The Dissertation Committee for Matthew Scott Farber
certifies that this is the approved version of the following dissertation:

Essays on the Economics of Education of Underserved Populations

Committee:

Leigh L. Linden, Supervisor

Carolyn J. Heinrich

Brendan A. Kline

Jane Arnold Lincove

Richard Murphy

Essays on the Economics of Education of Underserved Populations

by

Matthew Scott Farber, B.A.; B.S.B.A.; M.S. Econ.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2016

To Laura and Asher

Acknowledgments

I would like to thank Leigh Linden, Rich Murphy, Dan Hamermesh, Jane Lincove, Carolyn Heinrich, and Brendan Kline for their invaluable assistance. I would also like to thank the Texas Education Resource Center, and specifically Celeste Alexander and Cindy Corn, for providing data. Lastly, I'd like to acknowledge Laura's patience with not only my extended schooling, but also with the many research questions I brought to her due to her expertise in the subject matter.

Essays on the Economics of Education of Underserved Populations

Publication No. ____________________

Matthew Scott Farber, Ph.D.
The University of Texas at Austin, 2016

Supervisor: Leigh L. Linden

This dissertation examines how current targeted accountability and funding provisions under federal guidelines impact the academic outcomes of the country's more underserved populations.[1] The first chapter demonstrates that accountability at the race level leads to increased reading and math achievement for students. I investigate the impact of school-level accountability on racial subgroups within a school, using a regression-discontinuity design with student-level Texas panel data on third through eighth graders from 2004 through 2011. The targeted incentives increase passing rates by 1-2 percentage points and the scores by .03 standard deviations in both math and reading. These results persist for two to three years after intervention, but fade out by the fourth year. Furthermore, students outside the targeted group are not hindered, with no effect on passing rates and scores. A deeper analysis suggests that schools are not focusing on high-leverage students but rather implementing wide-ranging interventions. I also find that the majority of gains are due to gains among Black students, though it is not clear whether this is due to racial targeting.

[1]The research presented here utilizes confidential data from the State of Texas supplied by the Texas Education Research Center (ERC) at The University of Texas at Austin. The author gratefully acknowledges the use of these data. The conclusions of this research do not necessarily reflect the opinion or official position of the Texas Education Research Center, the Texas Education Agency, the Texas Higher Education Coordinating Board, the Texas Workforce Commission, or the State of Texas. Any errors are attributable to the author.

In the second chapter, I analyze the impact of federally designed and funded interventions on student achievement, both of targeted students and non-targeted students. Under the No Child Left Behind (NCLB) act of 2001, schools with less than 40% low-income students use federal Title I funds for a Targeted Assistance Program, where schools above 40% are free to use those same funds as general school money. This paper uses a fuzzy regression discontinuity design around the 40% threshold with student-level Texas panel data on third through eighth graders from 2004 through 2011 to investigate. The evidence suggests that there is no difference in student outcomes, on the whole or among subsamples, between the methods of using the federal funding.

The third chapter shows that the impact of Title I funding on student achievement is complex, benefiting certain subgroups of students while impacting others negatively. I use an instrumental variable research design in order to estimate impacts while keeping external validity through exploiting the large data set available, which includes student-level panel data on Texas public school students from the years 2004 through 2011. While the instrument does not satisfy the exclusion restriction in this case, with certain assumptions, the estimates are useful as lower bounds on the true point estimates. Title I funding increases math passing rates by a minimum of 3 percentage points and has an impact of equal to or greater than 0 on reading passing rates and standardized scores for both math and reading. The IV estimates among elementary school students are negative in both math and reading, while lower-performing and low-income middle school students show large, though insignificant, effects of the funding on both math and reading exams. Unfortunately, this study cannot speak to the impacts on older students due to a weak first stage among high schools.

Table of Contents

List of Tables

List of Figures

Chapter 1

Targets of Opportunity: The Role of School-Specific Targeted Incentives on Student Achievement

1.1 Introduction

Accountability systems have been prevalent in the US education sector since the mid 1990s, and became federally mandated under No Child Left Behind in 2001. They are usually based on passing rates of statewide tests, and are designed to give stakeholders more information as to how each school and district is performing. By holding schools accountable for their results, the process is thought to be transparent, increasing focus on student achievement. One other aim of increased accountability is to close the achievement gap, as minority students have lagged behind their peers in academic outcomes for decades. Indications are that this gap has increased recently, but the reason is as of yet unknown. While there is a deep literature on the effects of increased accountability, there is very little research on how specific, targeted accountability affects outcomes for students. It is important to know whether targeting specific subpopulations can help increase the achievement of the underperforming students in our schools. In this paper, I use a race-based policy portion of No Child Left Behind to examine how added school-level incentives impact student performance for both targeted and non-targeted students.

Lawmakers at all levels are currently attempting to frame the next big education initiative to replace the Elementary and Secondary Education Act (ESEA), after the staggered shutdown of

No Child Left Behind (NCLB) in the last year or two. The NCLB Act was enacted in order to ensure all students achieved at acceptable levels, including minority and low-income students. The theory behind the act was that if schools were simply assessed on how all their students were doing, they would then rise to meet those standards, bringing up test scores of their lowest-performing students to do so. The incentive structure of the policies bundled together under the act was such that schools needed to increase the ratio of students passing each year and had to worry about the students at the bottom of the distribution to meet the preset goals.

Increased accountability has had effects both in how schools run and their corresponding results. States first started implementing their own accountability measures in the late 1990's, allowing researchers to use a difference-in-difference strategy to find impacts. In general, higher accountability standards within states were found to increase test scores (Reback 2008, Hanushek and Raymond 2005, Carnoy and Loeb 2002). The rollout of NCLB soon afterward gave us more evidence that increased accountability leads to increased student achievement (Dee and Jacob 2011), as well as evidence that parents respond to school performance when published (Hastings and Weinstein 2007). There is also evidence that schools respond to these incentives by focusing on students near the current performance standard (Neal and Schanzenbach 2010). It seems clear that giving schools targets and attaching consequences to those targets changes school practice and increases student performance.

The achievement gap in our education system has been persistent, and there is evidence that it has grown over time (Bailey and Dynarski 2011, Reardon 2011). Minority and low-income students have not performed up to the standard of their peers. Since the beginning of the movement to resolve this educational inequity, a variety of policies and initiatives have been implemented to attempt to solve the problem. They have ranged widely in their success, but the gap persists and remains quite large. These efforts have included small scale initiatives aimed at students, state-level policies, and nationwide legislation tying federal funds to school and district performance.

Little is known about the effects of targeting specific students with targets and consequences and how this impacts student performance. No Child Left Behind is about general accountability, but there are provisions that target specific groups of students in order to ensure schools are not leaving anyone out. Kane and Staiger [2002b] posited that the NCLB subgroup rules would have no effect on performance among minority youth, and Figlio et al. [2009] used the Florida accountability system pre-NCLB to examine the question, confirming the earlier study. However, the research using NCLB-era data to determine the impacts of these policies on students is largely lacking.

With increased accountability and the incentive to focus on specific students comes the question of how this impacts other students and whether there are trade-offs with such policies. If schools are truly at full productivity, logic suggests there must be some such side-effect. To date, the literature is inconclusive. Current research is based on what happens when a school fails to meet AYP criteria. Students that are at the bottom of the distribution seem to gain more than normal as schools focus on them, but how this impacts the other students is up for debate. Springer [2008] and Ahn and Vigdor [2014] find no effect on other students, but Krieg [2011] finds that the higher-performing students score lower than expected on subsequent tests. Because accountability under NCLB is defined as the fraction of students passing in a school, it is important to understand how this affects students who are not around the passing threshold for a given test, either far below or far above.

This paper joins the literature on school accountability, but is one of few to examine the impact of very specific incentives on student achievement. Prior research on targeted incentives in schools uses data from the low accountability era before the advent of NCLB, using a difference-in-difference strategy, as opposed to the regression discontinuity design featured in this paper. I also add to the inconclusive literature on school trade-offs, through an examination of impacts on specific types of students outside of the targeted subgroup. Furthermore, I am able to look at how long the effects persist and whether the persistence varies by student achievement level.

The structure of the policy, in which subgroups that comprise more than 50 students and 10% of the student body qualify for AYP subgroup criteria, lends itself to a regression discontinuity design. If one were to compare students in subgroups above and below the threshold, it would be a comparison of students in very different settings. The optimal experiment would be one in which incentives are randomized across schools, or across years within schools. To approximate this, I use a regression discontinuity, comparing students in subgroups immediately below the threshold to those immediately above the threshold. If the number of students right around the threshold is truly idiosyncratic, this yields causal results. To examine those schools where incentives materially change through the policy, I limit the sample to schools in which the subgroup in question is a minority of the student population.

I find that students in targeted subgroups pass the math and reading exam at rates 1-2 percentage points higher than those in non-targeted subgroups. These students score .03 standard deviations higher on the math and reading exams. There is no evidence that schools are focusing on specific subsets of students by prior performance, such as those near the passing threshold. Nor is there evidence that students in low-performing subgroups are impacted differently from those in high-performing subgroups. However, I do find that Black students benefit more from treatment than Hispanic or White students. Schools do not seem to be targeting specific students, so this is likely due to student body composition across schools more than race-based targeting within schools.

The increased achievement persists for several years after treatment. In math, the impact is significant and positive for three years after treatment, with effect sizes of 1.84 percentage points and .03 standard deviations three years later, but the impact fades out in year four. In reading, the effect size is 1.06 percentage points and .02 standard deviations three years later, though the estimated effect on scores is not significant at that point.

I also find no drop in achievement in science or social studies due to the increased incentives for math and reading. Schools do not appear to be taking away instructional time from these non-AYP subjects in response to the added incentives. Contrary to prior research, students not targeted in these schools do not see a decrease in their performance levels. The effect sizes for both math and reading in both passing rates and standardized scores are precise zeros, indicating that instructional resources are not being taken from non-targeted students to boost the scores of targeted students.

The current policy debate, with the rolling back of NCLB after the lapse of the ESEA, is about how much accountability is needed in the new education framework for the nation. This paper and its findings give evidence that added incentives for schools lead to significant results for students. The targeted students benefit, at no expense to the other students in the school, indicating no presence of a trade-off. These results lead to the conclusion that accountability for schools should increase, not decrease. Schools push to meet these targets, and increased achievement is the result.

The paper is structured as follows. Section 2 provides an overview of the policy whereby the targeted incentives are applied to schools and Section 3 introduces the data and research design. Section 4 presents estimates of the impact of the added incentives on the targeted and non-targeted students, and Section 5 discusses the implications of these findings.

1.2 Institutional Background

The No Child Left Behind (NCLB) Act of 2001 was aimed at improving the scores of all students, but especially the scores of the traditionally under-performing groups. Basing school Adequate Yearly Progress (AYP) scores on an overall average left the potential for these subgroups to continue performing poorly, masked by a larger population of higher-performing students. Thus, there is a clause that states that schools are not only rated on the overall passing percentage, but that of any qualifying subgroups: African American, White, or Hispanic students, as well as special education students, Limited English Proficiency (LEP) students, and economically disadvantaged students.

The application of No Child Left Behind varied across states. States were allowed to create their own AYP passing standards by year, and with respect to the subgroup cutoff, were able to define the specific number, as long as they were assessing proficiency of each of the defined groups. Thus, while the cutoff for subgroup qualification varies by state, the general rule is applicable across all states, and Texas' policies do not differ much from the rest of the country.

In order to meet AYP in Texas, a school must pass a certain percentage of its students in both math and reading. These percentages are designed to increase over time, hence the name. The standards began at very low numbers in 2004, just 33% for math and 47% for reading, and rose over time, to 75% in math and 80% in reading in 2011. (See Table A1 for more detail.) If a school contains qualifying subgroups, each of those subgroups must also meet the AYP passing standards. If any single qualifying subgroup does not perform up to standard, the school fails AYP for that year.

Qualification as a subgroup depends on the size of the school. For Texas schools with more than 2,000 students, a subgroup must contain at least 200 students to qualify for AYP purposes. For Texas schools with fewer than 2,000 students, a subgroup must contain at least 50 students and 10% of the student body to qualify for AYP purposes. A student may count for multiple subgroups, i.e, Hispanic, LEP, and special education. If a subgroup does not have enough students to qualify officially, those students' scores still are part of the overall passing rate, but are not considered separately.

Enrollment counts are taken on a specific October day each year, and determine to which school students are allocated. To be counted, a student must be at the same school from the formal October count date through the testing date in March or April. Students that transfer schools are not counted for AYP purposes at either the new or former school.

For a subgroup to officially qualify, it must be comprised of at least 50 students and 10% of the tested student body who have been enrolled from October through the testing date. Thus, the count for analysis purposes could be taken as the amount of students enrolled in October who test in April. However, in order to improve student achievement, schools need time to implement an intervention, and must use some estimate of future testing counts earlier on in the year. It is not clear how administrators go about making this prediction. For the purposes of this paper, I use fall enrollment counts. These counts are officially done by the state, and are important to administrators on a variety of levels, including funding, so they serve as the best proxy for school administrators' predictions. I plot the probability of official spring treatment as a function of fall counts in Figure A1 for reference. As shown, it is very unlikely for a school just above the fall threshold to qualify for spring treatment. However, it is administrator perception that matters in planning out interventions moreso than the actual spring count.

Sanctions begin after two consecutive years of not meeting the AYP standards and increase in severity if a school continues to underperform. The first sanction requires that the underperforming schools give students the choice to transfer to other schools in the district. Later sanctions include placement on a watch list and the implementation of a school improvement plan. By the sixth consecutive year of missing the target, schools must undergo a complete restructuring. (See Table A2, for more detail.) In addition to the official AYP sanctions, school administrators most likely face consequences from district officials for continual underperformance.

1.3 Data & Research Design

My analysis is based upon data from the Texas Educational Resource Center (ERC), including student-level demographic, attendance, and test score data. The demographic files contain information on age, gender, free or reduced lunch status, ethnicity, special education status, English as a second language (ESL) status, limited English proficiency (LEP) status, and gifted and talented status (GT). The attendance files contain information on enrollment for each student across each school year, including if and when a student may have transferred mid-year.

The test score data is from the Texas Assessment of Knowledge and Skills (TAKS) exams, developed by Pearson, and conducted in the springs of third grade through eleventh grade. Students are tested in both math and reading every year, though in science, only in grades 5, 8, and 11, and social studies, only in grades 8 and 11. Writing is the sole exam subject not fully consisting of multiple choice questions and is assessed in grades 4 and 7. For the later years in the sample, students in grades 5 and 8 are required to pass both the math exam and the reading exam in order to be promoted, but have up to three attempts to do so. All public school students are required to take the TAKS exams unless they have a severe disability. Those with moderate disabilities are given accommodations, but still take the exams. In my analysis, the test scores are normalized to have a mean of zero and a standard deviation of one for each grade and year across the entire Texas sample.

Students were matched across these files both cross-sectionally and longitudinally using de-identified student ID numbers. Using the fall enrollment dates and the testing dates, I then constructed school-level variables, including gender, ethnicity, FRL, special education, ESL, LEP, and gifted/talented ratios. These variables are used as additional controls throughout the analysis. I also constructed a student count variable, both by race and overall school membership, for use as the running variable. I construct separate count variables for fall and spring to examine their differing effects.

Starting with all students in Texas that match across these files, a pooled cross-section of 16,820,061 students, I drop any schools exempt from AYP considerations, leaving 16,446,768 students. I then pare the sample down to the most relevant students for analysis. The NCLB policy mandates schools test students in grades 3-8 and one high school grade that states are free to choose. To make the results more generalizable, I exclude high schools, leaving me with students from elementary and middle schools tested in grades 3 through 8. This leaves 13,952,455 students.

Furthermore, I only include elementary and middle schools with more than 150 students. This is so that the students in the analysis present a minority of the school population. Comparing a subgroup of 49 students to one of 50 students will not matter if both schools have 60 total students to begin with as the group matters too much to the overall passing ratio for the subgroup policy to become relevant. This restriction leaves 13,370,364 students.

Due to the nature of a regression discontinuity design, I then restrict the sample to the relevant bandwidth; in this case, five. While I present basic results for bandwidth of ten, five, and three, throughout the paper I continue using a bandwidth of five, following the process of cross validation found in Imbens and Lemieux [2008]. Some rounding is applied due to the more discrete nature of the running variable in this research design. The final sample includes 239,464 students across 4,821 campus by race by year groups. Further detail on sample size due to my restrictions is available in Table A3. I also present results showing that my estimates are robust to sample restrictions in Table A4.

I present both student-level and school-level summary statistics in Table1.1, both for the relevant demographics and the baseline testing measures. As one might hope in an RD design, both students and the schools they attend are similar, above and below the threshold. The student- and school-level means are very close on either side of the threshold in both demographic characteristics and baseline test scores.

1.3.1 Research Design

The purpose of this paper is to determine how schools react to extra incentives, and whether this translates to student scores. Thus, I attempt to estimate the impact of subgroup qualification on student test achievement. However, using ordinary least squares (OLS) most likely will not give a causal effect of subgroup qualification on student outcomes due to the differences in school composition between the treatement and control groups. For this reason, I use a regression discontinuity approach to exploit the provision (described in detail earlier) whereby a subgroup's scores count for AYP if that subgroup has a minimum of 50 students and do not count if the subgroup consists of fewer than 50 students.

The RD design, in this case, simply compares student outcomes in racial subgroups of more than 50 students to those of fewer than 50 students for White, Black, and Hispanic student subgroups. With a large enough sample size, the comparison would be between subgroups with 49 students and subgroups with 50 students, but for precision, I expand the group to a bandwidth of five rather than one. However, this allows for the possibility of a relationship between the number of students in the subgroup and outcomes, so I control for the number of subgroup students with a separate function on either side of the threshold.

In a perfect experiment, subgroup qualification would be randomized within the bandwidth noted; another possibility is that treatment is randomly turned on and off for each subgroup within a school over time. To approximate this second version, I use school by race fixed effects. This also helps control for any long-standing performance differences between races within a school. One might be concerned about how many "switchers" there are in the sample. A very small amount would mean the results are based on a small sample of schools and not representative of the population. However, the large majority of school by race clusters appear on either side of the bandwidth over time, and thus the sample size is not a concern. I therefore estimate the following equation to study the impact of subgroup qualification on student achievement:

$$y_{isrt} = \alpha + \beta Qual_{rst} + f(Count_{rst}) + f(Count_{rst} * Qual_{rst}) + X_{isrt}\pi + Z_{sr}\gamma + T_t\eta + \epsilon_{isrt} \quad (1.1)$$

for student i in school s of race r in year t. $Qual$ is an indicator of the student's racial subgroup qualification status, $Count$ is a count of the number of students in the student's racial subgroup, X is a vector of student and school-level characteristics, Z is a vector of school x race fixed effects, T is a vector of time fixed effects, and ϵ is the error term. In the basic results, the function of count on either side of the threshold is simply a linear spline, but a quadratic spline is used for a larger bandwidth in a separate specification. Estimates are clustered at the school x race x year level, which is also the level of treatment.

1.4 Results

1.4.1 Validity of Regression Discontinuity Design

The first assumption that must be satisfied under a regression discontinuity design is a lack of manipulation of the running variable. Because there is no real advantage to having a subgroup qualify for AYP criteria, there is an incentive for school administrators to manipulate the count of students in a given race to ensure that the official count is a number less than 50. For this paper, I use fall counts as the running variable, but the count that actually qualifies is the spring count of students tested. I use fall counts because I believe that this is what administrators base their educational decisions on for the purpose of increasing student test scores to meet the added criterion. This gives them all year to intervene in one manner or another, and is generally a good predictor of number of students tested in the spring.

To test for manipulation in the fall counts, I provide a visual in Figure 1.1, plotting the distribution of the count of students in racial subgroups, using a quadratic specification on either side of the threshold. Evidence of manipulation would be shown through a discontinuity in the density around the cutoff. If there were manipulation, one might see a drop in the number of racial subgroups with a qualifying count of 50 or 51, and a corresponding increase in the number with counts of 48 or 49. As shown, such manipulation is not present. On the other hand, there is a small and statistically significant jump in the number of subgroups with 50 students. However, because there is no evidence that indicates an advantage to having a subgroup qualify, I believe there is no manipulation of student counts around the threshold.

However, there is another method by which manipulation is possible, whereby school administrators take notice of the fall counts and take action to make sure that the spring testing counts are below 50, thus disqualifying a racial subgroup from qualification. This could be done through suspensions, or strategic student movement between campuses, or a variety of other methods. In Figure 1.2, I plot the percent of students tested in the spring by the number of students in a subgroup present in the fall. If this type of manipulation were happening, there would be a negative discontinuity in the percent tested at the threshold. However, no such discontinuity appears. There is no statistically significant discontinuity in either direction at the threshold, though the coefficient would appear to be positive, if anything.

The second assumption is that the cutoff affects the students involved only through the treatment; in this case, subgroup qualification. Thus, I check whether the students and the schools the students attend are different on observable characteristics across the 50-student threshold. I run the base specification without any controls with each of the student-level and school-level covariates as an outcome. These are displayed in Figures A2 and A3, respectively. None of the covariates shows a statistically significant discontinuity.)For more detail, see Panel A of Table A5 in the Appendix.) Note that the covariates included are those at the student level and those at the school level. As another check, I do the same for prior year test scores for the sample of students that were tested the previous year. The sample is smaller, of course, but none of the prior scores or passing rates show a significant jump at the threshold, either. These results are also included in the Appendix, in Panel B of Table A5. Due to the lack of discontinuities in either the student-level covariates, school-level covariates, or baseline test scores, I am confident in the interpretation of observed discontinuities at the 50 student threshold as due solely to the AYP qualification and not due to sample composition on either side of the threshold.

