499
Eligible for free lunch	0.907	0.895	0.878	0.844	0.869	0.846
Eligible for reduced-price lunch	0.045	0.050	0.043	0.057	0.033	0.053
Currently in bilingual program	0.207	0.192	0.217	0.150	0.190	0.152
In bilingual program in the past	0.089	0.107	0.114	0.166	0.100	0.160
In special education program	0.025	0.019	0.013	0.011	0.013	0.015
Lives in foster care	0.069	0.063	0.050	0.048	0.058	0.068
Lives with relative other than parent	0.097	0.101	0.078	0.081	0.071	0.086
Social status of neighborhood (composite)	-0.442	-0.434	-0.470	-0.393	-0.473	-0.377
Neighborhood poverty index (composite)	0.453	0.412	0.422	0.365	0.437	0.348
Age	12.415	12.488	14.049	14.058	14.936	14.925
1997 Cohort	0.322	0.379	0.377	0.306	0.305	0.283
1998 Cohort	0.340	0.317	0.400	0.330	0.445	0.347
1999 Cohort	0.338	0.305	0.223	0.364	0.249	0.370
Failed reading only in May	0.408	0.514	0.393	0.399	0.299	0.363
Failed math only in May	0.157	0.221	0.153	0.306	0.171	0.300
Failed both math and reading in May	0.435	0.265	0.454	0.295	0.530	0.337
Retained	1.000	0.000	1.000	0.000	1.000	0.000

Assigned to transition center	0.002	0.000	0.018	0.000	0.974	0.000
Promoted	0.000	1.000	0.000	1.000	0.000	1.000
May Reading Score	4.481	4.724	6.201	6.681	6.228	6.661
May Math Score	5.152	5.444	6.969	7.189	6.826	7.104
August reading score	4.771	5.298	6.653	7.345	6.696	7.322
August math score	5.497	5.928	7.267	7.635	7.144	7.519
Educational attainment as of Fall 2005						
Graduated from CPS	0.296	0.404	0.385	0.479	0.256	0.327
Dropped out	0.572	0.527	0.610	0.516	0.742	0.668
Transferred to private school	0.036	0.035	0.032	0.021	0.026	0.025
Moved to another public school district	0.160	0.154	0.141	0.110	0.096	0.118
Still enrolled in the CPS	0.132	0.068	0.005	0.005	0.002	0.004
Age when started high school	15.218	14.909	15.143	14.393	15.256	15.228
Age when left high school	17.243	17.283	17.601	17.548	17.749	17.829
Number of credits completed when left high school	86.430	98.469	90.384	117.353	75.285	91.917
Age when dropped out	17.025	17.044	17.239	17.234	17.395	17.454
Number of credits completed in high school before dropping out	42.863	47.855	43.932	60.894	43.346	47.259
Observations	4,908	6,869	1,611	2,609	1,403	1,886

Notes: The statistics above reflect our analysis sample, which consists of students who were (a) enrolled in the sixth or eighth grades during the first three years of Chicago's accountability program (1997, 1998 and 1999), (b) took both the math and reading exams in the Spring, (c) were subject to the accountability policy, (d) failed the promotional cutoff and attended summer school, (e) took both the math and reading exams at the end of the Summer and scored between 1 grade equivalent below and 0.5 grade equivalents above the promotional cutoff, and (f) enrolled in the CPS the following Fall. For more details, see the text.

Table 2: The Effect of Grade Retention on the Probability of Dropping Out of School

		Dependent Variable = Dropped out of school		
		6th Graders	8th Graders, Younger than 14.4 years old in May	8th Graders, At least 14.4 years old in May
Estimation Strategy	Specification			
OLS	No covariates except for group dummies	0.039** (0.013)	0.079** (0.015)	0.066** (0.017)
	All covariates vary by group	0.027* (0.012)	0.045+ (0.026)	-0.004 (0.024)
IV, where excluded instruments include a connected spline that is allowed to vary across cohort and failure group	No covariates except for group dummies and the index score, which is allow to vary by group	-0.024 (0.035)	0.117** (0.034)	-0.029 (0.038)
	Same as above, but also includes student demographics and prior test scores, all of which are allowed to vary by group	-0.032 (0.038)	0.112** (0.034)	-0.011 (0.039)
Control group mean of the dependent variable		0.528	0.517	0.699
Observations		9,519	3,599	2,849