1.4.2 Impact of Subgroup Qualification on Students in Subgroup

The main results show that the targeted incentives have a positive impact on student achievement. Table 1.2 presents the basic results for the bandwidth of 5 students for both math and reading, and Figures 1.3 and 1.4 presents the graphical evidence. Students in subgroups that qualify for AYP status pass at a rate of 1.17 pp higher in math and 1.34 pp higher in reading, scoring about 0.03 standard deviations higher in both math and reading. In the basic sample, there are 239,464 students across 4,821 clusters at the race-by-year-by-campus level.

Table 1.3 presents the main coefficient for varying bandwidths, polynomial specifications, and control variables for the math exam. Results are presented with a linear specification, with the slope allowed to differ across the threshold, at bandwidths of three, five, and ten. A quadratic specification is presented at the bandwidth of ten. For all of these specifications, results are presented with no controls, student demographics, student and school demographics, and all demographics plus year fixed effects.

The coefficients for the math results are positive for both passing rates and scores across all sixteen specifications. The impact sizes are similar among the three bottom rows, excluding the model with a bandwidth of ten and a linear spline. Once all control variables are included, the majority of the results across these three model types are significant at the 10% level. It seems that the specification is not the driver of the results in this case.

Table 1.4 is organized in a similar manner. However, the coefficients are not as consistent as those from the math exam. At a bandwidth of ten, regardless of the polynomial or the controls included, the pass rate and score discontinuities are not significant, and in some cases, are actually negative. However, when the model is restricted to a smaller bandwidth, the effect sizes and their significance are remarkably similar. I am confident in the main result presented for both reading pass rates and reading scores, as well as for the math results.

As stated earlier, I check for the robustness of my results to the sample restrictions in Table A4, and show that the results do not vary much across samples. I also run the main specification on various placebo thresholds, shown in Table A6, from 35 to 65 by fives, and show that 50 is the unique threshold with significant and positive results.

The magnitude of the results, while positive and significant, is somewhat small. However, the important piece is that schools are responding to these incentives, and that student scores are increasing as a result. As long as schools are not losing sight of other students in the process, these results are a signal that targeted incentives do help students.

1.4.3 Test Administration and Count Timing

As stated earlier, the count that determines subgroup qualification for AYP purposes is actually the spring count of students tested. However, at that point it is too late for administrators to intervene to improve student test scores if necessary. Thus, I use the fall student enrollment count as the running variable through the analysis. However, in Table 1.5, I present results for both fall and spring counts, as well as for the first and last test taken by students. These might be different for various reasons: students moving in or out of the district, students transferring between schools within the district, students being absent, etc.

In grades five and eight in the later years of the NCLB regime, students were required to pass both the math and the reading exams in order to progress to the next grade[1]. They were given

[1]for more info, see http://tea.texas.gov/student.assessment/ssi/

three chances to do so. Thus, some students took the test just once, while others took the test three times. Administrators with subgroups including students in these grades might wait for the first round of test results to come in before acting. If that is the case, we would expect to see a stronger result using the last test for each student than using the first test.

For each outcome in the table, the first column shows the effect size using fall counts while the second column shows the effect using spring counts. The results using the spring counts as the running variable in Table 1.5 are insignificant for all outcomes measured, unlike the results using fall counts. It seems that administrators use the fall counts as a predictor of spring counts and act accordingly. This is logical, as the spring counts are unknown until the testing day, when it is too late to intervene. Administrators possibly use more complex predictions, but it is clear from this table that those predictions are much closer to the fall counts than to the spring counts.

The top row of the table shows effect sizes using the last test taken for each student within a subject, while the bottom row shows effect sizes using the first test taken. Results using the first test and those using the last test are not statistically different in any of the four outcomes. Thus, it appears that any interventions implemented by school administrators are coming during the school year, and are not ramped up after the first test results come back. This lack of action may be because of the timing of the tests (each round is a month apart, and results take a couple weeks to return), or may be because only a very small portion of students take a test more than once.

1.4.4 Impacts by Performance Levels

1.4.4.1 Students

Schools have quite a bit of information on each student for tested subjects. At a minimum, they have grades and prior year test scores. Many schools benchmark students, giving a series of tests aimed at mimicking the state exams as best as possible throughout the year in order to provide a better picture of where each student is academically. Either way, schools have an adequate picture of where each student stands in relation to the year-end exam. One might be interested, then, in which students schools target during interventions implemented in order to improve racial subgroup scores above the threshold.

The simplest intervention would be to give extra instruction to the lowest-performing students. Because 83-89% of students pass the state exams, the lowest quartile would be the easiest place to increase the passing rate, the only metric that matters in AYP calculations. Thus, in Table 1.6, I present basic results as well as results by quartile. I define quartiles by using each student's standardized state exam score the prior year. I then interact each quartile indicator with each independent variable, allowing all covariates to vary by quartile. The sample size is smaller in this table because it excludes the first year of the sample, any third graders, and students who did not sit for the exam the prior year, as they do not have prior year scores.

For this sample, the discontinuity in the math passing rate is 0.64 pp, while for reading it is 1.56 pp. The discontinuities in score are 0.025 standard deviations for math and 0.039 standard deviations for reading. These results are similar to those presented in Table 1.2. When examining results by quartiles, there is absolutely no pattern among the math results. The jumps by quartile are statistically similar, and I test for the equality of these four coefficients, showing the p-value for these tests in the bottom of the table. For math, the p-values are 0.503 and 0.298 for passing rates and scores, respectively, as expected. Among the reading results, it appears that the lower two quartiles benefit the most, with passing rate jumps of 1.82 and 1.56 pp for the first and second quartiles, respectively, and score jumps of 0.034 and 0.047. The p-values for equality of the four coefficients, however, are very high. The evidence for interventions focusing on the first quartile is tenuous at best.

To double-check, I split the sample by number of questions each student was from passing the previous year, and allow for the impact to vary across these subsamples. I present results in Figures 1.5 and 1.6. Each point on the plots is sized according to the number of students within the cell. There is no obvious pattern present in the plots for any of the four outcomes, other than more students passing the exams than failing them the prior year.

These results seem to disqualify the hypothesis that school administrators would simply focus on the lowest-performing students in order to increase the passing rate with the least amount of inputs. The results persist across all quartiles, and are certainly not monotonically decreasing in quartile. The question then remains as to what types of interventions are happening and if students of all abilities are benefiting from the added incentive.

1.4.4.2 Subgroups

Similarly to schools intervening more for certain students, one would expect larger impacts on subgroups around the passing threshold. If a racial subgroup the prior year all passed the exam easily, then there is no incentive for a school to add in extra instruction or give tutoring. However, if the subgroup the prior year performed very badly, then one might expect more interventions and larger impacts. Only the subgroups in the lowest quartile are really in danger of failing to meet the AYP standard.

I split the subgroups into quartiles by prior year's performance in order to analyze whether students in lower performing subgroups benefit more from the treatment. Results are presented in Table A7. As before, the sample sizes are different from the basic results. If a new school opens, there won't be any scores from the prior year. There are very few of these instances, though.

The results indicate the impact of the subgroup qualification does not vary by the subgroup's prior year performance. Effects may be stronger for the highest quartile, but the difference across quartiles is not statistically different except for a marginal difference in effect sizes for math scores, as I test for the equality of the coefficients. This is yet another insight into administrator planning, and suggests that perhaps the interventions implemented are not focused, but rather wide-ranging.

1.4.4.3 Rest of School

If the rest of the school performs extremely poorly, then there is no incentive to increase scores for a particular subgroup. Even if that subgroup passes, the school won't be meeting AYP regardless. On the other hand, in schools where the rest of the school performs quite well, if only one subgroup might hold them back, there is a large incentive to increase scores for those particular students. To examine whether this is true, I split the subgroups into quartiles by the performance of the rest of the school in the previous year and run the model for each quartile separately. (Results are found in Table A8 of the Appendix.) I find that the effects differ across quartiles for math scores, but not for any of the other three outcomes.

There are also many possibilities for the relationship between the subgroup in question and the rest of the school that may make such analysis difficult. These two are likely highly correlated, which means that subgroups in higher-performing schools may not be as low-achieving as their counterparts in overall low-performing schools. They may already be achieving at levels above the school passing threshold, and as such, no intervention would be necessary.

1.4.5 Results by Race

Given that the results thus far suggest that treatment leads to some type of intervention, but that the intervention impacts all students somewhat equally, without a focus on low-performing students or subgroups, there is no reason to believe that students of any one race will benefit more than those of another race. However, given the race-based impetus for the policy, it is prudent to split the results by the three races included in the policy and analyze results.

In Figure 1.7, I present mean math passing rates and scores by fall count, split by race. White students perform the best, then Hispanic students fall in the middle, with Black students performing far below the other two. There is a positive impact on both passing rates and scores for White students, while the impact for Hispanic students seems to be very small, if present at all. However, the impacts on passing rates and scores for Black students is quite large, especially in comparison to the other two. Figure 1.8 displays similar results for the reading exam, and shows a very similar pattern. White student see a small positive impact, Hispanic students see almost no impact, and Black students see a very large impact.

I formalize these results in Table 1.7. The first column for each outcome displays the joined impact, while the second displays coefficients from a regression in which I interact race with each covariate, allowing all coefficients to vary by race. The impacts across the four outcomes for Black students are the only significant coefficients, but just as importantly, are much larger in magnitude. On the math exam, the passing rate impact for Black students is 2.00 pp and the score impact 0.052 standard deviations, while the maximum of the other two races is 0.92 pp for passing rate and 0.026 standard deviations for score. The gap isn't quite as large for reading, but is still present.

Black students in the sample have lower test scores, and show the largest response to treatment. However, I have shown that schools are not targeting low-performing students or low-performing subgroups. The larger response is most likely not due to race-based targeting within schools. What, then, is driving this result? It is possible that, given the lower scores, school administrators see Black subgroups surpassing the qualification threshold as a bigger threat to AYP status than other subgroups, and respond in kind. It is also possible that, for some reason, Black students simply respond better to the interventions used in schools than the other students. Unfortunately, it is not clear in this analysis exactly why the response is higher.

1.4.6 Long-Run Results

A natural question is whether the results presented persist past the year of treatment. The math and reading tests from one grade to another cover some different material, but much of the content is based on the same general objectives. Thus, even if there are long-run results, it is not proof of true growth or any evidence that schools are not simply focusing on test-taking strategies. Even such strategies could show persistent results.

In Panels A and B of Table 8, I present long-run results for achievement levels and achievement growth, respectively. I restrict the sample to students for whom testing data exists up to four years after their subgroups are within the relevant bandwidth. This rules out any students from the original sample who were in sixth grade or above or students from the years 2008 or later. In the growth sample in Panel B, I further restrict to only students with a test score in the prior year. In both panels, each coefficient presented is from a different regression, and includes a variable for treatment in each year prior as well as treatment interaction terms. Thus, the coefficient presented is for those students who have treatment in year zero, but not in any year afterward, as compared to those without treatment in any year.

In Panel A, there are positive long-run results, persisting three years out for math and somewhere between two to three years out for reading. Students in qualified subgroups not only achieve at higher levels in the year of treatment, their improvement continues for about three years afterward, though the increased achievement fades over time. The increase in math passing percentage varies a bit, but the other outcomes show more consistency. Math scores are increased .040 standard deviations in the year of treatment, and decrease to .030 standard deviations three years later, though still significant at the 10% level. Reading scores show a less consistent pattern, where an increase of .036 standard deviations in the year of treatment declines to 0.019 one year later and jumps back up to 0.036 standard deviations two years out. There are two possibilities for how these higher achievement levels persist. The first is that schools keep the same interventions used in the year of treatment for two to three years afterward, continuing to raise scores. The second is that these students see a one-time increase in scores, and stay on that higher trajectory for awhile afterward.

To differentiate between the two, I create the exact same table in Panel B, but for achievement growth rather than levels. If schools are continuing with the same successful interventions, growth will be higher each year, whereas if it is more of a one-time increase in scores, there will only be an increase in growth in year zero and not beyond. The second hypothesis is supported quite strongly by the results from Panel B. In reading, there is achievement growth in the year of treatment, and this fades out by one year later. In math, there appears to be a large jump two years out that I cannot explain, but there is score growth in the year of treatment that falls to almost zero one year later.

Student achievement is impacted strongly in the year of treatment, and the affected students are shifted on to a higher learning curve through the school's interventions. They stay on this higher curve for two to three years afterward. However, this is much more of an intercept change than a slope change. Student test scores are not growing differentially beyond the year of treatment, indicating that schools are not continuing interventions aimed at these students. As before, curiosity remains as to which students benefit the most in the long-run. Are the students most in need showing persistence over several years? I have run specifications on long-run results by quartiles, but there are no discernible patterns as to which quartiles have more persistent or stronger results.

1.4.7 Spillover Effects

1.4.7.1 Impacts on Students Outside Subgroup

Schools may intervene in order to increase passing rates of specific racial subgroups due to increased pressure, but it is quite difficult to target only students of one race with an intervention. More likely, either students are all given the intervention, or they grouped by performance and targeted that way. Thus, results for other students could increase in tandem with those of the subgroup in question. In Table 1.9, I present results on other students in the schools where subgroups are in the bandwidth around the threshold. The sample sizes are much higher here, because I have focused in general on subgroups that comprise a minority population in the school in question. Because I constricted the sample to schools in which the treated subgroup is a minority, almost all of the non-treated students are part of a qualified subgroup, but well above the threshold. For the sample presented, the coefficients for all four outcomes are precise zeros and are not significant. It is apparent that the schools do not take resources away from non-targeted students to focus on the targeted students. Some schools may be using wider-ranging interventions, but perhaps not enough to affect the overall effect sizes.

In the second column for each outcome, I bin students into quartiles based on their prior year exam scores and interact those indicators across all other independent variables. In math, we see the largest results for the fourth quartile, with a pass rate increase of 0.76 pp and a score increase of 0.024 standard deviations. The coefficients for the lower three quartiles are much smaller. In reading, the first quartile students see a smaller effect than the other three, but there is not a larger pattern at play. This lack of pattern again among the quartiles is fully in line with the prior results. It would be strange for students outside the treated subgroup to have disparate impact among quartiles if the treated subgroup students did not. Schools do not seem to be targeting any specific academic group with interventions, and this holds for students outside of the subgroup as well.

1.4.7.2 Impacts on Non-AYP Subjects

Students are tested in science in grades five and eight and in social studies in grade eight. These exam scores are not considered for school AYP consideration nor for student progression. However, these subjects use elements of math and reading, and if students have truly grown in those subjects, they might be expected to perform better on these tests as well. In Table 1.10, I present these results. On the science exam, the passing rate coefficient is -0.18, though not significant, and the score coefficient is an insignificant -0.015. In social studies, the pass rate coefficient is an insignificant 1.21 and the score coefficient is 0.006 standard deviations. The standard errors for these results are quite high compared to the prior results due to the much smaller sample sizes. There does not appear to be any effect on test scores in these non-incentivized subjects. This could be because any score growth due to growth in reading and math is outweighed by time taken away from these subjects for math and reading classes, or because neither of these two mechanisms is occurring.

1.5 Discussion

Closing the achievement gap is one of the principle aims of the No Child Left Behind legislation. One of the main components to do so is the use of subgroup incentives within schools. However, the research on the effectiveness of these incentives is lacking. There has been plentiful research into how general accountability affects students, but the effects of targeted accountability incentives are unclear. Furthermore, the progress in closing the achievement gap has been slow, and information on policies that may help would be useful.

Added race-based incentives increase academic achievement on both math and reading scores, both in the passing rate and the actual scores. This added achievement does not seem to be concentrated among lower-performing students, nor among lower-performing racial groups, nor even among the lower-performing schools. It is concentrated among Black students, as opposed to White or Hispanic students. The effects persist throughout all populations, independent of prior performance. These effects are not only present in the treatment year, but persist for two to three years afterward, even if the incentives do not persist due to a falling student count for the race in question. Increased performance in the treatment group does not come at the expense of students outside of the treatment group, with the non-targeted students showing no effect.

The natural question arises as to what schools are doing in order to achieve these results. The simplest type of intervention to raise the passing rate for the treated group of students would be to focus on the lowest-achieving students, in the form of extra instruction or small group tutoring. However, this is not borne out in the results, where the highest performing students show similar achievement increases to the lowest-performing students. Also, schools cannot plausibly focus on one race of students, even if that is where the incentive lies. Because of the broad range of students and schools where positive results are shown, it seems that any intervention implemented is at a school-wide level. The long-run achievement growth results show that these interventions do not persist past the year of treatment. The fact that effects based on fall student counts are significant and positive, while those based on spring student counts are statistically zero, lends credence to this well. As to what these larger academic changes are, this paper does not have any methods for deciphering. Perhaps schools are simply adding more instructional time in math and reading, a common instructional intervention. In the future, it may be worth investigating what strategies were used, perhaps to determine how to further promote effective treatments.

Schools are responding to race-based incentives with some type of change that increases student test scores, regardless of prior individual or school performance. If schools were at full productivity, we would expect to see a drop in the achievement of other students in the school. However, instead we see no change in reading and math scores for the other students. This indicates that schools are not operating at full productivity, and that these incentives perhaps push them in that direction, making them a useful policy lever.

Tables & Figures

Table 1.1: Summary Stats

	Student-Level		School-level	
	Below Threshold	Above Threshold	Below Threshold	Above Threshold
Demographic Variables				
Ratio Male	0.503	0.503	0.512	0.512
	(0.500)	(0.500)	(0.031)	(0.031)
Ratio FRL	0.569	0.575	0.525	0.539
	(0.495)	(0.494)	(0.259)	(0.250)
Ratio Special Education	0.084	0.081	0.114	0.114
	(0.277)	(0.273)	(0.042)	(0.044)
Ratio ESL	0.096	0.097	0.061	0.063
	(0.432)	(0.434)	(0.071)	(0.073)
Ratio Gifted	0.082	0.083	0.093	0.090
	(0.267)	(0.273)	(0.097)	(0.088)
Ratio Black	0.420	0.401	0.161	0.172
	(0.494)	(0.490)	(0.156)	(0.161)
Ratio Hispanic	0.351	0.354	0.367	0.375
	(0.477)	(0.478)	(0.250)	(0.238)
Ratio White	0.229	0.245	0.423	0.407
	(0.420)	(0.430)	(0.265)	(0.254)
N	104,615	134,849	2,124	2,457
Baseline Scores				
Math Pass Ratio	0.820	0.817	0.805	0.813
	(0.384)	(0.387)	(0.396)	(0.390)
Standardized Math Score	-0.077	-0.089	-0.086	-0.099
	(0.994)	(1.002)	(0.398)	(0.397)
Reading Pass Ratio	0.886	0.883	0.877	0.882
	(0.318)	(0.322)	(0.328)	(0.323)
Standardized Reading Score	-0.014	-0.026	-0.029	-0.040
	(0.974)	(0.983)	(0.356)	(0.354)
N	73,068	94,184	2,110	2,439

Notes: Below the threshold refers to students in subgroups or schools with subgroups with 45-49 students, while above the threshold refers to students in subgroups or schools with subgroups with 50-54 students. Statistics displayed are the mean, and below each mean, the standard deviation in parentheses.

Table 1.2: Basic Results

	Math		Reading	
	Pass %	Score	Pass %	Score
Subgroup Qualifies	1.17**	0.032***	1.34***	0.034***
	(0.50)	(0.012)	(0.36)	(0.010)
Student demographics	✓	✓	✓	✓
School demographics	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
N	239,464	239,464	239,464	239,464
r^2	0.177	0.325	0.128	0.302
Mean	82.16	-0.084	89.00	-0.020
Clusters	4,821	4,821	4,821	4,821

Notes: Each column is from a separate regression with a linear function of the count on each side of the threshold, allowing the slope to vary on either side. All regressions include year, grade, and school-by-race fixed effects, gender, race, FRL status, special education status, and LEP status at the student level, as well as gender, FRL, special education, LEP, and student count at the school level.
*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table 1.3: Specifications – Math

	Passing Rate				Score			
Bandwidth 10, linear spline								
Subgroup qualifies	0.29	0.24	0.12	0.24	0.009	0.005	0.005	0.007
	(0.36)	(0.37)	(0.35)	(0.33)	(0.008)	(0.008)	(0.008)	(0.008)
	450,270	450,270	450,270	450,270	450,270	450,270	450,270	450,270
Bandwidth 10, quadratic spline								
Subgroup qualifies	0.91	1.15**	0.82	0.91*	0.016	0.023*	0.019	0.019
	(0.58)	(0.58)	(0.56)	(0.52)	(0.013)	(0.013)	(0.013)	(0.012)
	450,270	450,270	450,270	450,270	450,270	450,270	450,270	450,270
Bandwidth 5, linear spline								
Subgroup qualifies	1.26**	1.51***	0.93*	1.17**	0.030**	0.036***	0.029**	0.032***
	(0.56)	(0.57)	(0.54)	(0.50)	(0.013)	(0.012)	(0.012)	(0.012)
	239,464	239,464	239,464	239,464	239,464	239,464	239,464	239,464
Bandwidth 3, linear spline								
Subgroup qualifies	0.71	0.80	0.66	1.32*	0.038*	0.037*	0.035*	0.040**
	(0.85)	(0.87)	(0.84)	(0.79)	(0.020)	(0.020)	(0.019)	(0.019)
	154,444	154,444	154,444	154,444	154,444	154,444	154,444	154,444
Student demographics		✓	✓	✓		✓	✓	✓
School demographics			✓	✓			✓	✓
Year FE				✓				✓

Notes: Each result is from a separate regression with a linear or quadratic function of the count on each side of the threshold, allowing the slope to vary on either side. Regressions include various combintations year and grade fixed effects, gender, race, FRL status, special education status, and LEP status at the student level, as well as gender, FRL, special education, LEP, and student count at the school level, as indicated. All regressions include school-by-race fixed effects.

*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table 1.4: Specifications – Reading

	Passing Rate				Score			
Bandwidth 10, linear spline								
Subgroup qualifies	-0.009	-0.06	-0.10	-0.05	0.012	0.009	0.009	0.009
	(0.27)	(0.27)	(0.26)	(0.26)	(0.008)	(0.008)	(0.008)	(0.008)
	356,758	356,758	356,758	356,758	356,758	356,758	356,758	356,758
Bandwidth 10, quadratic spline								
Subgroup qualifies	0.23	0.50	0.37	0.49	0.009	0.019	0.016	0.018
	(0.44)	(0.43)	(0.43)	(0.42)	(0.014)	(0.013)	(0.013)	(0.012)
	356,758	356,758	356,758	356,758	356,758	356,758	356,758	356,758
Bandwidth 5, linear spline								
Subgroup qualifies	1.22***	1.44***	1.20***	1.34***	0.029***	0.036***	0.032***	0.034***
	(0.38)	(0.38)	(0.37)	(0.36)	(0.011)	(0.011)	(0.010)	(0.010)
	239,464	239,464	239,464	239,464	239,464	239,464	239,464	239,464
Bandwidth 3, linear spline								
Subgroup qualifies	1.83***	2.01***	1.97***	2.14***	0.036**	0.036**	0.037**	0.042***
	(0.59)	(0.59)	(0.58)	(0.56)	(0.017)	(0.016)	(0.016)	(0.016)
	154,444	154,444	154,444	154,444	154,444	154,444	154,444	154,444
Student demographics		✓	✓	✓		✓	✓	✓
School demographics			✓	✓			✓	✓
Year FE				✓				✓

Notes: Each result is from a separate regression with a linear or quadratic function of the count on each side of the threshold, allowing the slope to vary on either side. Regressions include various combintations year and grade fixed effects, gender, race, FRL status, special education status, and LEP status at the student level, as well as gender, FRL, special education, LEP, and student count at the school level, as indicated. All regressions include school-by-race fixed effects.

*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table 1.5: Results by Test Administration & Count Timing

	Math Pass Rate		Math Score		Reading Pass Rate		Reading Score	
	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring
Last Test								
Subgroup Qualifies	1.17**	0.17	0.032***	0.018	1.34***	0.06	0.034***	0.006
	(0.50)	(0.49)	(0.012)	(0.013)	(0.36)	(0.37)	(0.010)	(0.011)
	239,464	224,385	239,464	224,385	239,464	224,385	239,464	224,385
First Test								
Subgroup Qualifies	1.29**	0.13	0.031**	0.015	1.33***	0.03	0.030***	0.003
	(0.51)	(0.49)	(0.012)	(0.013)	(0.39)	(0.41)	(0.010)	(0.011)
	239,464	224,385	239,464	224,385	239,464	224,385	239,464	224,385
Student demographics	✓	✓	✓	✓	✓	✓	✓	✓
School demographics	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Each result is from a separate regression with a linear function of the count on each side of the threshold, allowing the slope to vary on either side. All regressions include year, grade, and school-by-race fixed effects, gender, race, FRL status, special education status, and LEP status at the student level, as well as gender, FRL, special education, LEP, and student count at the school level.
*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table 1.6: Results by Student Performance Prior Year

	Math				Reading			
	Pass	Pass	Score	Score	Pass	Pass	Score	Score
Subgroup Qualifies	0.64		0.025*		1.56***		0.039***	
	(0.57)		(0.013)		(0.42)		(0.011)	
Subgroup Qualifies * 1st Quartile		-0.26		0.016		1.82*		0.034
		(1.24)		(0.022)		(1.10)		(0.024)
Subgroup Qualifies * 2nd Quartile		0.106		0.007		1.56**		0.047***
		(0.86)		(0.022)		(0.67)		(0.016)
Subgroup Qualifies * 3rd Quartile		1.32***		0.005		1.18**		0.027*
		(0.57)		(0.015)		(0.47)		(0.014)
Subgroup Qualifies * 4th Quartile		1.09*		0.035**		1.51***		0.018
		(0.63)		(0.015)		(0.42)		(0.013)
Student demographics	✓	✓	✓	✓	✓	✓	✓	✓
School demographics	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
p-value for equality of coefficients		0.503		0.298		0.821		0.497
Mean	83.60	83.60	-0.039	-0.039	88.47	88.47	0.021	0.021
N	167,253	167,253	167,253	167,253	167,253	167,253	167,253	167,253
Clusters	4,795	4,795	4,795	4,795	4,795	4,795	4,795	4,795

Notes: Each column is from a separate regression with a linear function of the count on each side of the threshold, allowing the slope to vary on either side. All regressions include year, grade, and school-by-race fixed effects, gender, race, FRL status, special education status, and LEP status at the student level, as well as gender, FRL, special education, LEP, and student count at the school level. For columns with quartiles, each independent variable is interacted with each of the quartile dummies in one regression, allowing for the coefficients to vary by quartile.
*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table 1.7: Results by Race

	Math				Reading			
	Pass	Pass	Score	Score	Pass	Pass	Score	Score
Subgroup Qualifies	1.16**		0.033***		1.29***		0.035***	
	(0.50)		(0.012)		(0.36)		(0.010)	
Subgroup Qualifies * Black		2.00**		0.052***		1.73***		0.040**
		(0.86)		(0.018)		(0.62)		(0.016)
Subgroup Qualifies * Hispanic		0.92		0.006		1.09*		0.024
		(0.82)		(0.022)		(0.60)		(0.018)
Subgroup Qualifies * White		-0.44		0.026		0.78		0.038**
		(0.77)		(0.022)		(0.58)		(0.018)
Student demographics	✓	✓	✓	✓	✓	✓	✓	✓
School demographics	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
p-value for equality of coefficients		0.090		0.211		0.589		0.579
Mean	82.16	82.16	-0.084	-0.084	89.00	89.00	-0.020	-0.020
N	293,464	293,464	293,464	293,464	293,464	293,464	293,464	293,464
N Black	98,069	98,069	98,069	98,069	98,069	98,069	98,069	98,069
N Hispanic	84,467	84,467	84,467	84,467	84,467	84,467	84,467	84,467
N White	56,928	56,928	56,928	56,928	56,928	56,928	56,928	56,928
Clusters	4,821	4,821	4,821	4,821	4,821	4,821	4,821	4,821

Notes: Each column is from a separate regression with a linear function of the count on each side of the threshold, allowing the slope to vary on either side. All regressions include year, grade, and school-by-race fixed effects, gender, race, FRL status, special education status, and LEP status at the student level, as well as gender, FRL, special education, LEP, and student count at the school level. For columns with quartiles, each independent variable is interacted with each of the race dummies in one regression, allowing for the coefficients to vary by quartile.
*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table 1.8: Long-Run Results

	Panel A: Achievement Levels			
	Math		Reading	
	Pass	Score	Pass	Score
Current Year	2.14***	0.040**	1.21**	0.036**
	(0.70)	(0.018)	(0.50)	(0.015)
+1 Year	1.51**	0.038**	1.24**	0.019
	(0.65)	(0.018)	(0.55)	(0.016)
+2 Years	3.02***	0.040**	0.93*	0.036**
	(0.71)	(0.017)	(0.53)	(0.015
+3 Years	1.84**	0.030*	1.06*	0.021
	(0.76)	(0.016)	(0.61)	(0.015)
+4 Years	0.14	0.001	-0.51	0.014
	(0.81)	(0.016)	(0.59)	(0.014)
N	98,519	98,519	98,519	98,519
Clusters	3,263	3,263	3,263	3,263

	Panel B: Achievement Growth			
	Math		Reading	
	Pass	Score	Pass	Score
Current Year	0.89	0.036*	1.68**	0.038**
	(0.80)	(0.019)	(0.76)	(0.019)
+1 Year	-0.21	0.008	0.97	0.016
	(0.81)	(0.020)	(0.63)	(0.018)
+2 Years	3.04***	0.035*	0.18	0.006
	(0.97)	(0.018)	(0.72)	(0.018)
+3 Years	0.08	-0.010	0.96	0.012
	(0.96)	(0.017)	(0.77)	(0.016)
+4 Years	-0.49	-0.025	-0.56	0.009
	(0.94)	(0.017)	(0.63)	(0.015)
N	52,160	52,160	52,160	52,160
Clusters	2,845	2,845	2,845	2,845

Notes: Each result is from a separate regression with a linear function of the count on each side of the threshold, allowing the slope to vary on either side. All regressions include year, grade, and school-by-race fixed effects, gender, race, FRL status, special education status, and LEP status at the student level, as well as gender, FRL, special education, LEP, and student count at the school level. Each model in Panel B also includes the prior year's corresponding subject exam score. The sample in this table is restricted to students with a score in the year prior to the year of treatment and the four years following the treatment year.
*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table 1.9: Effect on Students Outside Qualifying Subgroup

	Math				Reading			
	Pass	Pass	Score	Score	Pass	Pass	Score	Score
Subgroup Qualifies	0.07		0.011		0.03		0.007	
	(0.40)		(0.009)		(0.24)		(0.006)	
Subgroup Qualifies * 1st Quartile		-0.80		0.010		-1.09		-0.014
		(1.01)		(0.017)		(0.084)		(0.017)
Subgroup Qualifies * 2nd Quartile		-0.65		-0.010		0.33		0.013
		(0.62)		(0.013)		(0.39)		(0.010)
Subgroup Qualifies * 3rd Quartile		0.10		0.009		0.40		0.010
		(0.43)		(0.010)		(0.29)		(0.008)
Subgroup Qualifies * 4th Quartile		0.76*		0.024**		0.22		0.015*
		(0.43)		(0.012)		(0.27)		(0.009)
Student demographics	✓	✓	✓	✓	✓	✓	✓	✓
School demographics	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
p-value for equality of coefficients		0.235		0.129		0.430		0.377
Mean	84.33	84.33	0.058	0.058	90.00	90.00	0.066	0.066
N	1,237,738	1,237,738	1,237,738	1,237,738	1,237,738	1,237,738	1,237,738	1,237,738

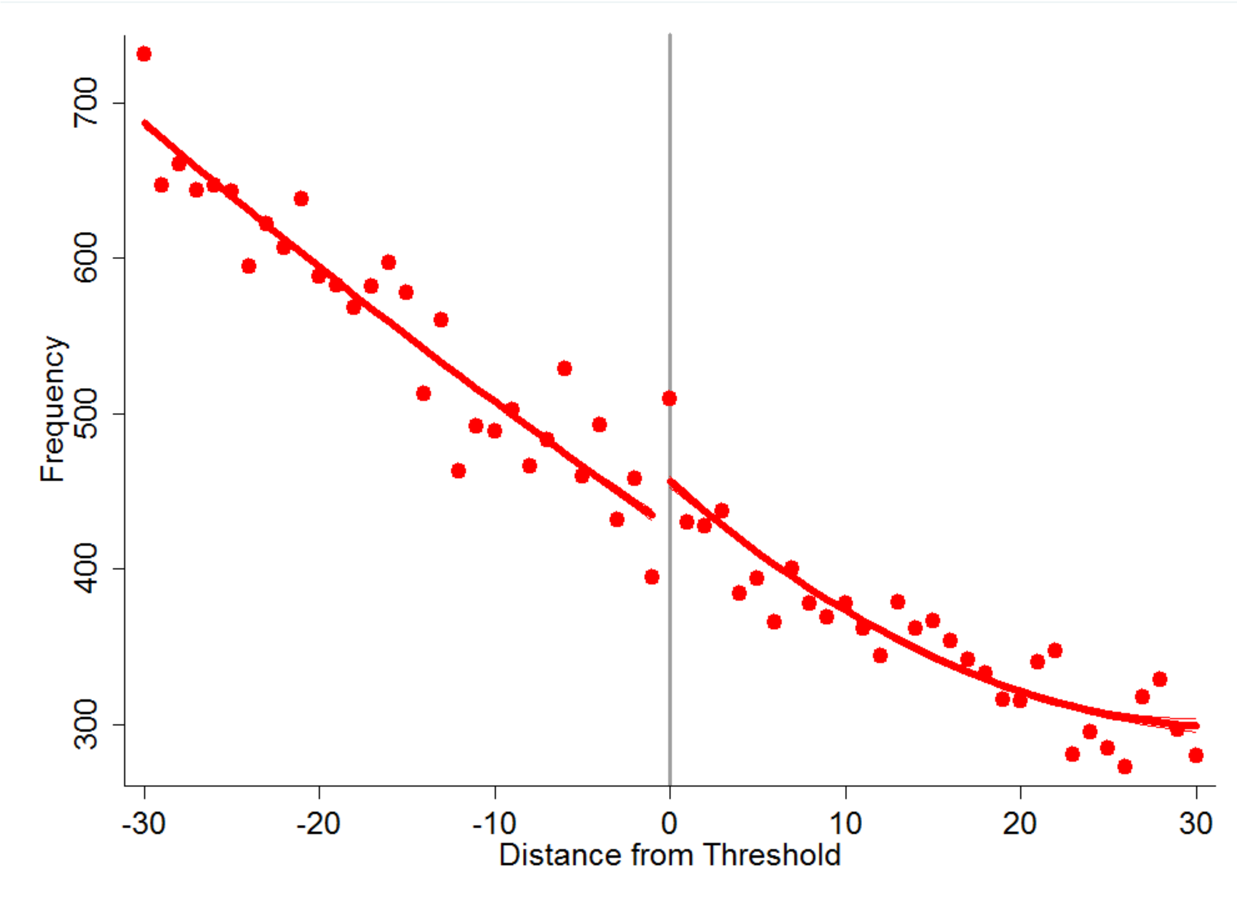
Notes: Each column is from a separate regression with a linear function of the count on each side of the threshold, allowing the slope to vary on either side. All regressions include year, grade, and school-by-race fixed effects, gender, race, FRL status, special education status, and LEP status at the student level, as well as gender, FRL, special education, LEP, and student count at the school level. For columns with quartiles, each independent variable is interacted with each of the quartile dummies in one regression, allowing for the coefficients to vary by quartile.
*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table 1.10: Spillover Effects on Non-AYP Subjects

	Science		Social Studies	
	Pass	Score	Pass	Score
Subgroup Qualifies	-0.18	-0.015	1.21	0.006
	(0.87)	(0.019)	(1.22)	(0.036)
Mean	75.13	-0.049	89.54	-0.033
N	67,134	67,134	22,126	22,126
Clusters	4,001	4,0001	998	998

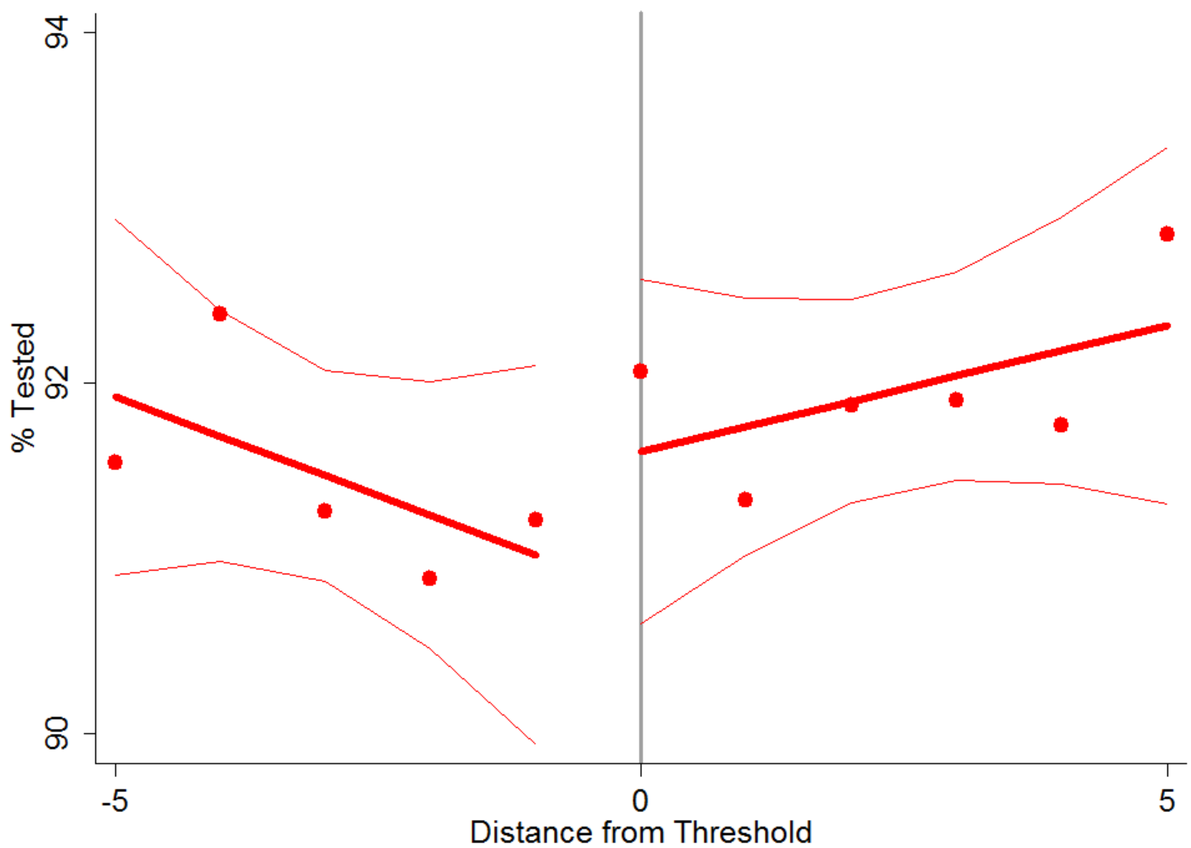
Notes: Each column is from a separate regression with a linear function of the count on each side of the threshold, allowing the slope to vary on either side. All regressions include year, grade, and school-by-race fixed effects, gender, race, FRL status, special education status, and LEP status at the student level, as well as gender, FRL, special education, LEP, and student count at the school level.
*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Figure 1.1: Check for Manipulation of Counts



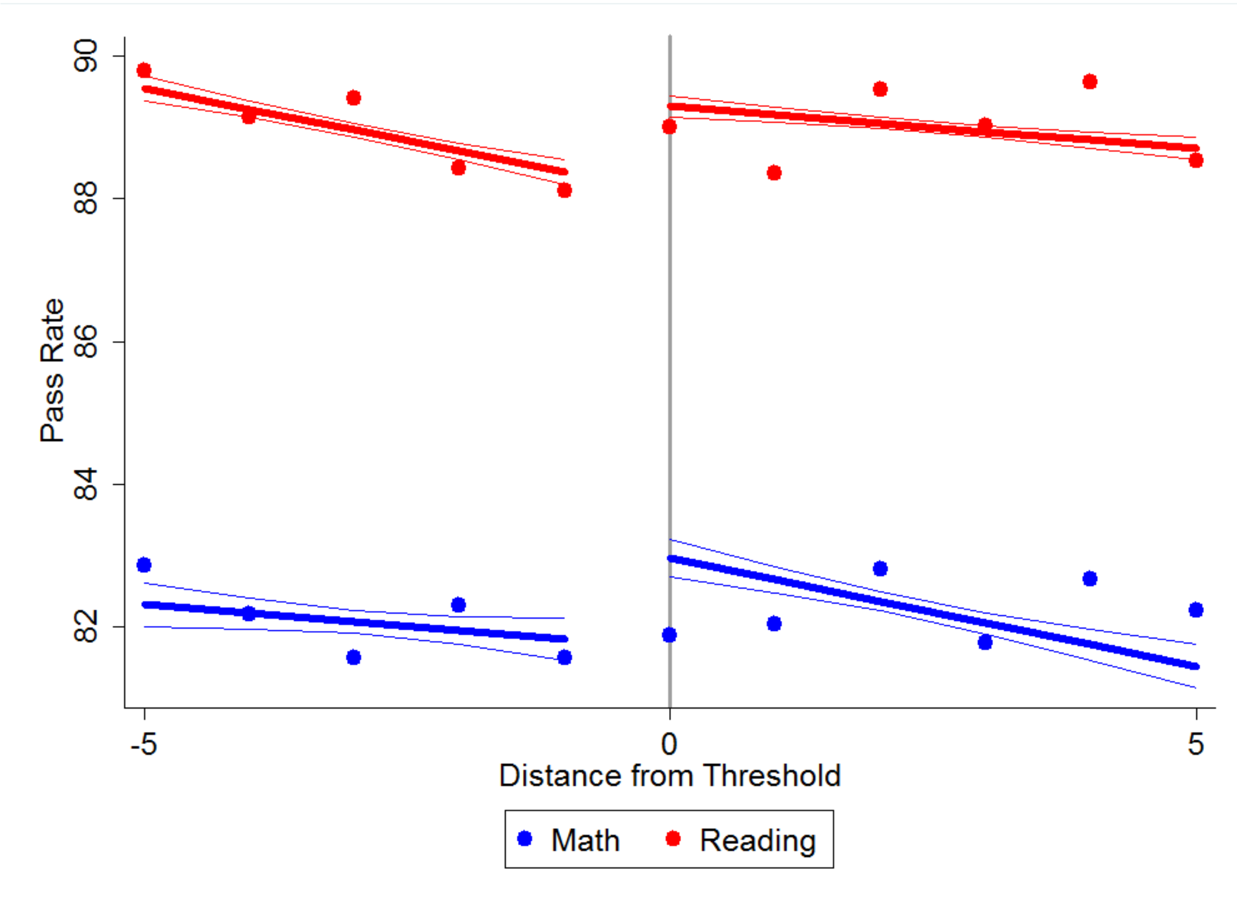
Notes: The figure shows the amount of race x school subgroups at each distance from the threshold, along with a fitted quadratic function on each side of the threshold.