Notes: Sample includes all students in the analysis sample described in the notes for Table 1, but also drops students who leave the school district prior to the time when we can calculate their dropout status. The specifications shown in rows 1 and 3 include controls indicating the subject the student failed (math, reading, or both math and reading) and the year in which the student failed (1997, 1998, or 1999), these are interacted to create a total of nine different summer school groups. In row 3, we also control for the index score interacted with each summer school group. The specifications shown in rows 2 and 4 include the covariates described in rows 1 and 3 in addition to a host of prior achievement and student demographic characteristics described in the text. Huber-White standard errors clustered by index value x failure group are shown in parentheses. + significant at 10%; * significant at 5%; ** significant at 1%.

Table 3: Robustness Checks

		Dependent Variable = Dropped out of high school		
		6th Graders	8th Graders, Younger than 14.4 years old in May	8th Graders, At least 14.4 years old in May
	Estimation Strategy			
(1)	Baseline	-0.032 (0.038)	0.112** (0.034)	-0.011 (0.039)
(2)	Larger sample (-1.5 to +1.0)	-0.019 (0.032)	0.089** (0.025)	0.036 (0.031)
(3)	Larger sample (-2.0 to +1.5)	-0.003 (0.026)	0.068** (0.022)	0.037 (0.027)
(4)	Larger sample (-2.0 to +1.5), including 3rd order polynomial in the index score	-0.040 (0.037)	0.073* (0.030)	0.032 (0.034)
(5)	Larger sample (-2.0 to +1.5), include 3rd order polynomial in the index score that varies above and below the official cutoff	0.159 (0.100)	0.139** (0.041)	-0.029 (0.052)
(6)	Larger sample (-2.0 to +1.5), above cutoff terms are not excluded instruments – just use the slope in the marginal area	-0.002 (0.029)	0.074** (0.024)	0.035 (0.025)
(7)	Baseline sample, instead of a spline specification for the first stage, use a single binary variable indicating whether the individual passed the cutoff. The coefficient on the pass dummy is allowed to vary across cohort x failure group	-0.025 (0.037)	0.089* (0.039)	-0.010 (0.043)
(8)	Same as row (7), using a larger sample (-2.0 to +1.5) and including a 3rd order polynomial in the index score, which is allowed to vary by group and to vary above and below the cutoff	0.108 (0.099)	0.117** (0.041)	-0.031 (0.059)

Notes: Except where indicated, the sample is the same as described in Table 2 (sixth graders, n=9,519; younger eighth graders, n=3,599; older eighth graders, n=2,849). Except where indicated, we include the same covariates described in row 4 of Table 2. Huber-White standard errors clustered by index value x failure group are shown in parentheses. + significant at 10%; * significant at 5%; ** significant at 1%

Table 4: IV Estimates of Grade Retention on Changing Public School Districts or Attending Private School

	Sixth Graders			
	Changed public school district	Transferred to private school	Changed public school district or transferred to private school	Graduated from Chicago Public Schools
Retained in Grade	-0.004 (0.024)	-0.000 (0.011)	-0.005 (0.025)	-0.007 (0.034)
Control group mean	.154	.035	.189	0.328
Observations	11,777	11,777	11,777	11,777
	8th Graders, Younger than 14.4 years old in May			
	Changed public school district	Transferred to private school	Changed public school district or transferred to private school	Graduated from Chicago Public Schools
Retained in Grade	0.000 (0.023)	0.010 (0.011)	0.011 (0.024)	-0.099** (0.033)
Control group mean	.110	.021	.131	0.416
Observations	4,220	4,220	4,220	4,220
	8th Graders, At least 14.4 years old in May			
	Changed public school district	Transferred to private school	Changed public school district or transferred to private school	Graduated from Chicago Public Schools
Retained in Grade	-0.003 (0.028)	-0.007 (0.012)	-0.010 (0.034)	0.013 (0.040)
Control group mean	.118	.025	.143	0.280
Observations	3,289	3,289	3,289	3,289

Notes: The sample is the same as described in Table 2 except that we include students who leave the Chicago Public Schools prior to the time we can determine their dropout status. For this reason, the sample sizes are larger. We include the same covariates described in row 4 of Table 2. Huber-White standard errors clustered by index value x failure group are shown in parentheses. + significant at 10%; * significant at 5%; ** significant at 1%