Figure 1.2: Student Test Rates as a Function of Subgroup Count



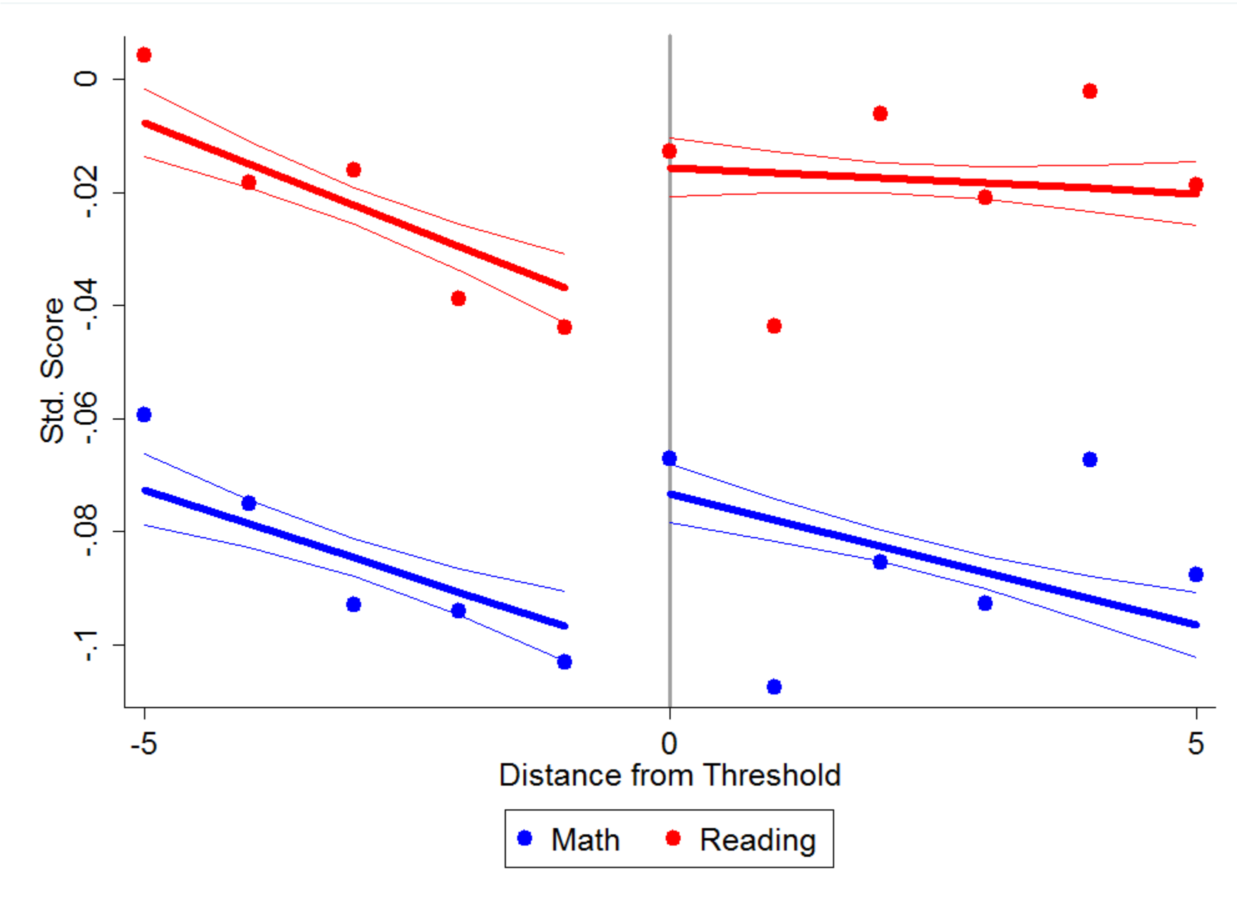
Notes: The figure shows the percentage of students tested at each distance from the threshold along with the predicted values generated by the default regression specification described in the text.

Figure 1.3: Passing Rates



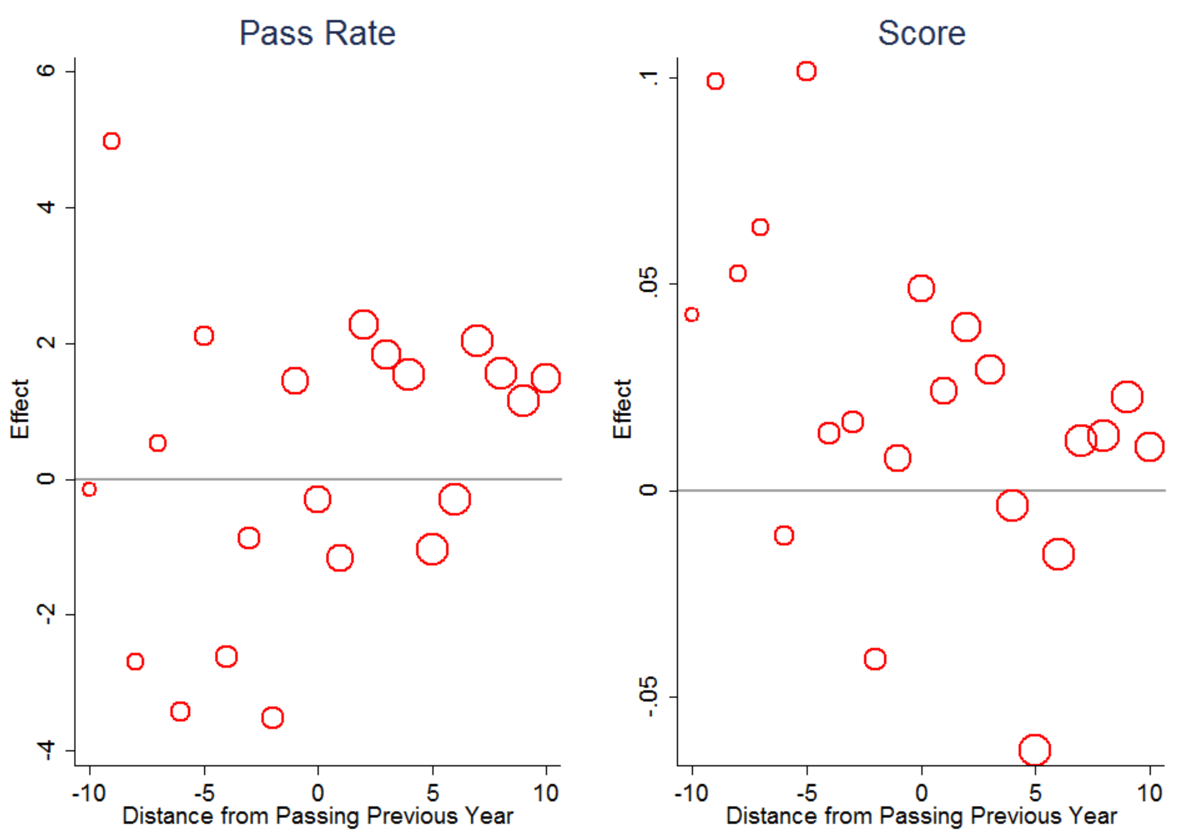
Notes: The figure shows the mean passing rate by the distance from the threshold, as well as the predicted values generated by the default regression specification described in the text.

Figure 1.4: Scores



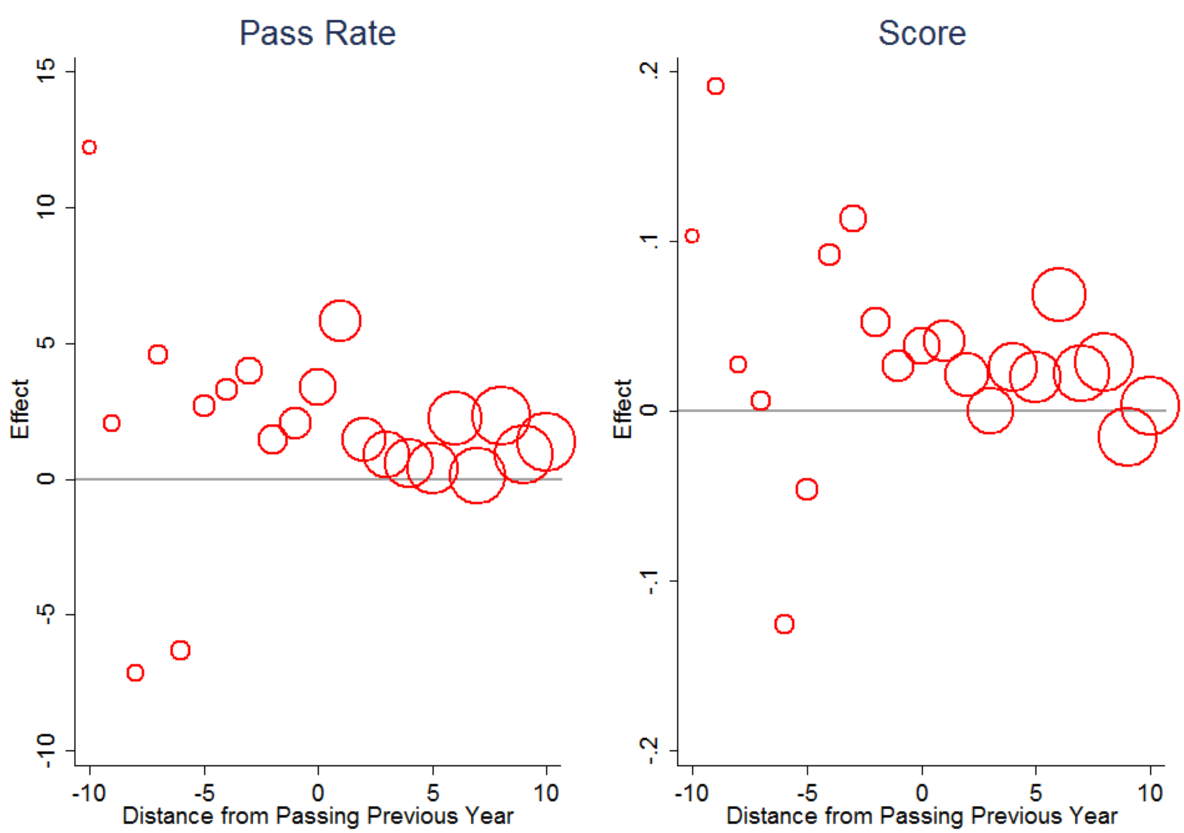
Notes: The figure shows the mean standardized score by the distance from the threshold, as well as the predicted values generated by the default regression specification described in the text.

Figure 1.5: Math Impacts by Distance From Passing Previous Year



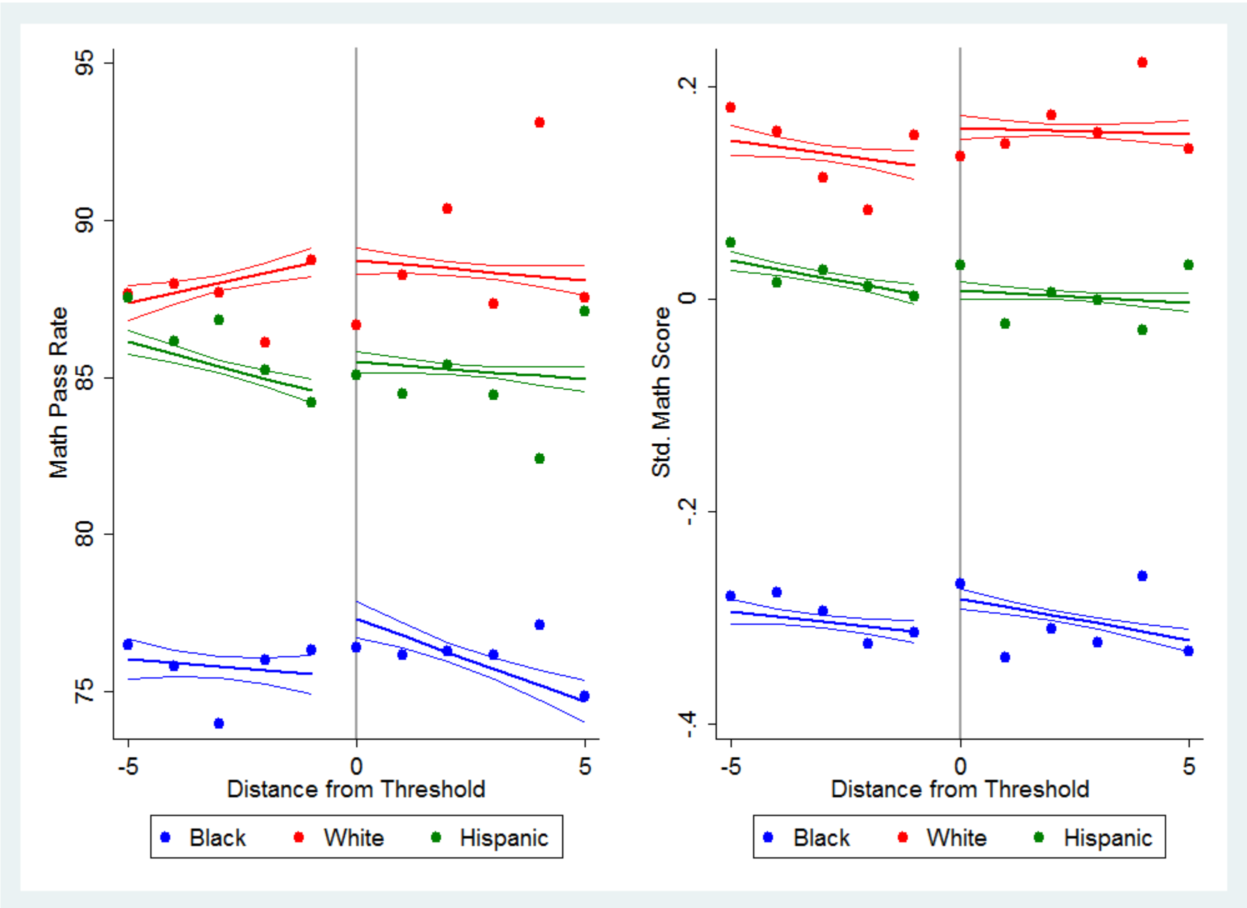
Notes: The figures above display the mean impacts when the sample is split by the distance each student was from passing the previous year. The data points are weighted by the number of students within each cell.

Figure 1.6: Reading Impacts by Distance From Passing Previous Year



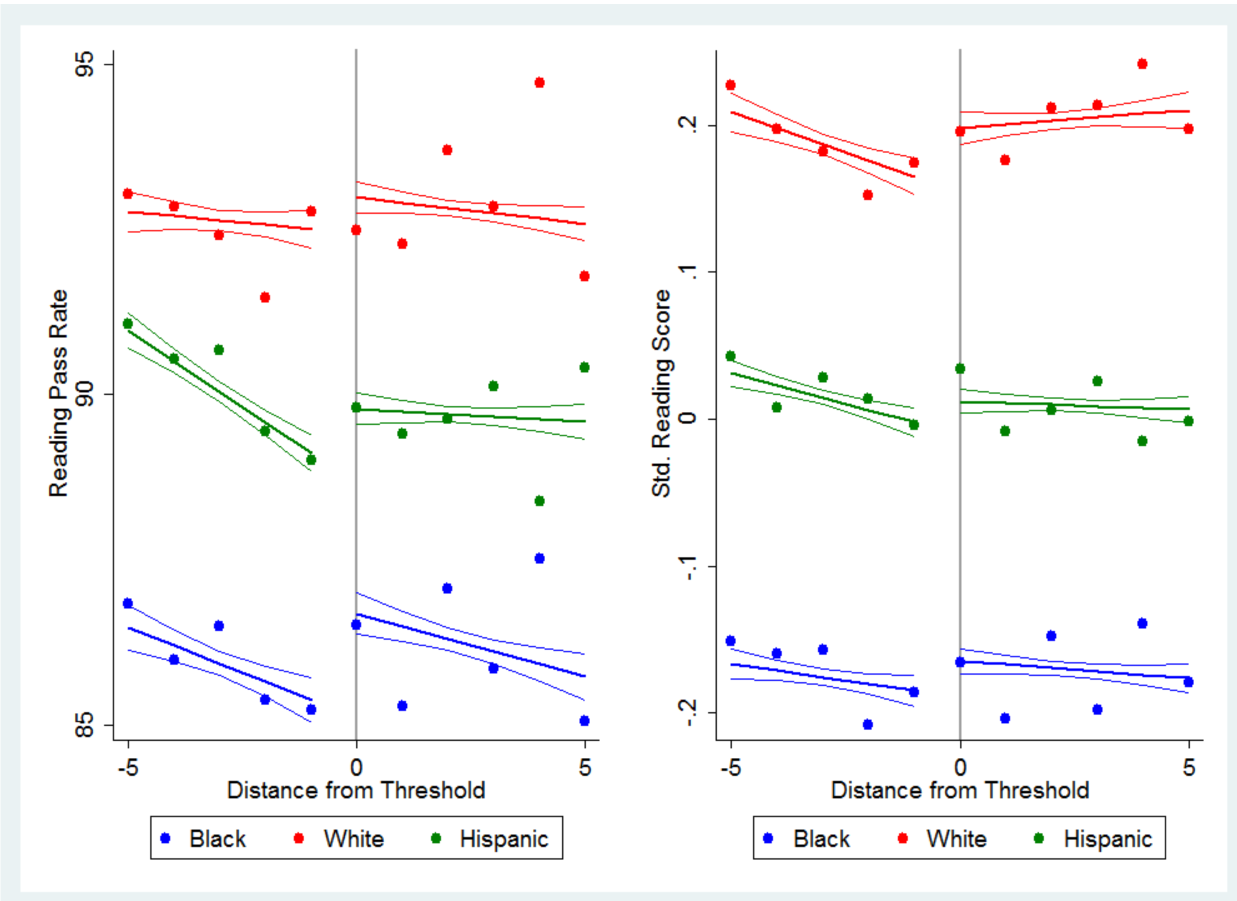
Notes: The figures above display the mean impacts when the sample is split by the distance each student was from passing the previous year. The data points are weighted by the number of students within each cell.

Figure 1.7: Math Impacts by Race



Notes: The figures above show results when running the main specification with each covariate interacted with race indicators. For formal results, see Table 1.7.

Figure 1.8: Reading Impacts by Race



Notes: The figures above show results when running the main specification with each covariate interacted with race indicators. For formal results, see Table 1.7.

Chapter 2

Teacher Knows Best? A Study on Interventions for At-Risk Students

2.1 Introduction

It has been established that low-income and minority students lag behind their better off peers in academic achievement using various metrics, such as test scores and completion rates at different levels. According to Reardon [2011], the test score gap is 30 to 40% wider than it was a mere 25 years ago. Bailey and Dynarski [2011] find that there are growing inequality gaps in college entry, persistence, and graduation. Even though the achievement gap has been studied in detail in the past years, the gap has not begun to close, per Hemphill and Vanneman [2011]. Given that the minority share of the population in the United States is growing, this issue figures to remain important for the foreseeable future.

The federal government gives school districts extra funding for low-income students under Title I, Part A in order to help school districts raise the achievement of these students. For schools with lower concentrations of low-income students, this money comes with strict guidelines for how it is to be used. However, in schools with higher concentrations of low-income students, this money can be treated as part of a school's general funds. While the money per student is similar, the treatment of these at-risk students varies between these two programs, called a Targeted Assistance Program (TAP) and a Schoolwide Program (SWP), respectively. In a TAP, at-risk students are targeted by specific interventions. In a SWP, at-risk students are considered as simply a part of the general population, and the school may or may not use interventions of its own design. These programs are described more in Section 2.1.

In this paper, I examine whether funding set aside for a specific purpose helps to achieve that purpose in schools. To do so, I identify the causal effect of schoolwide programs on the academic achievement of both at-risk and non at-risk students using quasi-experimental variation in the program provision generated by the Title I policy that dictates school program provision. Under the policy (most recently updated in No Child Left Behind, 2001), schools with a low-income population of less than 40% must use a Targeted Assistance Program, while those with concentrations at or above 40% may use a Schoolwide Program. This suggests a regression discontinuity design (RDD) in which I compare students in schools immediately above the threshold to students in schools immediately below the threshold in order to obtain the difference in outcomes between schools implementing Schoolwide Programs as opposed to those implementing Targeted Assistance Programs. I give more details on this strategy in Section 5.

This paper adds to the literature evaluating the effects school funding (described in Section 2.2) in three ways. First, it addresses how the money is used in schools, rather than simply extra money to assist with disadvantaged students. Even if papers studying the effect of additional money in schools show results, the mechanism for these results is not clear. In this paper, the money does not vary across treatments, but the interventions used for at-risk students do vary. Thus, I study slightly more of the actual practices within schools, rather than simply the broad financial inputs and academic outputs. Secondly, this paper evaluates whether money set aside for a specific purpose in schools achieves that purpose. Most school funding is general purpose, and often targeted funding is given for boutique interests. In this case, the targeted funding is aimed at a large and under-performing student population. Thirdly, the previous papers use a regression discontinuity design at a much lower threshold of low-income students. This paper uses a threshold much closer to the median concentration of low-income students across public schools in the United States, and thus is generalizable to a broader range of schools than the previous research. Earlier papers focusing on schools on the threshold for receiving Title I money are observing schools on one end of the poverty distribution. As the majority of schools have growing concentrations of low-income students, the median concentration is likely to grow, making this earlier research even less applicable.

I implement my empirical strategy using data from the Educational Resource Center (ERC) at the University of Texas at Austin. While this data is described in more detail in Section 4, I use reading and math test scores for students from third grade through eleventh grade between 2004 and 2011 as outcomes. I use the student-level demographic data to construct school concentrations of low-income students.

To preview the results in Section 5, I find an increase of about 17% in the likelihood of a school offering a schoolwide program at the threshold, but do not find many significant changes in school- or student-level covariates at the threshold. The distribution is smooth across the threshold as well. I use this increase in school provision of a SWP as an instrumental variable to identify the causal effect of SWP provision on student achievement. The use of an instrumental variable provides estimates of the local average treatment effect of these programs in schools where the provision of a SWP or TAP is completely governed by the policy.

The findings show no impact at all of the provision of a schoolwide program on student test scores. This holds for all students, as well as for low-income and non-low-income students separately. There is no pattern when splitting students by quartiles, either. The lack of significant impacts persists across multiple bandwidths and polynomial degrees, showing that this lack of impact is not due to a specification choice. This may be due to a smaller-than-expected discontinuity in the first stage, but more likely is indicative of the true impact of the program, given the small point estimates and standard errors.

2.2 Background and related literature

2.2.1 Legislative background

Title 1 was originally signed by Lyndon B. Johnson in 1965 as part of the Elementary and Secondary Education Act (ESEA) (Pub. L. 89-10, 79 Stat. 27, 20 U.S.C. ch. 70). Since 1965, it has been the largest source of federal government funding for K-12 education. In 2011, more than $14 billion of federal spending was allocated to Title I, and currently more than 50% of public schools receive public funding. Titled "Better Schooling for Educationally Deprived Children", its purpose was to give financial assistance to local educational agencies with high amounts of low-income children. The thought was that these students, being low-income themselves, who attended schools with other low-income students, had a much more difficult path to success than their educational peers.

There are many guidelines on how funding may be used at a particular school. For schools with high concentrations of low-income students, they may use the money for schoolwide programs (SWPs). The threshold for determining whether a high concentration existed was originally set in 1975 at 75%, meaning that schools with a student body composed of more than 75% low-income students could use the money as general funding. In 1994, Improving America's Schools Act (IASA) added standards to assess student progress and, more importantly for this paper, reduced the schoolwide program threshold to 50%. In 2001, No Child Left Behind (NCLB) lowered this threshold once again to 40%, along with other changes like mandating yearly standardized tests and adding in an Adequate Yearly Progress (AYP) indicator.

For a school to qualify for a schoolwide program, they must meet the threshold for eligibility and then apply. The application process takes a year, during which the school submits a comprehensive plan for the use of the funding to improve instruction at the school. This is important because it is actually the previous year's low-income student count that indicates eligibility. The funds can only be used to supplement funds available from non-federal sources, and cannot be used to replace other funding. Thus, results are not due to any crowding-out effect. The program must promote schoolwide reform and upgrade the entire educational operation of the campus. Lastly, once a school becomes eligible, the school keeps its eligibility, as long as it is eligible for Title I, Part A funds, which is a lower threshold. Given that most schools have seen a rise in poverty, this is not often an issue. It is also important to note that the threshold for schoolwide program eligibility does not have anything to do with the amount of funding for a school, only how that funding may be used.

Schools below the threshold for schoolwide program eligibility must use the funding for Targeted Assistance Programs (TAPs). These programs meet the intent and purpose of Title I if they serve only students failing or most at risk of failing and provide supplementary services designed to meet the needs of the students in the program. Schools must maintain records that document that the Title I, Part A funds are spent on activities and services for only Title I, Part A children (those failing or most at-risk of failing). Targeted Assistance Programs include many items, such as extended learning to allow at-risk students to catch up with grade level peers. These include an extended school year, before- and after-school programs, or summer program opportunities. They may include small group pull-outs or tutoring before, during, or after school. The funding may also be used to pay for paraprofessionals in classrooms containing mostly at-risk students, such as remedial classes, or to pay for extra instruction taught by a designated Title I teacher. The important consideration is that the funds are targeted toward at-risk students and not the general population.

2.2.2 Related Literature

There is a substantial body of literature on the relationship between funding and student performance, and many of these papers are on the use of Title I funding to supplement funding for schools with high concentrations of low-income students. However, the majority of this research is on the effect of receiving the extra money and not how it is used. Both Card and Payne [2002] and Gordon [2004] examine the effects of extra funding from the state and the federal government, respectively. Gordon finds that federal funding is completely crowded out within three years, but Card and Payne find that state funding actually increases district spending. More importantly, they show evidence that equalization of spending across districts narrows the gap in test score outcomes between income groups. This paper does not explicitly address Title I Part A funding, but more recent research does.

Van der Klaauw [2008] uses data from New York City public schools from 1993, 1997, and 2001 to look at schools directly above and below the threshold for Title I funding. Because he has data from a specific school district with explicit eligibility rules, he can exploit the discontinuity at the eligibility threshold, which in this case was the average poverty level across the district. Using an RD design, he finds that Title I funding was ineffective at raising student performance, and actually finds negative estimates in the first two samples. He considers multiple outcomes, including grade retention, suspension, school attendance, student mobility rates, and a set of reading and math test scores. Matsudaira et al. [2012] perform a very similar analysis to Van der Klaauw in a large urban school district. The authors also use an RD design on the within-district threshold for eligibility for Title I funds. However, they first look at how much that eligibility impacts funding for schools, finding that it raises federal revenue by about $460 per student, but only raises education spending by $360 per student. They find no impact on overall school-level test scores as well as no impact on those subgroups most targeted by Title I funding. They also find that Title I eligibility results in no significant increase in total direct expenditures because such small amounts of money are involved compared to the state and local funding. While more detailed than Van der Klaauw's analysis, the methodology and results are exactly the same. The authors in these papers use a regression discontinuity approach around the threshold for Title I funding, examining the effect of giving more money to schools on academic outcomes.

Outside of the United States, Leuven et al. [2007] look at the effects of two different subsidies for primary schools with sharp cutoffs at a threshold of 70% disadvantaged students. One subsidy is aimed at extra funding for personnel, while the other is used for technology. They use a modified regression discontinuity (RD) design [Hahn et al., 2008], which in this case is a difference-in-difference with a narrow bandwidth around the threshold of 70%. They actually find negative point estimates, and that the extra technology funding is especially detrimental to girls' achievement. The personnel estimates are explained by suggesting that the schools above the threshold already had sufficient personnel, and that more educators had a very low marginal value. Ooghe [2011] examines a program similar to Title I in Belgium using both regression discontinuity and difference-in-difference, finding positive effects for mathematics, reading, and spelling, but only significant effects for spelling. He finds somewhat larger effects for disadvantaged students and smaller effects for lower initial performers.

Thus far, the literature has been focused on finding a relationship between extra funding for disadvantaged students and their academic outcomes. However, in this paper I use strings attached to Title I funding to examine how school choice in the use of supplementary federal funds affects both disadvantaged and non-disadvantaged students. There is a small, non-causal body of research on this topic (Wong and Meyer [1998], Sunderman [2001], Boland et al. are some examples), but these are almost entirely qualitative in nature and do not attempt to quantify causal effects. In this paper, I use a threshold of 40% low-income students, while previous research is based around a much lower threshold. This paper will be about schools much closer to the typical public school in the US, which has a student body with about 50-55% low-income students. Furthermore, earlier papers observe the school as a ”black box”, only observing what goes in and what comes out. In this paper, I attempt to at least subdivide the larger black box into two smaller black boxes inside the school and observe how a specific mechanism impacts students differentially.

2.3 Empirical strategy

2.3.1 Conceptual framework

The direction of the impact of targeted assistance programs (as compared to schoolwide programs) on student achievement is theoretically ambiguous for at-risk students, and positive for non at-risk students. The ambiguity for at-risk students is due to the differences between the three channels through which targeted assistance programs differ from schoolwide programs: dedicated staff increases, student pull-outs during the school day, and student pull-outs outside the regular school day or year. For one, the quality of the new staff dedicated to the at-risk students may suffer. Secondly, and perhaps more importantly, there is an increased probability of tracking due to the targeted interventions.

I switch terminology often between at-risk and free-or-reduced lunch (FRL) status because schools decide how to divide students into these two groups based on academic measures. While this may not be exactly aligned with income, this data is not available, and thus I am using FRL status as a highly correlated indicator.

Dedicated staff increases and mid-day student pull-outs work similarly. Dedicated staff increases mean that there are classes full of at-risk students taught by Title I teachers, while mid-day student pull-outs are usually groups of at-risk students being pulled out of class to be taught by a Title I paraprofessional or teacher. These dedicated staff members are brought on for a full year at a time, just like any other staff member. However, even for the Title I teachers, one might posit that they are of lower quality than already established classroom teachers, for one of two reasons: They would already be hired if they were of higher quality, and most often the least experienced, newest teachers are given these classrooms full of at-risk students. The paraprofessionals are not teachers, though for pull-outs they will effectively become so. In either case, the at-risk students are learning from staff members of lower quality than they would under a schoolwide program. This would cause their test scores to suffer, but they are also learning in smaller groups, which might mitigate this effect somewhat.

In either of these situations, there will be increased tracking of students, which might outweigh effects by any other channel. By forcing schools to dedicate staff solely to at-risk students, the guidelines mean that there will be classrooms or pull-out groups full of at-risk students who are achieving poorly academically. When this happens, the non at-risk students will also be grouped. The TAP effectively mandates that schools track students academically. For the non at-risk students, it seems logical that this can only improve their scores. For the at-risk students, this may help them academically due to teaching more focused on their needs, but may hurt them moreso due to behavioral concerns and peer effects. They also might not get any more of the teacher's time individually, due to there being more students needy of that time.

The one intervention that should have clear, unambiguous effects is student pull-outs outside of the normal school day or school year. At-risk students are given tutoring or summer school on top of their normal classes. This should have only positive impacts on academic outcomes, as it supplements rather than supplants normal classroom instruction.

The percentage of low-income students in a school most often increases over time, so it is more useful to think of the outcomes in this fashion. As schools move from targeted assistance programs to schoolwide programs, the effects are most likely negative for the non at-risk students, who benefit from the tracking inherent in the targeted assistance programs. However, for at-risk students, the effects are much less clear. They benefit from supplemental instruction outside of the normal school day or year, but the effect of dedicated staff could go either way. They might benefit from smaller, more focused instruction, but the composition of these classes or small groups might hurt them through peer effects.

2.3.2 Identification strategy

Due to the difficulty in predicting how moving from a targeted assistance program to a schoolwide program will impact students academically, an empirical framework is necessary. I want to estimate the impact of schoolwide programs on achievement (among Title I schools), for which the following equation would be used:

$$y_{ist} = \alpha + \omega Y_{is_{t-1}} + \beta SWP_{st} + \delta FRL_{ist} + \pi X_{ist} + \gamma Z_{st} + \epsilon_{ist} \tag{2.1}$$

for student i in school s in year t. y is an academic measure, SWP is an indicator for the school having a schoolwide program, X is a set of student characteristics including FRL status, Z is a set of school characteristics, and ϵ is the error term. The parameter of interest is β, which should measure how income students are impacted by schoolwide programs. However, schools using schoolwide programs have a higher percentage of low-income students, who score lower academically, on average. Thus, one cannot interpret the ordinary least squares (OLS) estimate of β as a causal effect of schoolwide programs. Due to the compositional differences, these schools would most likely have lower test scores regardless of the presence of a schoolwide program. To address this issue, I use a regression discontinuity approach based on the policy rule determining which schools are eligible to run schoolwide programs. Under this provision of NCLB, schools are allowed to apply for schoolwide program use when their student body is greater than or equal to 40% low-income. Once they have a schoolwide program in place, they may keep even if they fall below the threshold, but this rarely happens in practice.

I will essentially be comparing student outcomes for both groups in schools immediately above the 40% threshold to those immediately below it. It is unlikely that these schools differ much, but due to the policy rule, only those above the 40% threshold are allowed to offer the schoolwide program, while those below it are mandated to use a targeted assistance program. While I would like to examine only those schools in a very narrow bandwidth, for precision I expand the bandwidth around the threshold. As stated before, there is a relationship between the percentage of FRL students and academic outcomes, so I must control for this at the school level. Therefore, I estimate the following first-stage equation to study the impact of the policy on the decision to offer a schoolwide program at the school level:

$$SWP_{st} = \alpha^{FS} + f(FRL\%_{st}) + \delta^{FS} Above40_{st} + f(FRL\%_{st}) * Above40 + \pi^{FS} X^{FS}_{ist} + \gamma^{FS} Z^{FS}_{st} + \epsilon^{FS}_{ist} \tag{2.2}$$

for school s in year t. $FRL\%$ is the school percentage of FRL students and $Above40$ is an indicator for the percentage being equal to or greater than 40. Because it takes a year to apply for a schoolwide program and schools cannot lose the program once they have it, I actually use the following for $FRL\% : FRL\%_{st} = max\{FRL\%_{s2001,} FRL\%_{s2002}, FRL\%_{s(t-2)}, FRL\%_{s(t-1)}\}$. I find that $\delta^{FS}$is positive and statistically significant, showing that some schools that otherwise would have had targeted assistance programs were induced to move to schoolwide programs. I then instrument the endogenous regressor in equation (1), SWP, with $Above40$. Xin equation (1) contains, X^{FS}, $f(FRL\%)$,and $f(FRL\%) * Above40$. This is following a standard fuzzy regression discontinuity framework Imbens and Lemieux [2008], and yields a causal estimate of β. Equation (2) is the first-stage question associated with the 2SLS estimation of equation (1).

Below I also report the results of estimating the reduced-form equation,

$$y_{ist} = \alpha^{RF} + \omega Y_{is_{t-1}} + f(FRL\%_{st}) + \delta^{RF} Above40_{st} + f(FRL\%_{st}) * Above40 + \pi^{RF} X^{RF}_{ist} + \gamma^{RF} Z^{RF}_{st} + \epsilon^{RF}_{ist}. \tag{2.3}$$

This reduced-form equation estimates the effect of being just above the 40% cutoff, δ^{RF}, on achievement. However, to obtain the effect of schoolwide programs on achievement, I rescale this reduced-form effect, equivalent to the 2SLS estimation of β. In yet a further specification to find differential impacts of the policy on FRL students and non-FRL students, I allow both the discontinuity the functions on either side of the threshold to vary by FRL status, allowing for different coefficients for the two groups of students.

The fuzzy RD strategy identifies the local average treatment effect (LATE) for schools close to the 40% cutoff. These schools are not far off the mean for % FRL students by school in the United States, at around 50% (in 2010). Thus, the schools used for identification in this paper can be thought of as characteristic of generic public schools. However, due to the choice inherent in the policy, identification of compliers is more complicated. Schools do not have to apply once they cross the threshold; they may stay with their targeted assistance program if administrators choose to do so. The results identified in this paper are based on schools that are above the threshold and have chosen to apply. There are two possibilities for which schools these are: 1) schools in which the administrators believe that they can provide better interventions for at-risk students than those prescribed by the targeted assistance program, or 2) schools in which the administrators believe a schoolwide program is simply easier to administer than the targeted assistance program. However, due to the length and difficulty of the application process, there is a deterrent to those schools trying to switch over simply for ease of administration. Therefore, I believe that the effects identified are from those schools in which administrators believe their own interventions are better than those mandated by a TAP. However, there is no data on which interventions are used in the schools that have a SWP, so the reader should think of the identified effects as the effect of allowing schools choice in how they help their at-risk students in place of specific programs within the targeted assistance programs. This paper speaks, in that respect, to the issue of local versus federal control of school practices. In a SWP, schools have control over how to use the money, and in a TAP, the government mandates specific interventions to be used with the money given.

2.4 Data

The data used for implementation come from the Texas Education Resource Center (ERC) at the University of Texas at Austin. At the student level, the data contains demographics, school information, and reading and math test scores from the end-of-year exams called the Texas Assessment of Knowledge and Skills (TAKS). Using the student information, I construct a ratio for low-income students for each year. Because the application process takes ones year, I lag this ratio. The minimum threshold for a schoolwide program changed with the passage of NCLB in 2001, so I use data from 2001 forward for each school. Due to the policy that schools cannot "lose" a schoolwide program once they have one, I then take the max of these low-income measures for each school between 2001 and the lagged year in the dataset.

I merge this data with the publicly available Common Core of Data from the National Education Center for Education Statistics (NCES)to get more information on each school. Namely, I use the indicators in this dataset for both Title I status and whether the school runs a Targeted Assistance Program or a Schoolwide Program. I drop any schools without Title I status, but as these generally have very low FRL ratios, this does not materially affect my results either way.

For academic outcomes, I standardize the TAKS scores in math and reading by year and grade. I also use the passing indicator in the data, though the threshold for passing varies by subject, grade, and year. The test is given in grades 3-11 from 2003 through 2011. I control for baseline scores throughout the paper, so this eliminates third graders and any exams from 2003.

Once the datasets are merged and only the students in schools within the 10% bandwidth kept, I have a total sample of 2,138,719 students across 5,745 schools, about 370 students per school sampled. As one can see in Table 2.1, there are more non-FRL students than FRL students due to the 40% threshold by school. The FRL ratio is certainly not equal across the threshold, but this also follows directly from the research design. About two-thirds of the students in the sample attend schools above the threshold, due to the distribution of FRL percentage by schools in the sample. The students in schools above the threshold appear less likely to be white, and more likely to be black or hispanic. Students in schools above the threshold score about .1 standard deviation lower on both math and reading. This could be due to schoolwide program status or due to the demographic composition.

2.5 Results

2.5.1 Discontinuity / First Stage

Per the guidelines pertaining to schoolwide program provision in Title I schools, one should see a discontinuity in the ratio of schools using a schoolwide program at the 40% low-income threshold. I provide visual evidence for this in Figure 2.1, plotting the percentage of schools with schoolwide programs by the percentage of low-income students in each school as well as a linear approximation. One might expect a jump from zero to 100% at the cutoff, but for several reasons this is not the case.

The principal reason for the smaller jump present in the figure is that schools are allowed to use a measure other than free-and-reduced lunch count for their low-income count. They may use Medicaid numbers, Supplemental Nutrition Assistance Program (SNAP) numbers, or Census data. They may also use the low-income percentage from a feeder school. For example, a middle school with 34% low-income students that receives students from an elementary school with 43% low-income students may implement a schoolwide program. Lastly, schools above the 40% threshold do not have to implement a schoolwide program; they may choose not to apply for one, preferring the targeted assistance program for one reason or another. As stated earlier, this is why the results must be interpreted as the effect of the freedom of the school to choose different interventions than those prescribed under Targeted Assistance Programs. For these reasons, the jump is not as sharp as one might hope, but there is still a significant discontinuity present.

Table 2.5 provides the first-stage results corresponding to Figure 2.1 as part of the general results. The figure includes both a linear spline specification and a quadratic spline specification. For the purposes of the paper, I base my results off the linear specification, but results from various bandwidths and polynomials are included in Tables 2.8 and 2.9. The coefficient for the linear discontinuity is positive and significant in all specifications. Other specifications were done, with similar results – only the preferred specifications are included. Schools with more than 40% FRL students are about 17 percentage points more likely to have a schoolwide program than districts below the threshold.

Though a larger discontinuity would allow us to see impacts more readily, it is just not feasible in this case. There are several different methods of measurement districts are allowed to use, as well as several exceptions. I believe those account for the small discontinuity. I did the first stage analysis for several subsamples of schools to determine if there were one type of school that complies more readily with the policy than others, and the 17 percentage point increase is pretty constant across the subsamples. However, even with a smaller discontinuity than expected, because of the size of the data set, the F-test value for the first stage is quite large, at about 39, well above the generally accepted minimum value of 10 for a proper IV analysis.