Table 5: The Heterogeneity of Grade Retention Effects on Dropout Rates

	6th Graders	8th Graders, Younger than 14.4 years old in May	8th Graders, At least 14.4 years old in May
Baseline estimates	-0.032 (0.038)	0.112** (0.034)	-0.011 (0.039)
Year			
1997 Cohort	0.047 (0.062)	-0.008 (0.046)	-0.140* (0.059)
1998 Cohort	-0.104* (0.050)	0.144** (0.048)	0.037 (0.050)
1999 Cohort	0.018 (0.062)	0.346** (0.096)	0.076 (0.106)
Chi-Square Test of Equal Coefficients [p-value]	4.37 [0.11]	13.02 [0.00]	6.32 [0.04]
May Performance			
Failed Reading	-0.025 (0.062)	0.050 (0.043)	-0.058 (0.055)
Failed Math	-0.081 (0.084)	0.157 (0.120)	-0.059 (0.095)
Failed Reading and Math	-0.022 (0.054)	0.186** (0.059)	0.052 (0.059)
Chi-Square Test of Equal Coefficients [p-value]	0.39 [0.82]	3.68 [0.16]	2.12 [0.35]
School Quality			
School Is in Top Half of Reading Distribution	0.055 (0.068)	0.134* (0.068)	-0.177** (0.058)
School Is in Bottom Half of Reading Distribution	-0.070* (0.039)	0.120* (0.057)	0.056 (0.058)
Chi-Square Test of Equal Coefficients [p-value]	2.60 [0.11]	0.03 [0.86]	7.95 [0.00]
Race			
Black	-0.022 (0.036)	0.120** (0.040)	0.012 (0.057)
Hispanic	-0.080 (0.077)	0.019 (0.071)	-0.042 (0.075)
Chi-Square Test of Equal Coefficients [p-value]	0.47 [0.49]	1.80 [0.18]	0.45 [0.50]
Gender			
Male	-0.039 (0.049)	0.015 (0.039)	-0.096 (0.059)
Female	-0.009 (0.061)	0.181** (0.052)	0.049 (0.057)
Chi-Square Test of Equal Coefficients [p-value]	0.14 [0.71]	6.57 [0.01]	3.16 [0.08]
Family Income			
Free Lunch	-0.059 (0.041)	0.094* (0.040)	0.010 (0.052)
No Free Lunch	0.136 (0.129)	0.087 (0.157)	0.149 (0.167)

Chi-Square Test of Equal Coefficients	2.05	0.05	0.09
[p-value]	[0.15]	[0.83]	[0.76]