2.5.2 Tests of the validity of the RD design

Schools receive a constant marginal amount of federal Title I money per low-income student around the threshold in question. Thus, they have an incentive to enroll as many students as possible in the FRL program. This would not affect the distribution around the 40% threshold. However, if school administrators believe a schoolwide program is beneficial, there might be a second order incentive to push their school count above the cut-off. To assess this, I plot the distribution of schools around the 40% cut-off in Figure 2.2. A discontinuity in the distribution would mean that schools are manipulating the running variable, thus invalidating the research design [**?**]. In the figure, one can see a possible negative discontinuity at the threshold, which in this case is not concerning. There is no reason for schools to want to drop below the 40% threshold, as there is no discernible benefit from doing so. Being above the threshold allows for more freedom, and no other policies are in place around the 40% number. I also formally complete a Mccrary test, finding no significant discontinuity in the distribution around the threshold. For this test, I follow the general Mccrary procedure, binning the counts by FRL ratio, constructing a fourth-order polynomial on either side, and checking for a discontinuity at the 40% threshold. In this case, the discontinuity is extremely small and insignificant. Results for this test are found in Table 2.2. Thus, I do not believe there is precise control of the running variable.

Next, I check for differences in observable characteristics across the threshold, both at the school level and at the student level, using the preferred specification with each covariate as the outcome in a separate regression. This is found in Table 3.5. At the student level, I examine the propensity to be white, black, hispanic, female, FRL status, English Language Learner status, and special education status. At the school level, I examine the number of students, percentage breakdowns for white, black, and hispanic students, and the student/teacher ratio. As previously mentioned, students above the threshold are more likely to be White, and less likely to be Hispanic. Schools above the threshold show a similar pattern in their student bodies. Also, schools above the threshold are about 50 students smaller than those below the threshold. The fact that several of the coefficients in the table are significant is not ideal, and it is necessary to think through how this might bias results. Based on correlations between race, student count, and test scores, it appears that these would tend to bias the results upward.

I also check for differences in individual baselines student test scores for both math and reading across the threshold, shown in Table 2.4. Baselines scores in this case are comprised of student test scores from the prior year TAKS exam. Here, the numbers are not what one would expect based solely on the demographics in the previous paragraphs. Students above the threshold score 1% of a standard deviation lower in reading, and pass the math test at a 0.6% lower rate. This differences are quite small, and most likely won't impact the results much, but I do include baseline math and reading scores in the main specifications for the analysis. Due to this, the coefficients in question should be thought of as the effects on student growth, not student achievement levels.

2.5.3 Effects of Schoolwide Programs on Student Achievement

In the principal specification, I examine the impact on all students on both math and reading exams of a schoolwide program. Figures 2.3 and 2.4 show the reduced-form relationship between low-income student percentage and achievement on both the math and reading tests. In math, there appears to be a positive, though very small, discontinuity at the threshold. In reading, there is an extremely small discontinuity, one that appears to almost not be present at all. Achievement decreases as the percentage of low-income students within a school increases throughout the bandwidth. This relationship does not appear to be of a higher order, leading me to control for a linear function with differing slopes on each side of the threshold in my preferred specification.

The corresponding estimates are found in Table 2.5, and show exactly how small the discontinuities in the figures are. The first column on each side of the table shows results for the standardized score, while the second shows results for the passing rate. The first row shows the IV estimate under all three specifications, while the second row shows the reduced form estimate and the third the first stage estimate. These results come from my preferred specification, with a linear spline on low-income ratio, and control for student demographics, school demographics, and baseline test scores.

The reduced-form estimates are bounded around zero, with a minimum of -0.004 for the math score, and a maximum of 0.001 for the reading scores. As one might expect, this leads to IV estimates very close to zero as well. None of the four IV estimates are significant at the 10% level, and the t-stat with the largest magnitude of the four is about -0.6. This is not due to large standard errors but rather to extremely small coefficients. In all cases, the first stage is constant, and shows a jump of 17.1 pp at the threshold in the probability of implementing a schoolwide program.

It is apparent from these results that, on average, using a schoolwide program over a targeted assistance program has no impact whatsoever on test scores for either math or reading. However, due to the nature of the two programs, this could be possible due to varying effects across groups canceling each other out, as the targeted assistance program targets specific students, while the schoolwide program does not.

2.5.3.1 Results by FRL Status

While general test results are an important outcome for these programs, the nature of schoolwide programs call for an investigation into how the the students targeted by the policy are impacted differentially. In the control state (a TAP), schools must use Title 1 money for at-risk students. One approximation for this in the data is FRL status. In a schoolwide program, the money can be used across the schools as general funds. Thus, I examine the impact of schoolwide program on both FRL and non-FRL students for both tests using the specification described earlier where I interact all the independent variables with FRL status, as well as include a variable for this status. In this model, I allow both the discontinuity and the slopes to vary by FRL status. Non-FRL students score much higher, about .4 standard deviations on each test, on average. This is true for both tests. Results for this are found in Table 2.6.

Across the four outcomes and two groups, not one coefficient is significant. Moreover, there is no discernible pattern where any type of conclusion might be drawn. Similarly to the previous table, all the results are bunched quite tightly around zero. Not only does it appear that there is not any type of impact on either group individually, but the results do not vary across groups, as the policy might suggest. The bottom row shows a p-value from a test of equality of the two coefficients displayed above. A low p-value would indicate that the two coefficients can be rejected as equal, while a higher p-value would indicate that the two cannot be rejected as equal. The p-value for these tests are given in the bottom row. The coefficients cannot be rejected as equal for any of the four outcomes.

2.5.3.2 Results by Quartile

While the definition of targeted students is somewhat nebulous and is largely left up to the schools, I also investigate potential impacts across groups by splitting the students into quartiles based on baseline test scores. If schools are targeting the lowest-performing students, then one would expect these students to do worse in a schoolwide program than in a targeted assistance program, while the higher-performing students would score higher. In Table 2.7, I interact each of the independent variables with an indicator for each of the four quartiles, allowing for differential analysis.

In math, there is no significant impact on the lower-performing students, but the highest quartile actually performs worse in a schoolwide program than in a targeted assistance program. Students in the fourth quartile show an impact of -0.086 standard deviations in test score and -0.033 percentage points lower in passing rate in a schoolwide program. This is the opposite of what one might expect, and is difficult to explain.

In reading, the pattern is less clear. The fourth quartile students score 0.048 standard deviations higher in a schoolwide program, but do not pass at a higher rate. There are two significant coefficients among the reading estimates, and both only at the 10% level. Conclusions are tough to come. Once again, in the bottom row of the table, I test for whether the estimates can be rejected as equal. In this case, the impacts on reading score can be rejected as equal at the 10% level, but the estimates for the other outcomes cannot.

2.5.3.3 Compliers

As in any instrumental variables model, it is important to think about which subjects fall into the complier category, as opposed to the never-takers and the always-takers. In this case, the never-takers are the schools who keep a targeted assistance program even though they are above the 40% threshold. There are several possible reasons for this. They might believe that the interventions required within a TAP are more beneficial than the alternative. They might not want to go through the long application process for a schoolwide program. Always-takers are much harder to figure out, though according to Figure (2.1) there are some schools well below the threshold that do indeed have schoolwide programs. They may be using a feeder school's low-income ratio or a different measure than FRL.

Compliers, then, are those schools who are using their own FRL percentage as their measure and who apply for a schoolwide program immediately at or above the 40% threshold. These schools most likely have administrators who believe that they can do better by their students under their own control of the Title I funds than they can under the required TAP interventions. They may use the same interventions, but use the funds differently. They may have completely different interventions than those required in TAPs, or they may integrate the funds into the general school finance structure and not use any interventions at all. The important difference between schools with TAPs and schools with SWPs among the complier category is the lifting of the required interventions. Thus, the results may be interpreted as more about school choice than about the interventions themselves. Are students better off in schools where some of their educational structure is dictated by Title I guidelines or in schools where their individual schools or districts are free to structure the education as they see fit?

2.5.4 Sensitivity Analysis

In Tables 2.8 and 2.9, I examine the robustness of my results to several concerns regarding the choice of specification for the regression discontinuity design on math and reading outcomes, respectively. Each cell shows the 2SLS RD estimate of the effect of a schoolwide program on achievement levels from a separate regression. The first column in each table shows the estimates with no controls, while the second includes student-level covariates. In the third, I add school-level covariates, and in the fourth, I add baseline test scores. The first row displays baseline 2SLS estimates in which I use a bandwidth of 0.1 and a linear spline specification. The estimates from the preferred specification used throughout the paper are found in the first row and fourth column for each outcome.

In rows 2 and 3, I check the sensitivity of my estimates to the chosen bandwidth of 0.1. While the main analysis includes students in schools with a student body of between 30% and 50% FRL members, in row 2 I reduce the bandwidth to 0.05 and in row 3 I increase the bandwidth to 0.2. The smaller bandwidth leads to larger estimates with even larger standard errors, using a sample size of about half of the original. The larger bandwidth leads to more precise and smaller estimates for math, while the estimates for reading are larger. The precision in this specification is helpful, but it is hard to believe that schools at 20% FRL membership and those at 60% FRL membership are similar enough for a true RD comparison. For math, none of the estimates using these bandwidths are significant for any of the covariate groups included. For reading, using a bandwidth of 0.2 gives significant negative estimates for a couple of the outcomes. Overall, it looks like the bandwidth is not a factor in driving my results.

In the baseline model, I assume that the relationship between school FRL ratio and student achievement levels is piecewise linear. The relationship is shown in Figures 2.3 and 2.4, and this assumption looks correct. However, if it is not correct, my estimates may be biased in one direction or the other. Thus, in rows 4 and 5, I run the model using a piecewise quadratic function (allowing the slopes to differ across the threshold). In row 4, I do this at the original bandwidth of 0.1. The estimates for math are very similar, though with much larger standard errors, while the estimates for reading are again larger, though also with large standard errors. In row 5, I expand the sample to a bandwidth of 0.2. These estimates for both math and reading are similar to those in the original analysis, and the standard errors are of similar magnitude. Of the 32 coefficients across these two rows between the two tables, one is significant at the 10% level. It appears that my polynomial specification, too, is not driving the results showing a complete lack of impact of the policy.

While the coefficients do vary somewhat by specification and bandwidth, none of the changes are large enough to cause concern that the preferred specification is leading to biased estimates. There are very few significant coefficients, and the magnitudes are all very small. The estimates presented are indicative of the wider analysis, and not due to the preferred specification only.

2.6 Conclusion

In this paper, I examine the effects of schoolwide programs (versus targeted assistance programs) on the achievement of both FRL and non-FRL students using a regression discontinuity approach that exploits a policy in Title I, Part A determining school eligibility for these programs. I find that schoolwide programs have no impact whatsoever on student achievement in both math and reading. I examine whether the effects vary differentially across student groups, first analyzing by FRL status and then by quartiles of academic achievement. The lack of any impact is constant across groups, showing that it is not an average covering up more varied results.

Three possibilities arise from this. In one, schools eligible for a schoolwide program are simply continuing the use of the same interventions as those in a targeted assistance program, or at least of a generally similar nature. The other possibility is that schools below the threshold are not following the guidelines with respect to the strings of the Title I funds, and thus the students in question are not getting the interventions the policy mandates. There are reports due to ensure that these interventions are occurring, so this would mean a lack of a good reporting system or falsification of those reports. The last possibility is that schools on opposite sides are indeed treating students differently, but that these differences have no impact on student achievement.

The debate concerning the difference in educational achievement between socioeconomic groups has grown in the educational realm, and few remedies have been found. Furthermore, there has not been a causal study of the difference between the two program types on student achievement. These results contribute to the debate by doing just that – using a plausible causal analysis to examine how the program types effect students differentially. These results can also be applied to a wider debate on whether education should be a local or a broader concern. Even though the interventions mandated in a targeted assistance program are meant to help at-risk students, it appears that they have no effect when compared to programs designed by the schools themselves.

Tables

Table 2.1: Summary Statistics

	Full Sample	Below Threshold	Above Threshold
Student			
Female	0.496	0.494	0.497
White	0.528	0.580	0.498
Black	0.125	0.109	0.135
Hispanic	0.309	0.270	0.331
Spec. Ed	0.062	0.062	0.062
ESL	0.050	0.044	0.053
FRL	0.426	0.370	0.458
School			
Female	0.487	0.486	0.488
White	0.517	0.569	0.487
Black	0.128	0.111	0.138
Hispanic	0.314	0.276	0.337
Spec. Ed	0.115	0.113	0.116
ESL	0.043	0.038	0.046
FRL	0.442	0.385	0.475
Student Count	947.862	930.937	957.604
Baselines			
Baseline Math Score	0.106	0.159	0.076
Baseline Reading Score	0.133	0.183	0.104
Baseline Math Pass	0.800	0.816	0.791
Baseline Reading Pass	0.895	0.906	0.888
Students	2,138,719	781,321	1,357,398
Schools	5,745	2,130	3,615

Table 2.2: McCrary Test for Manipulation of Running Variable

Log discontinuity	0.008
	(0.061)
Bandwidth	0.084
Bin size	0.003
Students	1,820,149
Schools	4,849

Notes: Estimates from formal McCrary test in which the optimal bandwidth and bin size are determined using the data, after which the data is binned, allowing for the estimation of a quartic on both sides and the log of the discontinuity at the threshold.
*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table 2.3: Tests on Observables at Threshold

	Student Level	School Level
Female	0.000	0.000
	(0.002)	(0.002)
White	0.047***	0.025**
	(0.016)	(0.011)
Black	-0.006	-0.002
	(0.010)	(0.006)
Hispanic	-0.037***	-0.020**
	(0.013)	(0.009)
Spec. Ed	-0.004	0.001
	(0.003)	(0.002)
ESL	0.002	0.001
	(0.004)	(0.002)
FRL	-0.011***	-0.004
	(0.004)	(0.003)
Student Count		-54.317**
		(25.100)
Observations	2,138,719	5,745

Notes: All specifications clustered at school x year level. Each coefficient and associated standard error comes from a separate regression with a separate linear term on either side of the threshold and a dummy for being above the threshold without other controls.
*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table 2.4: Tests on Baseline Scores at Threshold

	Student Level	School Level
Baseline Math Score	-0.013	-0.006
	(0.009)	(0.010)
Baseline Reading Score	-0.012**	-0.004
	(0.006)	(0.007)
Baseline Math Pass	-0.006*	-0.003
	(0.003)	(0.003)
Baseline Reading Pass	-0.002	-0.000
	(0.002)	(0.002)
Observations	2,138,719	5,745

Notes: All specifications clustered at school x year level. Each coefficient and associated standard error comes from a separate regression with a separate linear term on either side of the threshold and a dummy for being above the threshold without other controls.
*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table 2.5: Estimates of Effect of Schoolwide Programs on Test Scores

	Math		Reading	
	Score	Pass Rate	Score	Pass Rate
SWP	-0.022	-0.004	0.007	-0.007
	(0.036)	(0.016)	(0.024)	(0.010)
Reduced Form	-0.004	-0.001	0.001	-0.001
	(0.006)	(0.003)	(0.004)	(0.002)
First Stage	0.171***	0.171***	0.171***	0.171***
	(0.027)	(0.027)	(0.027)	(0.027)
Student Controls	✓	✓	✓	✓
School Controls	✓	✓	✓	✓
Baseline Scores	✓	✓	✓	✓
N	2,138,719	2,138,719	2,138,719	2,138,719
Clusters	7,446	7,446	7,446	7,446
F-Test	38.970	38.970	38.970	38.970

Notes: All specifications clustered at school x year level. Each coefficient and associated standard error comes from a separate regression with a separate linear term on either side of the threshold and a dummy for being above the threshold.
*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table 2.6: Differential Estimates by FRL Status

	Math		Reading	
	Score	Pass Rate	Score	Pass Rate
SWP*Non-FRL	-0.030	-0.010	0.013	-0.003
	(0.035)	(0.014)	(0.024)	(0.009)
SWP*FRL	-0.002	0.007	0.001	-0.013
	(0.041)	(0.021)	(0.031)	(0.015)
Student Controls	✓	✓	✓	✓
School Controls	✓	✓	✓	✓
Baseline Scores	✓	✓	✓	✓
N	2,138,719	2,138,719	2,138,719	2,138,719
Clusters	7,446	7,446	7,446	7,446
P-value for equality	0.456	0.359	0.488	0.360

Notes: All specifications clustered at school x year level. Each coefficient estimated using a linear regression with separate dummies for FRL and non-FRL students, allowing the slope to vary on each side. FRL is also controlled for in these specifications. The p-value comes from a test of equality of the two parameters in the table.
*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table 2.7: Differential Estimates by Quartile

	Math		Reading	
	Score	Pass Rate	Score	Pass Rate
Quartile 1	0.024	0.033	0.016	-0.045
	(0.056)	(0.043)	(0.055)	(0.030)
Quartile 2	0.022	0.015	-0.050	-0.024*
	(0.050)	(0.025)	(0.035)	(0.014)
Quartile 3	-0.021	-0.030**	-0.004	0.005
	(0.036)	(0.015)	(0.023)	(0.008)
Quartile 4	-0.086**	-0.033*	0.048*	0.013
	(0.033)	(0.018)	(0.026)	(0.009)
Student Controls	✓	✓	✓	✓
School Controls	✓	✓	✓	✓
Baseline Scores	✓	✓	✓	✓
N	2,138,719	2,138,719	2,138,719	2,138,719
Clusters	7,446	7,446	7,446	7,446
P-value for equality	0.139	0.349	0.089	0.113

Notes: All specifications clustered at school x year level. Each column includes dummies by quartile, as well as the interaction of each covariate with each of the quartiles.
*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table 2.8: Estimates on Math Test Scores

	Scores				Pass Rate			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Baseline								
SWP	-0.001	-0.086	-0.078	-0.026	0.037	-0.019	-0.022	-0.006
	(0.079)	(0.061)	(0.060)	(0.036)	(0.039)	(0.022)	(0.022)	(0.016)
N	2,138,719	2,138,719	2,138,719	2,138,719	2,138,719	2,138,719	2,138,719	2,138,719
Bandwidth=.05								
SWP	-0.035	-0.229	-0.224	-0.113	0.029	-0.085	-0.089	-0.054
	(0.184)	(0.163)	(0.148)	(0.088)	(0.090)	(0.060)	(0.057)	(0.040)
N	1,149,178	1,149,178	1,149,178	1,149,178	1,149,178	1,149,178	1,149,178	1,149,178
Bandwidth=.2								
SWP	-0.007	-0.061	-0.051	-0.026	-0.006	-0.015	-0.016	-0.008
	(0.050)	(0.038)	(0.037)	(0.023)	(0.024)	(0.014)	(0.014)	(0.010)
N	3,762,090	3,762,090	3,762,090	3,762,090	3,762,090	3,762,090	3,762,090	3,762,090
Bandwidth=.1, quadratic								
SWP	-0.069	-0.292	-0.312	-0.158	0.053	-0.119	-0.130	-0.081
	(0.261)	(0.233)	(0.213)	(0.124)	(0.127)	(0.090)	(0.085)	(0.060)
N	2,138,719	2,138,719	2,138,719	2,138,719	2,138,719	2,138,719	2,138,719	2,138,719
Bandwidth=.2, quadratic								
SWP	-0.029	-0.115*	-0.106	-0.043	0.025	-0.031	-0.034	-0.013
	(0.090)	(0.069)	(0.067)	(0.040)	(0.044)	(0.025)	(0.025)	(0.018)
N	3,762,090	3,762,090	3,762,090	3,762,090	3,762,090	3,762,090	3,762,090	3,762,090
Student Controls		✓	✓	✓		✓	✓	✓
School Controls			✓	✓			✓	✓
Baseline Score				✓				✓

Notes: All specifications clustered at school x year level. Each coefficient and associated standard error comes from a separate regression with a separate linear term on either side of the threshold and a dummy for being above the threshold.
*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table 2.9: Estimates on Reading Test Scores

	Scores				Pass Rate			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Baseline								
SWP	0.031	-0.046	-0.043	-0.001	-0.028	-0.016	-0.019	-0.008
	(0.056)	(0.040)	(0.039)	(0.025)	(0.020)	(0.012)	(0.013)	(0.010)
N	2,138,719	2,138,719	2,138,719	2,138,719	2,138,719	2,138,719	2,138,719	2,138,719
Bandwidth=.05								
SWP	0.114	-0.037	-0.040	0.018	-0.003	-0.004	-0.012	0.002
	(0.137)	(0.094)	(0.086)	(0.056)	(0.046)	(0.029)	(0.027)	(0.023)
N	1,149,178	1,149,178	1,149,178	1,149,178	1,149,178	1,149,178	1,149,178	1,149,178
Bandwidth=.2								
SWP	-0.001	-0.051**	-0.042*	-0.010	-0.020*	-0.009	-0.008	-0.000
	(0.035)	(0.026)	(0.025)	(0.016)	(0.012)	(0.008)	(0.008)	(0.006)
N	3,762,090	3,762,090	3,762,090	3,762,090	3,762,090	3,762,090	3,762,090	3,762,090
Bandwidth=.1, quadratic								
SWP	0.154	-0.029	-0.043	0.040	0.004	0.009	-0.001	0.020
	(0.200)	(0.128)	(0.113)	(0.075)	(0.065)	(0.040)	(0.036)	(0.032)
N	2,138,719	2,138,719	2,138,719	2,138,719	2,138,719	2,138,719	2,138,719	2,138,719
Bandwidth=.2, quadratic								
SWP	0.012	-0.066	-0.062	-0.012	-0.031	-0.018	-0.021	-0.008
	(0.064)	(0.045)	(0.044)	(0.028)	(0.023)	(0.014)	(0.014)	(0.011)
N	3,762,090	3,762,090	3,762,090	3,762,090	3,762,090	3,762,090	3,762,090	3,762,090
Student Controls		✓	✓	✓		✓	✓	✓
School Controls			✓	✓			✓	✓
Baseline Score				✓				✓

Notes: All specifications clustered at school x year level. Each coefficient and associated standard error comes from a separate regression with a separate linear term on either side of the threshold and a dummy for being above the threshold.
*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Figures

Figure 2.1: First Stage

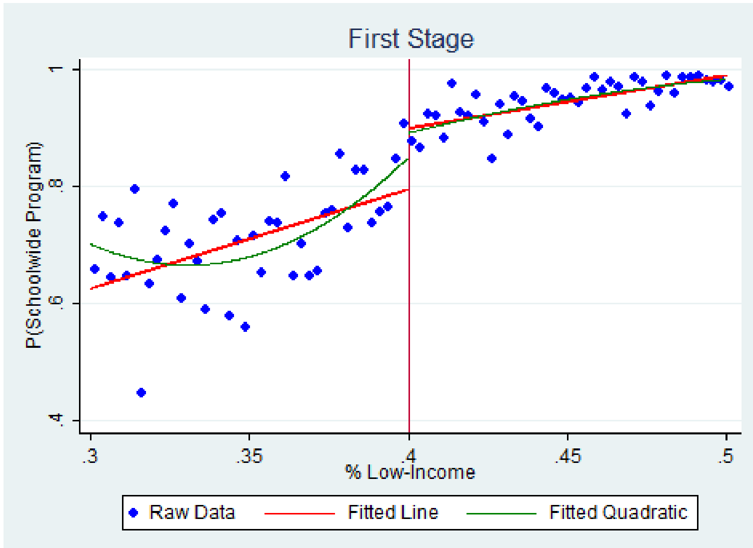


Figure 2.2: McCrary Test for Manipulation of Running Variable

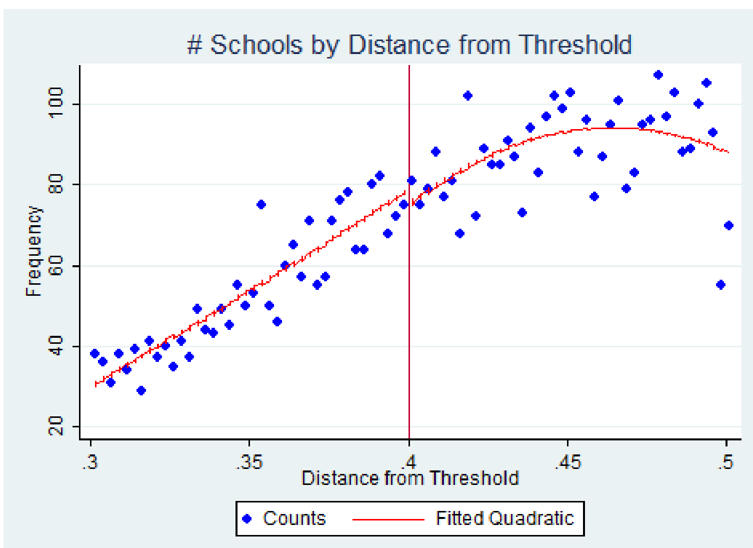


Figure 2.3: Passing Rates

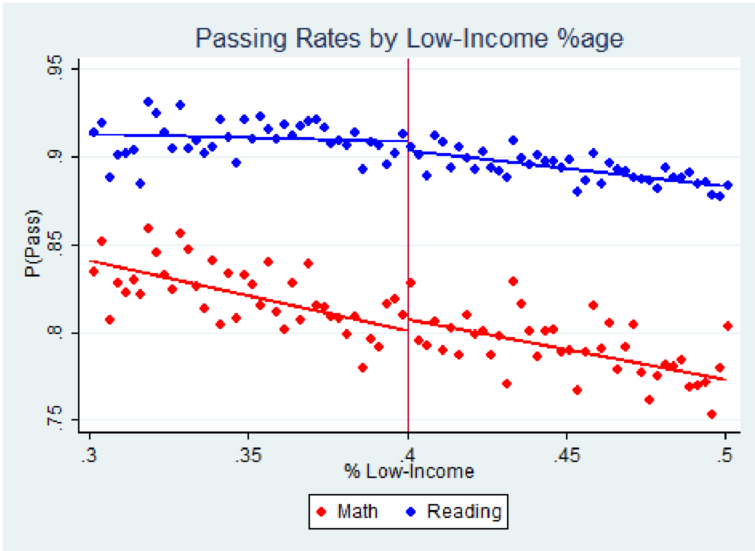
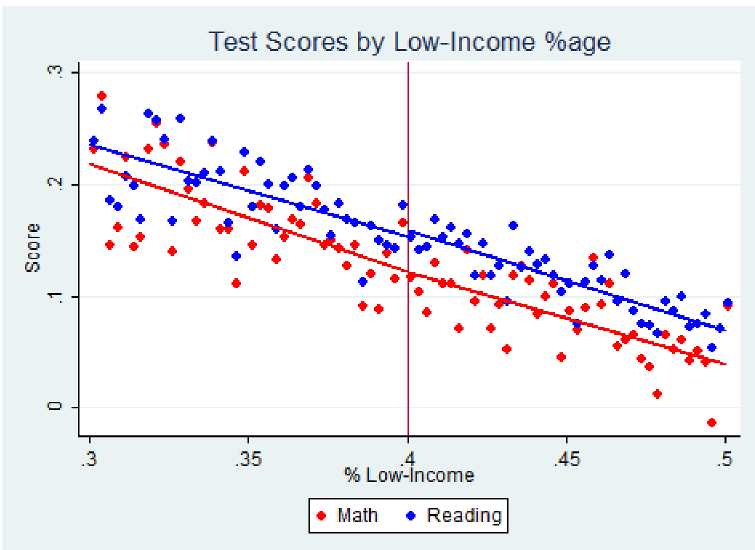


Figure 2.4: Scores



Chapter 3

Billions for What? The Impact of Title 1 Money on Student Achievement

3.1 Introduction

The federal government spends more than $14 billion per year on Title I funding to public schools across the country. There are various levels of accountability that accompany this large amount of money, but it is very difficult to calculate an estimate of the return on this spending.

Title I began in 1965 with the enactment of the Elementary and Secondary Education Act (ESEA), an act that largely redefined the federal government's role in local education. A piece of President Lyndon B. Johnson's ”War on Poverty”, the law created a large source of funding to help schools with high ratios of low-income children. ”In recognition of the special educational needs of low-income families...the Congress hereby declares it to be the policy...to provide financial assistance...to local educational agencies serving areas with concentrations of children from low-income families” (Section 201, Elementary and Secondary School Act, 1965)

While different versions of the Elementary and Secondary School Act have been passed by Congress throughout the years (most recently the Every Student Succeeds Act of 2015) with varying amounts of strings attached, Title I funding has been a large source of school district revenue since the original enactment in 1965. Since its inception, researchers have attempted to determine how Title I funding has impacted student achievement, the original stated goal of the program. This paper follows in this line of research, attempting to improve upon the previous studies in the quest to determine the relationship between spending and student outcomes.

Hanushek (1997) reviews the literature with respect to the relationship between school resources and student performance, concluding that "simple resource policies hold little hope for improving student outcomes." Borman and D'Agostino (1996) review the literature with respect to Title I more specifically, including 17 federal studies from 1966 to 1993, and find small positive effects that have become larger over time, perhaps because of increased oversight. However, many of the studies included in these meta-analyses cannot truly address the endogeneity of low-income student concentrations with respect to test scores.

To overcome this endogeneity, recent papers have used the within-district threshold for Title 1 status in regression discontinuity designs (Van der Klaauw 2008, Matsudaira et al. 2012). Van der Klaauw uses school-level data from New York City public schools to calculate the low-income student ratio threshold the district uses each year to determine which schools receive Title I funding and which do not. Matsudaira et al. use the same strategy for an unnamed urban school district, also using school-level data. Neither study finds positive impacts of Title I funding on student achievement, and Van der Klaauw actually finds negative impacts for two of his three study years. Both of these papers show high internal validity, but as is the case with regression discontinuity designs, suffer from a lack of external validity. Their estimates only apply to students in schools very much like those within the school district in question and very close to the threshold within that district for a given year.

In this paper, I attempt to answer the same question of Title I's impact on student achievement, but for a much wider range of schools. The obvious added benefit here is increased external validity – the results are much more applicable to students across the United States as a whole.

Furthermore, I have student-level data available to me, allowing for the inclusion of student-level covariates, increasing the sample size as well as the precision of my estimates. This level of data allows me to determine the impact of Title I funding on student growth, instead of just achievement, and lets me split the data into subgroups for which one may expect to see different impacts of Title I funding.

To achieve the aforementioned benefits, I have devised a strategy that allows for estimating the impact of Title I funding on student achievement across my entire sample. The data available is from public schools in Texas from the years 2004 through 2011. This is a large sample of very different looking schools and students, and as such, an innovative methodology is necessary. I exploit the funding provisions of Title I using an instrumental variables strategy. Each district must rank their schools in terms of low-income student ratio, then allocate Title I funds from the top down. There are exceptions, but in general schools lower in the district's distribution have a much lower probability of receiving Title I funding. I use each school's percentile within district as an instrument, conditional on district-by-year fixed effects, in an attempt to overcome the endogeneity present as well as take advantage of the large data set available. If each school's percentile within district-by-year grouping impacts student achievement only through Title I status, then this instrument is valid and my results are causal. However, in performing the analysis, I find that the instrument violates the exclusion restriction, and as such, the estimates are not causal, and may only be thought of as a correlative lower bound.

That the estimates presented represent correlative lower bounds comes from Nevo and Rosen [2012], in which the authors build on several previous papers to determine that under certain conditions, imperfect instruments may be helpful in providing bounds for point estimates rather than being thrown out altogether. In this case, several assumptions are met, and because the estimates from a simple OLS model are less than those in the IV model, the IV point estimates are lower bounds on the true point estimates.

I find that a school being designated as Title I increases Title I funding an average of $309 per student, and that this is not accompanied with an increase or decrease in state or other federal funding. However, when looking at funding by school level, I find only a significant increase in Title I funding per student among middle schools, not elementary or high schools. Along with that increase in funding, I find an increase in math passing rates of 2.9 percentage points and no change in reading passing rates. I see no overall impact on standardized score in either math or reading. As lower bounds, this simply informs that Title I money does not have a negative impact on scores, but does have a positive impact on math passing rates.

I then split the sample by low-income student status, school level, and student quartile (per previous year's scores). I find that Title I funding may decrease passing rates and scores among elementary school students, regardless of low-income status. On the other hand, I find large, though insignificant, impacts on both math and reading exams for middle school students, particularly those who are low-income or in the lower two quartiles of the test score distribution. I see no impact among high school students, though this is likely due to a very weak first stage at this level.

One might hope, given the total amount of federal expenditures on the Title I program, to see better returns than those detailed above. However, there is evidence of large, positive effects among the middle school students at whom the program is most targeted. More investigation is necessary to determine why such effects are found among middle school students in contrast to those found among elementary school students. Lastly, while I cannot currently devise a method to solve the lack of adherence to the exclusion restriction in this paper, it may be possible with more ingenuity.

This paper is structured as follows. Section 2 provides an overview of the funding guidelines for the Title I program and Section 3 introduces the data and research design. Section 4 presents estimates of the impact of Title I funding on student achievement and Section 5 discusses the implications of these findings.

3.2 Institutional Background

Title I was originally signed by Lyndon B. Johnson in 1965 as part of the Elementary and Secondary Education Act (ESEA) (Pub. L. 89-10, 79 Stat. 27, 20 U.S.C. ch. 70). Since 1965, it has been the largest source of federal government funding for K-12 education. In 2011, more than $14 billion of federal spending was allocated to Title I, and currently more than 50% of public schools receive public funding. Titled "Better Schooling for Educationally Deprived Children", its purpose was to give financial assistance to local educational agencies with high amounts of low-income children. The thought was that these students, being low-income themselves, who attended schools with other low-income students, had a much more difficult path to success than their educational peers.

In order to determine which schools receive Title I, Part A funds, a district must follow detailed guidelines. The focus of this paper is on larger districts, as smaller districts (those with fewer than 1,000 students) can choose to serve any or all of their campuses with Title I funds. In larger districts, a campus is eligible to receive Title I, Part A funds if the percentage of low-income students residing in the attendance area is equal to or greater than that of the district as a whole. Campuses with greater than 75% poverty must be served first. After that consideration, districts must rank their campuses in terms of low-income percentage and then serve the campuses from the top down. This ranking serves as the method of identification in this paper. However, there are many exceptions, which may explain why the instrument used is not perfect.

First, districts may choose to rank order within grade span rather than as an entire district. These are very clearly correlated, and thus, for simplicity's sake, I will use district-wide ranking. Secondly, a district may not serve any campus that has a low-income ratio of less than 35%. Lastly, a district may choose to serve any ineligible campus that was eligible last year for one additional transition year.

In terms of how these exceptions impact my analysis, they do not bias my results. They describe more who falls into the complier and non-complier category, and how to interpret the results. The school districts without exceptions will fall into the complier category, and the results will describe the impact of the additional funding in these districts.

In terms of how much money to allocate, a district does not have to allocate the same amount per child for each campus. However, a district may not allocate more per child to a campus with a lower poverty rate than to a campus with a higher poverty rate. The amount varies by state, by district, and by school, and thus, there is no one single amount.

3.3 Data and Research Design

My analysis is based upon data from the Texas Educational Resource Center (ERC), including student-level demographic, attendance, and test score data. The demographic files contain information on age, gender, free or reduced lunch (FRL) status, ethnicity, special education status, English as a second language (ESL) status, limited English proficiency (LEP) status, and gifted and talented status (GT). The attendance files contain information on enrollment for each student across each school year, including if and when a student may have transferred mid-year.

The test score data is from the Texas Assessment of Knowledge and Skills (TAKS) exams, developed by Pearson, and conducted in the springs of third grade through eleventh grade. Students are tested in both math and reading every year. All public school students are required to take the TAKS exams unless they have a severe disability. Those with moderate disabilities are given accommodations, but still take the exams. In my analysis, the test scores are normalized to have a mean of zero and a standard deviation of one for each grade, subject, and year across the entire Texas sample. For simplicity's sake, I focus on the math and reading exams.

Students were matched across these files both cross-sectionally and longitudinally using de-identified student ID numbers. Using the fall enrollment dates and the testing dates, I then constructed school-level variables, including gender, ethnicity, FRL, special education, ESL, LEP, and gifted/talented ratios. These variables are used as additional controls throughout the analysis.

I match this data with the school-level data set that includes an indicator for whether a school is a Title I school. I use the school outlay data set and collapse the school expenditures down by source of funding, so that I have amounts spent by each school each year using Title I funding, state funding, and other federal funding. After merging the data, I divide these amounts by the student counts to create a per capita Title I spending variable, a per capita state funds spending variable, and a per capital total spending variable.

Finally, I construct a district-level low-income ratio and rank schools by low-income ratio within district and year. I use this rank to generate a percentile, which will be used as the conditional instrument for analysis.

I then restrict the schools per the various guidelines pertaining to which schools are eligible to receive Title I funding. I exclude any schools from small districts due to their many exemptions. Schools with student bodies below 35% low-income are not eligible and schools with student bodies above 75% low-income are guaranteed funding, so I include only schools within these two thresholds. Lastly, schools with student populations below the district low-income ratio are not eligible, so I exclude them as well. After one more exclusion I explain in detail in the next section, the sample remaining includes 1,054,283 students in 2,122 campus-by-year groups.

As seen in Table 3.1, my resulting sample is somewhat similar to that of Texas as a whole during the period. My sample has more low-income students but less of both gifted and special education students. As one of the principal contributions of this paper is the external validity as compared to earlier work, it is important that this sample be representative of the wider student population as a whole.

3.3.1 Research Design

The purpose of this paper is to determine the impact that Title I funding has on student achievement. Ordinary least squares is fraught with issues, as Title I funding is given to schools with high poverty ratios, and high poverty is strongly correlated with lower performance on exams. I demonstrate the relationship between Title I designation and a school's poverty ratio in Figure 3.1. This type of analysis would not be causal, and would be very negatively biased. To overcome this endogeneity, I exploit a funding provision with an instrumental variables approach.

As stated earlier, Title I designations and funding at the school level are decided by district administrators. They rank their schools in terms of low-income student ratios, and then must apply the Title I designation from the top down. This allows me to use school rank within district as an instrument for Title I funding, conditional on school and student characteristics. I operationalize this rank by constructing a percentile for each school of their free-and-reduced lunch ratio within each district by year group.

In the first stage, as detailed here,

$$TitleI_{sdt} = \alpha + \beta_1 f(Percentile)_{sdt} + X_{isdt}\pi + Z_{st}\gamma + D_{dt}\eta + \epsilon_{ist} \tag{3.1}$$

I estimate the impact of the school's percentile within district on the probability of receiving Title I funding, conditional on school and student characteristics and district-by-year fixed effects. This is for student i in school s in district d in year t. X is a vector of student characteristics, including gender, race, grade, free-and-reduced-lunch status, gifted status, special education status, English as a secondary language status, and previous year's math and reading standardized scores. Z is a vector of school-level characteristics, including ratios for gender, race, free-and-reduced lunch, special education, English as a secondary language, and student body count. D is a vector of district-by-year fixed effects, and ϵ is the error term. The coefficient of interest is β_1, which measures the increase in probability of Title I designation corresponding to an increase in the school's percentile within district-by-year group.

In the second stage,

$$y_{isdt} = \alpha + \beta_2 \widehat{TitleI}_{sdt} + X_{isdt}\pi + Z_{st}\gamma + D_{dt}\eta + \epsilon_{ist} \tag{3.2}$$

,

I exclude the percentile variable, but include the predicted probability of Title I funding estimated in the first stage. All other controls are the same, and the coefficient of interest is β_2. This coefficient measures the impact of Title I funding on my outcomes, which include probability of passing math and reading exams and standardized scores on both. All estimates in both stages are clustered at the school-by-year level, the level of treatment.

3.3.2 Instrument Validity

In order for the research design to be considered causal, it must satisfy the two main assumptions of an instrumental variables design: relevance and the exclusion restriction. The first piece is easier to show. In Figure 3.2, I group schools into bins by their percentile of low-income ratio within the appropriate district-by-year group and display the means of Title I probability for each of those bins. The simple relationship shown is very strong and positive. In Table 3.2, I display the coefficient from Equation 3.1 using a variety of covariates at the school-by-year level. The first model, with simple district-by-year fixed effects, shows a very strong relationship between percentile and Title I probability. However, this model ignores school demographics as well as the school's FRL ratio, and most likely violates the exclusion restriction. The second column includes FRL ratio, and percentile is no longer relevant to Title I probability.

This is likely because district-by-year fixed effects and a school's FRL ratio together completely determine a school's percentile within district. Columns three and four correspond to one and two, but include a vector of school demographics. Column three is my preferred specification going forward. While I would like to include the school's FRL ratio, the instrument is no longer strong enough once I do so in column 4.

A school's percentile within district and FRL ratio are closely related, and I cannot include both, as shown in column 4. However, when I drop FRL ratio, one might worry about violating the exclusion restriction. For this not to be a worry, I need to use districts in which the schools have very similar FRL ratios. Then the percentile is informative as to Title I designation, but not to the school's demographic composition. In Figure 3.4, I plot each district's variance of school FRL ratios over its overall FRL ratio, weighted by the number of students in the district. There does not appear to be any relationship present, so excluding districts with large within-district variances should not pose any problem.

The next question is how to cut the sample. In Figure 3.5, I cut the sample four times and plot the first stage coefficient for each of them. The first sample includes schools in districts with variance in only the first quartile of district variances, the second sample includes schools in districts with variances below the median, the third sample includes schools in districts in the first three quartiles, and the last sample is all schools. In Figure 3.6, I plot the coefficients for each of the four outcomes by the same four samples. In each of these, it seems that the smallest and largest samples have estimates much different from the rest.

The smallest sample lacks precision due to the size, and is small enough to impact external validity. The largest sample, which includes districts with large within variances, is most impacted by the exclusion of the FRL ratio as a covariate. For the remainder of the paper, I use the third sample, in which I exclude schools in districts in the top quartile of the within-district variance measure. This sample leaves enough students for precise estimates and external validity, but does not suffer from a violation of the exclusion restriction as much as the complete sample.

In Table 3.3, I replicate Table 3.2, but at the student level. Column five here is similar to column three in the previous table, but includes a vector of student-level covariates. Using column five as the preferred specification, it appears that a 10 percentage point increase in a school's percentile of FRL ratio within district increases the school's probability of receiving Title I funding by 8.38 percentage points.

In Table 3.4, I show the IV estimate of Title I designation on federal Title I money per student, state money per student, and total spending per student minus Title I in separate regressions. Column 3 shows the results from the preferred specification, and it appears that a Title I designation increases Title I funding per student by $309. Some evidence for the exclusion restriction is shown in the second and third column, in that Title I schools do not receive differing amounts of state or total non-Title I funding per student. The point estimates here are positive, but with very large standard errors, which is most likely because these sources of funding are determined by a large variety of outside factors for which this study cannot control. The first column in this table shows why the simplest model is not valid. Without controlling for student and school characteristics, a school's percentile is highly correlated with factors that determine other sources of funding.