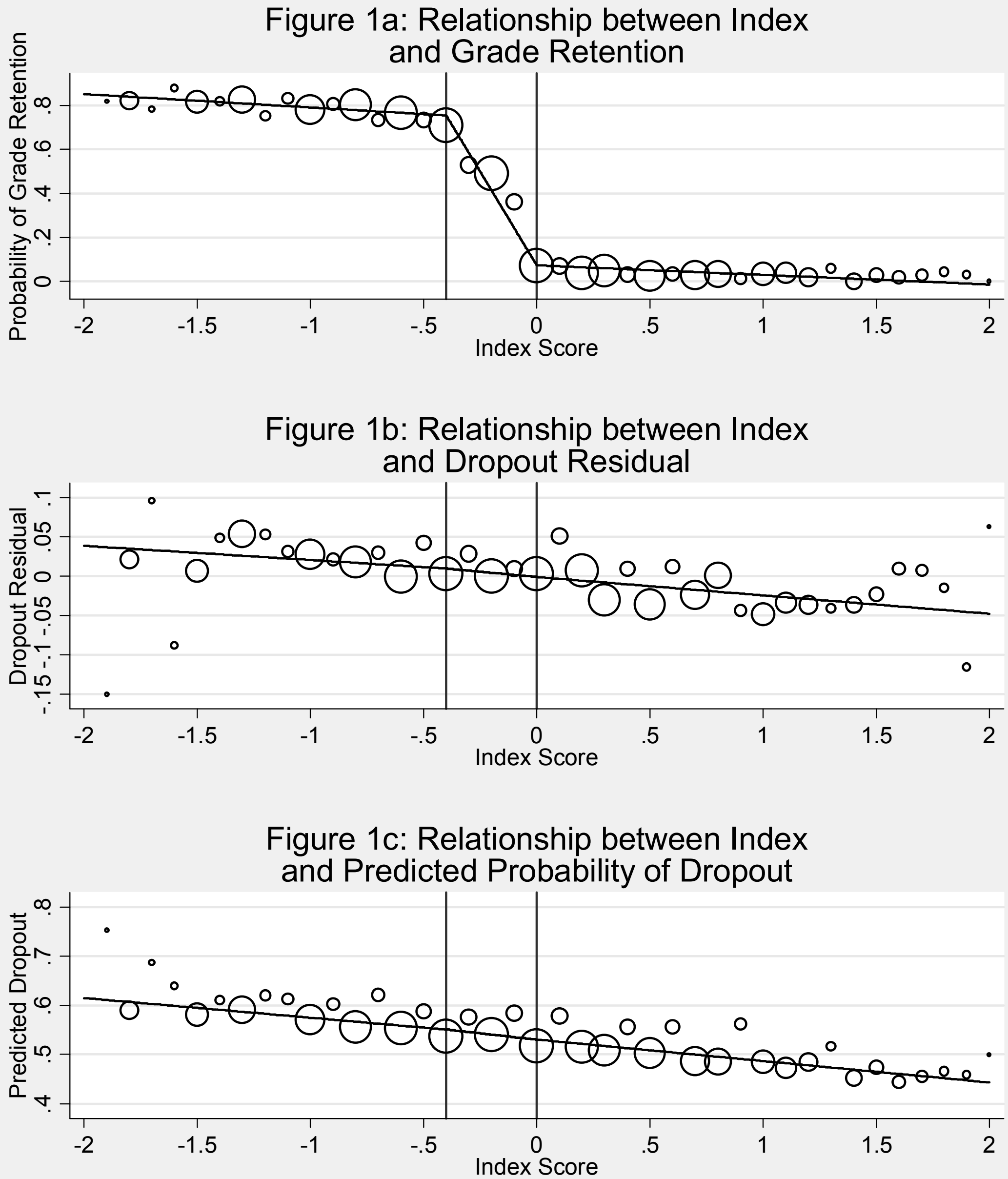
Notes: The sample is the same as described in Table 2 (sixth graders, n=9,519; younger eighth graders, n=3,599; older eighth graders, n=2,849). We include the same covariates described in row 4 of Table 2. Huber-White standard errors clustered by index value x failure group are shown in parentheses. + significant at 10%; * significant at 5%; ** significant at 1%.

Table 6: IV Estimates of the Impact of Grade Retention on the Academic Progress of All Students

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Enrolled 2 years later	Conditional on being enrolled 2 years later, in 8th grade or higher	Conditional on being enrolled 3 years later, retained in 8th grade	Entered high school in the CPS	Age began high school in CPS	Age left school (for any reason)	Total credits upon leaving school (for whatever reason)
Sixth Graders							
Retained in Grade	-0.004 (0.013)	-0.347** (0.045)	-0.088** (0.030)	-0.046 (0.030)	0.135** (0.052)	0.051 (0.104)	0.628 (4.260)
Control group mean	0.957	0.942	0.143	0.862	15.0	17.7	112.6
Observations	9,519	9,091	7,754	9,519	7,977	8,617	9,244
8th Graders, Younger than 14.4 years old in May							
Retained in Grade				-0.083** (0.020)	0.870** (0.038)	0.079 (0.103)	-27.107** (6.205)
Control group mean				0.971	14.4	17.8	125.8
Observations				3,599	3,358	3,582	3,516
8th Graders, At least 14.4 years old in May							
Retained in Grade				-0.145** (0.025)	0.738** (0.075)	0.161 (0.110)	3.736 (6.022)
Control group mean				0.941	15.2	18.0	97.4
Observations				2,849	2,424	2,840	2,769

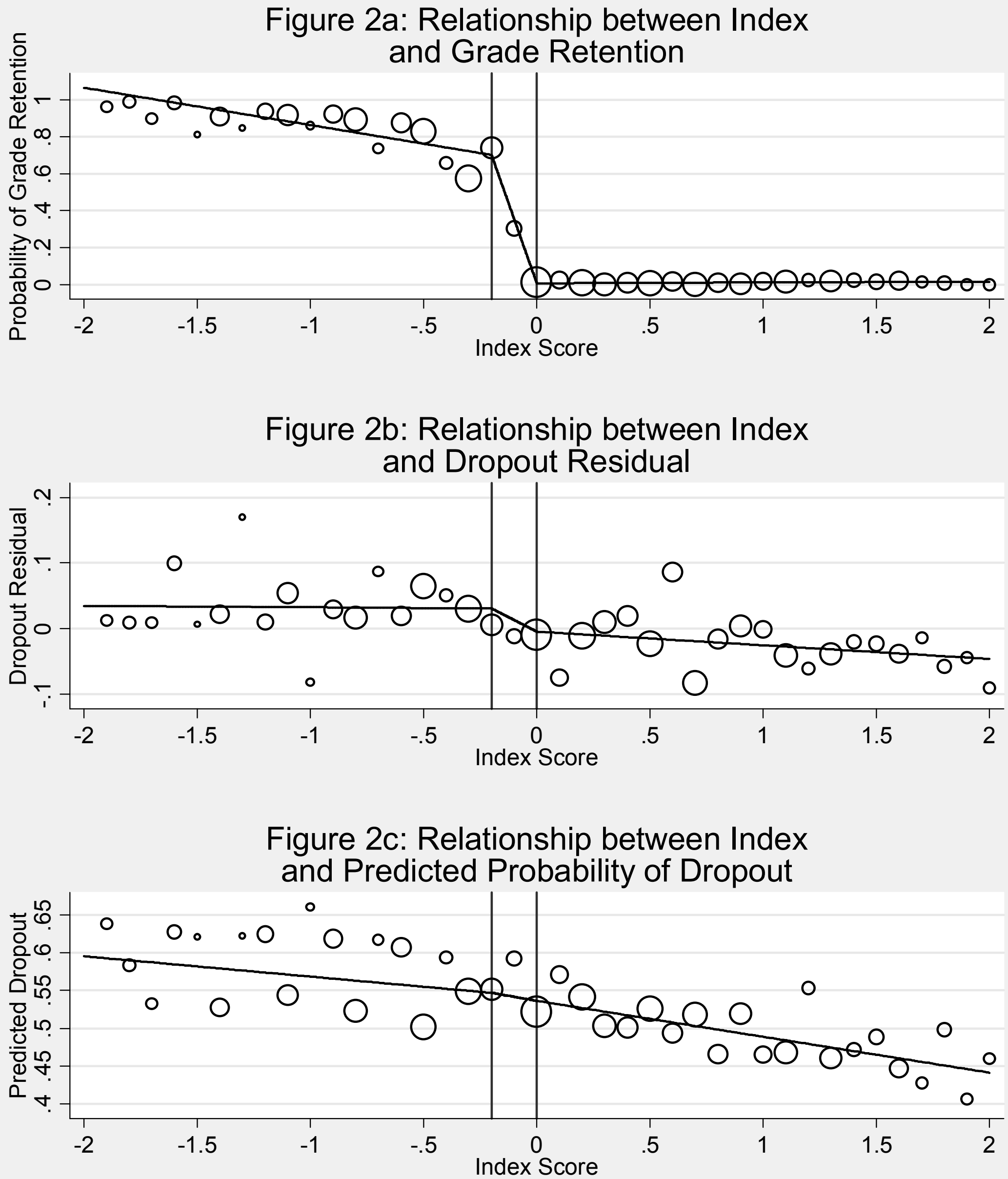
Notes: The sample is the same as the described in Table 2 except the sample sizes are in some cases larger because an outcome is observed even for those students for whom we cannot identify their dropout status. For this reason, the sample sizes are larger. We include the same covariates described in row 4 of Table 2. Huber-White standard errors clustered by index value x failure group are shown in parentheses. + significant at 10%; * significant at 5%; ** significant at 1%.

Figure 1: Sixth Grade



Notes: The number of observations at each point is proportional to the size of the bubble. The right vertical line is the formal cutoff and the left vertical line is the bottom of the marginal area. The solid line going through the raw data is a spline with knot points at the bottom of the marginal area and the cutoff. The spline was estimated using OLS. The predicted dropout rates and the dropout residuals are obtained through a regression of the dropout rate on student demographics, prior period test scores, dummy variables indicating the student's failure group (e.g. failed math in 1997), and the interaction of these dummy variables with the other covariates.

Figure 2: Eighth Grade-Younger Students



Notes: The number of observations at each point is proportional to the size of the bubble. The right vertical line is the formal cutoff and the left vertical line is the bottom of the marginal area. The solid line going through the raw data is a spline with knot points at the bottom of the marginal area and the cutoff. The spline was estimated using OLS. The predicted dropout rates and the dropout residuals are obtained through a regression of the dropout rate on student demographics, prior period test scores, dummy variables indicating the student's failure group (e.g. failed math in 1997), and the interaction of these dummy variables with the other covariates.

Figure 3: Eighth Grade-Older Students

Figure 3a: Relationship between Index and Grade Retention

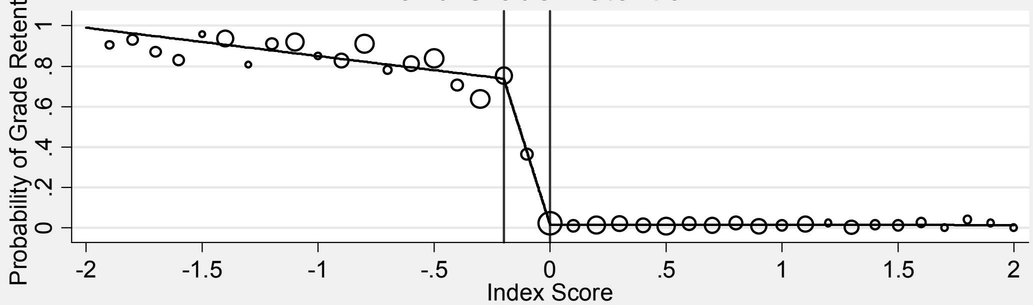


Figure 3b: Relationship between Index and Dropout Residual

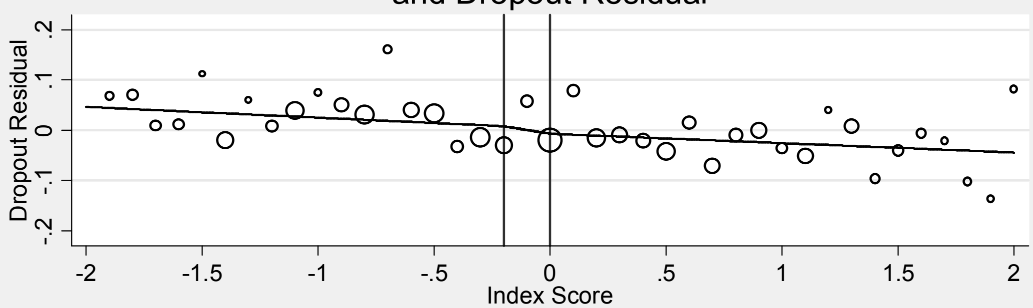
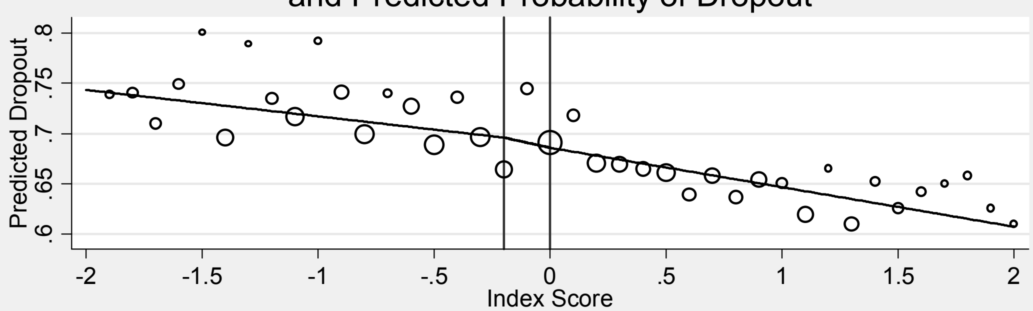


Figure 3c: Relationship between Index and Predicted Probability of Dropout



Notes: The number of observations at each point is proportional to the size of the bubble. The right vertical line is the formal cutoff and the left vertical line is the bottom of the marginal area. The solid line going through the raw data is a spline with knot points at the bottom of the marginal area and the cutoff. The spline was estimated using OLS. The predicted dropout rates and the dropout residuals are obtained through a regression of the dropout rate on student demographics, prior period test scores, dummy variables indicating the student's failure group (e.g. failed math in 1997), and the interaction of these dummy variables with the other covariates.