In Table 3.5, I regress the model separately on each student covariate as an outcome. Unfortunately, while none of the coefficients on individual gender or race are significant, a 10% increase in percentile corresponds with a .24 pp decrease in likelihood of being an ESL student, a 1.3 pp increase in likelihood of being a low-income student, and a 0.3 pp decrease in being a special education student. The most worrisome of all of these is the high magnitude and significance level of the FRL variable. This points to the idea that the instrument, even with restricting the sample on variance, is not accounting for all the difference in FRL ratios between schools. This obviously violates the exclusion principle, and as such, the results discussed hereafter must be taken as correlative more than causal.

In fact, following the thought process set out in Nevo and Rosen [2012], the instrumental variable estimates presented in this papers may be thought of as lower bounds for the true estimates. Under three assumptions, the authors show that in this case, the true point estimates are equal to or greater than the minimum of the OLS estimates and the IV estimates. These assumptions all hold, starting with A1: random sampling. A2 states that the imperfect instrument Z is correlated with the unobservable U, but that it is less so than the endogenous regressor X. In this case, that means that a school's percentile within district, conditional on district-by-year fixed effects, is less correlated with U than whether or not the school receives Title I money. Finally, A3 states that the correlations between the instrument and the error term and the endogenous regressor and the error term are both positive or both negative. In this case, both are negative. Thus, all three assumptions hold, which means that we can apply Lemma 1. In this case, that states that because the correlation between a school's percentile and its Title I status is positive, then because the correlation between Title I status and the error term are negative, $\beta \geq \max\{\beta^{OLS}, \beta_z^{IV}\}$. In this case, the OLS estimates are all below the IV estimates, and thus, the IV estimates presented here may be thought of as lower bounds on the true beta.

Because the coefficients arising from IV models are local average treatment effects (LATE), it is necessary to discuss to which schools and districts the estimates in this paper apply. The advantage of this research design over an RD design is the external validity. In the RD design papers, the estimates only speak to schools right around the thresholds in the given district. However, in this research design, the estimates only speak to districts that follow the rules without many exceptions. Due to the explicit exceptions, this includes schools in districts with more than 1,000 students that have a low-income student ratio of greater than 35% but less than 75% and higher than the district average. The nationwide average falls just above 50%, so these schools are somewhat centered around the national average. In general, the LATE in this case applies to schools in non-small districts who are neither very high nor very low in low-income students.

3.4 Results

Due to the lack of compliance with the exclusion restriction, but with the assistance of Nevo and Rosen [2012], it is important to remember that the results shown in the paper present lower bounds on the true point estimates. The impact of Title I funding on student achievement is shown for all students in Table 3.6. I compute the impact of the funding for four different outcomes, math and reading passing rates and math and reading standardized scores. These impacts are shown in the first row. A Title I designation increases a student's probability of passing the math exam by 2.9 percentage points, from a baseline of about 76%. There is no effect on math scores, however. On the reading exam, there is no impact on either passing rates or scores due to a Title I designation. The second row shows the first stage, which is the same for all outcomes. As lower bounds, these zeroes simply show that Title I money does not have a negative impact on testing outcomes, but does have a significant, positive impact on math passing rates.

In Table 3.7, I show the same results, but with the amount of Title I money as the first stage outcome. In this case, the first stage estimate is 0.259, which means that a 10% increase in percentile leads to $25.9 more Title I dollars per student. However, the first stage in this case is not strong enough, with an F-stat of only 7.1. As a consequence, while the coefficients on the second stage are perhaps larger, the standard errors are too.

Both tables show average impacts only, and as Title I is designed to assist specific groups of students, it is necessary to dig deeper. Because of this, I examine the results by various levels of heterogeneity, including low-income status, school level, and student quartile. In the rest of this section, I discuss these results in attempt to solve this puzzle.

3.4.1 Differential Impacts

3.4.1.1 FRL Status

Table 3.8 displays results split by free-and-reduced lunch status. In each column, I interact percentile with both outcomes of an FRL dummy variable as instruments for the Title I indicator interacted with both outcomes of the FRL dummy variable. I do not interact all the controls, but do include the individual's FRL status as a covariate. In math, it appears that Title I funding benefits the low-income students, though not at the expense of the non low-income students. Low-income students in Title I schools pass the math exams at a rate of 6.7 pp higher than their non Title I peers. The difference in standardized score is .03σ, though not significant. Low-income students in Title I schools pass reading exams at a rate of 2 pp lower than their non-Title I peers, and score slightly worse, though again, not significantly so. It appears that a Title I designation increases performance on the math test for low-income students, and does not have a negative impact on reading scores, while not negatively impacting the non low-income students. It is possible that Title I money even increases scores for these students, as these are lower bounds.

3.4.1.2 School Level

The disparate results between math and reading are not solved by splitting across FRL status, but there is a much clearer pattern when splitting the impacts by school level. In Table 3.9, I split the sample into elementary schools, middle schools, and high schools per the Texas Education Agency's designations. The far right column details the amount of additional money schools in each level receive per student when designated Title I. For both elementary and middle schools, it appears they may receive more money, but the standard errors are too large for statistical significance. The magnitude increase for elementary schools due to Title I designation is \$391 per student, and for middle schools it is \$225 per student.. The point estimate on high schools is quite high, at at additional \$1,365 per student, but the standard errors for all the high school outcomes are quite high. The first stage for high schools is not very precise, which leads to imprecise second stage estimates. I believe this is because high school Title I designations are more prone to using exceptions mentioned in Section 2, which are too numerous to detail again here.

In the first row, it is clear that a Title I designation at the elementary level hurts both scores and passing rates. Scores of elementary school students in Title I schools are 0.37σ lower in math and 0.27σ lower in reading, while passing rates are 0.03 pp lower in math and 0.17 lower in reading. Even as lower bounds, these impacts are quite large in magnitude. These schools do not receive less money, so there might be something about the designation itself that is causing these scores to be lower.

The middle school results show a different story. Students in Title I middle schools pass math at a rate 8 pp higher and reading at a rate of 15 pp higher than those in non-Title I middle schools. Middle school Title I students score almost 0.4σ higher in math and $.26\sigma$ in reading, though again, none of these estimates are significant at even the 10% level. However, these are also very large impacts, and more investigation into how Title I money is implemented at the elementary school and middle school levels is necessary.

3.4.1.3 Quartiles

In order to determine if the Title I money is helping the students most in need of intervention, I split the students into quartiles based on their previous year's scores. Before paring down the sample, I average each student's standardized scores across math and reading, then split the students into quartiles based on this average.

Once again, the pattern in math is much more clear than the pattern in reading. Title I students in the lower two quartiles perform much better than their peers in non-Title I schools in math, and this comes at the expense of the students in the upper two quartiles. In the first quartile, students in Title I schools pass at rates 24 pp higher than their peers and score 0.09σ higher. Second quartile students show similar, strong impacts, with coefficients of 19 pp and 0.18σ. However, students above the median are harmed in terms of math scores by Title I status. Third quartile students in Title I schools pass math exams at a rate of 10 pp lower than their peers, while fourth quartile students pass at a rate of 19 pp lower than their peers and score 0.22σ lower. If Title I funding is meant to help lower-performing students, it certainly hits the mark in math exams, but at the expense of the higher-performing students.

However, in reading, this same pattern does not hold. While the estimates for first quartile students show at worst a modest negative impact on reading performance, both second and third quartile students are harmed by Title I money. Second quartile students in Title I schools pass reading exams at a rate at worst 4 pp lower and score 0.07σ lower than their peers, while third quartile students in Title I schools score at worst 0.06σ lower, though there is no impact on their passing rates. Oddly enough, fourth quartile students seem to benefit from Title I designation, with those in Title I schools scoring $.08\sigma$ higher than those outside of such schools.

It appears that in math, Title I funding increases test scores for low-income or low-performing students at the expense of the non low-income or high-performing students. The positive effects are shown mostly at the middle school level. However, in reading, both non low-income and low-income students show lower scores in Title I schools, and this impact occurs mostly in the middle of the distribution. Unfortunately, this pattern is tough to decipher, and in tables not shown here, I estimate impacts by the interaction of school level, prior performance, and low-income designation. These tables yield no more information.

3.5 Discussion

The federal government spends over $14 billion per year on the Title I program in hopes of assisting schools with high concentrations of low-income students. However, conclusions as to the impacts of this money are lacking. While there have been many reports and various research studies attempting to determine the program's effects, most cannot overcome issues related to the strong relationship between low-income concentration and test scores. Authors that have been able to overcome this issue as of late have used a regression discontinuity design, writing papers with high internal validity that suffer from weak external validity. The instrumental variable design of this paper allows me to use a sample from across the state of Texas, greatly increasing the external validity. However, the internal validity of this paper is not nearly as high as those previous, and estimates within should be regarded as more correlative than causal.

Title I funding does not have consistent impacts for math and reading, nor does it have consistent impacts across heterogeneous student groupings. Overall, the funding appears to increase math passing rates, with no impact on reading passing rates nor on standardized scores on either test. The impacts in math are highly dependent on previous scores. Students who scored below the median the previous year show large increases in passing rates and scores, while students above the median show large decreases in both outcomes. In reading, there is a slight negative impact on both passing rates and scores, but not significant among almost any group. The one group shown to have decreased performance on reading exams due to Title I designation is those in the second quartile. While the estimates presented represent lower bounds on the true point estimates, the large magnitude of the negative estimates for these students cannot rule out true negative estimates. These results cannot speak to which level of school most effectively uses Title I funds, as the sample size is simply not large enough. Furthermore, the instrument used is not strong enough to estimate impacts among high schools.

Without knowing exactly how Title I funding is used in schools, it is difficult to dive into these results much more. It is possible that in middle schools, where students are much less closely tied to particular teachers, more money is easily used. As to why extra funding can be linked to negative outcomes, perhaps there is some type of stigma associated with a Title I designation that affects the student or teacher sample. However, as these estimates are lower bounds, the negative estimates may not be closely tied to the true point estimates.

The results shown indicate that funding may be working as advertised in middle schools. The point estimates are quite large among middle school students, though not significant. It is important to determine the mechanism for these types of impacts. If such an investigation can lead to a deeper understanding of how schools can use additional money to best impact student achievement, students all over will benefit, and the large amounts of money spent on public schools in the United States can be put to use more productively.

Tables & Figures

Table 3.1: Summary Statistics

	All Texas Students	Sample
Female	0.49	0.49
	(0.50)	(0.50)
White	0.34	0.32
	(0.48)	(0.47)
Black	0.16	0.20
	(0.37)	(0.40)
Hispanic	0.43	0.42
	(0.50)	(0.49)
English as Second Language	0.07	0.06
	(0.25)	(0.23)
Free or Reduced Lunch	0.50	0.55
	(0.50)	(0.50)
Gifted	0.10	0.08
	(0.31)	(0.27)
Special Ed	0.10	0.06
	(0.30)	(0.25)
N	13,705,645	1,054,283

Table 3.2: First Stage Estimates – School Level

	(1)	(2)	(3)	(4)
Percentile	1.389***	0.236	1.095***	0.198
	(0.105)	(0.187)	(0.123)	(0.181)
N	2,122	2,122	2,122	2,122
Mean	.701	.701	.701	.701
F-stat	173.8	1.6	79.1	1.2
District-by-Year FE	✓	✓	✓	✓
School demographics			✓	✓
FRL Ratio		✓		✓

Notes: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the campus level.

Table 3.3: First Stage Estimates – Student Level

	(1)	(2)	(3)	(4)	(5)	(6)
Percentile	1.511***	0.307*	0.992***	0.298*	0.838***	0.067
	(0.101)	(0.166)	(0.116)	(0.159)	(0.112)	(0.158)
N	1,054,283	1,054,283	1,054,283	1,054,283	1,054,283	1,054,283
Mean	.611	.611	.611	.611	.611	.611
F-stat	225.8	3.4	73.6	3.5	56.0	0.2
District-by-Year FE	✓	✓	✓	✓	✓	✓
School demographics			✓	✓	✓	✓
Student demographics					✓	✓
FRL Ratio		✓		✓		✓

Notes: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the campus level.

Table 3.4: Impact on $ Per Capita ($1000s)

	(1)	(2)	(3)
Impact on Title I $ Per Capita			
Title I Status	0.321***	0.263*	0.309*
	(0.079)	(0.150)	(0.177)
Mean	0.354	0.354	0.354
Impact on State $ Per Capita			
Title I Status	1.353***	0.397	0.932
	(0.398)	(0.722)	(0.834)
Mean	6.818	6.818	6.818
Impact on All Non Title I $ Per Capita			
Title I Status	2.273**	0.723	1.936
	(1.051)	(1.952)	(2.299)
Mean	9.902	9.902	9.902
First Stage	1.511***	0.992***	0.838***
	(0.101)	(0.116)	(0.112)
N	1,054,283	1,054,283	1,054,283
District-by-Year FE	✓	✓	✓
School demographics		✓	✓
Student demographics			✓

Notes: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the campus-by-year level.

Table 3.5: Impact of Instrument on Observables

	Female	White	Black	Hispanic
Percentile	0.007	0.005	0.002	0.002
	(0.005)	(0.005)	(0.004)	(0.004)
Mean	0.495	0.323	0.199	0.423

	ESL	FRL	Gifted	Special Ed
Percentile	-0.024***	0.132***	0.011*	-0.030***
	(0.004)	(0.015)	(0.006)	(0.010)
Mean	0.059	0.548	0.081	0.064
N	1,054,283	1,054,283	1,054,283	1,054,283
School demographics	✓	✓	✓	✓
District-by-Year FE	✓	✓	✓	✓

Notes: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the campus-by-year level.

Table 3.6: Main Results

	Math		Reading	
	Pass	Score	Pass	Score
Title I Receipt	0.029**	0.016	-0.011	0.004
	(0.015)	(0.033)	(0.009)	(0.021)
First Stage	0.838***	0.838***	0.838***	0.838***
	(0.112)	(0.112)	(0.112)	(0.112)
N	1,054,283	1,054,283	1,054,283	1,054,283
Mean	0.762	-0.016	0.849	0.014
First Stage F-stat	56.0	56.0	56.0	56.0
Student demographics	✓	✓	✓	✓
School demographics	✓	✓	✓	✓
District-by-Year FE	✓	✓	✓	✓

Notes: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the campus-by-year level.

Table 3.7: Main Results Using Title 1 Money as First Stage

	Math		Reading	
	Pass	Score	Pass	Score
Title I Money ($1000s)	0.095	0.050	-0.034	0.011
	(0.060)	(0.111)	(0.034)	(0.067)
First Stage	0.259***	0.259***	0.259***	0.259***
	(0.097)	(0.097)	(0.097)	(0.097)
N	1,054,283	1,054,283	1,054,283	1,054,283
Mean	0.762	-0.016	0.849	0.014
First Stage F-stat	7.1	7.1	7.1	7.1
Student demographics	✓	✓	✓	✓
School demographics	✓	✓	✓	✓
District-by-Year FE	✓	✓	✓	✓

Notes: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the campus-by-year level.

Table 3.8: Estimates by FRL Status

	Math		Reading	
	Pass	Score	Pass	Score
Title I Receipt * FRL	0.067***	0.028	-0.020*	-0.023
	(0.017)	(0.034)	(0.011)	(0.024)
Title I Receipt * Not FRL	-0.014	-0.002	-0.000	0.033
	(0.016)	(0.035)	(0.011)	(0.023)
N	1,054,283	1,054,283	1,054,283	1,054,283
Mean	0.762	-0.016	0.849	0.014
Student demographics	✓	✓	✓	✓
School demographics	✓	✓	✓	✓
District-by-Year FE	✓	✓	✓	✓

Notes: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the campus-by-year level.

Table 3.9: Estimates by School Level

	Math		Reading		Title I $$
	Pass	Score	Pass	Score	Per Capita
Title I Receipt * Elementary	-0.027	-0.372	-0.170	-0.265	0.391
	(0.096)	(0.276)	(0.111)	(0.178)	(0.292)
Title I Receipt * Middle	0.082	0.394	0.151	0.263	0.225
	(0.133)	(0.368)	(0.154)	(0.234)	(0.379)
Title I Receipt * High	-0.634	-1.724	-0.787	-1.087	1.365
	(1.679)	(4.546)	(2.011)	(2.931)	(3.769)
N	1,054,283	1,054,283	1,054,283	1,054,283	1,054,283
Mean	0.762	-0.016	0.849	0.014	0.354
Student demographics	✓	✓	✓	✓	✓
School demographics	✓	✓	✓	✓	✓
District-by-Year FE	✓	✓	✓	✓	✓

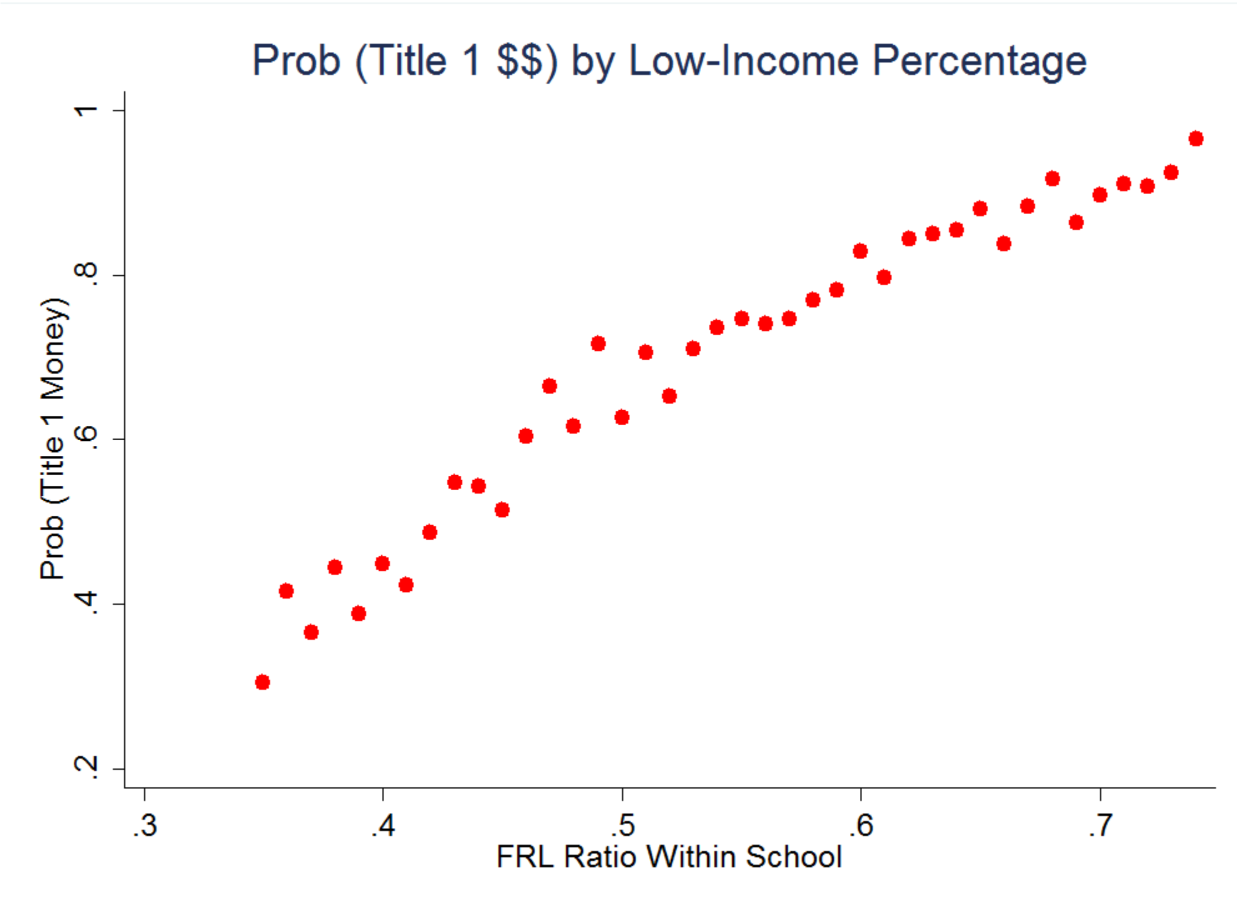
Notes: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the campus-by-year level.

Table 3.10: Estimates by Prior Student Performance

	Math		Reading	
	Pass	Score	Pass	Score
Title I Receipt * 1st Quartile	0.237***	0.086*	-0.042	0.066
	(0.047)	(0.048)	(0.028)	(0.042)
Title I Receipt * 2nd Quartile	0.185***	0.180***	-0.044***	-0.065***
	(0.035)	(0.050)	(0.015)	(0.032)
Title I Receipt * 3rd Quartile	-0.104***	0.027	0.012	-0.056*
	(0.025)	(0.037)	(0.013)	(0.030)
Title I Receipt * 4th Quartile	-0.193***	-0.222***	0.022	0.075*
	(0.043)	(0.056)	(0.017)	(0.041)
N	1,054,283	1,054,283	1,054,283	1,054,283
Mean	0.762	-0.016	0.849	0.014
Student demographics	✓	✓	✓	✓
School demographics	✓	✓	✓	✓
District-by-Year FE	✓	✓	✓	✓

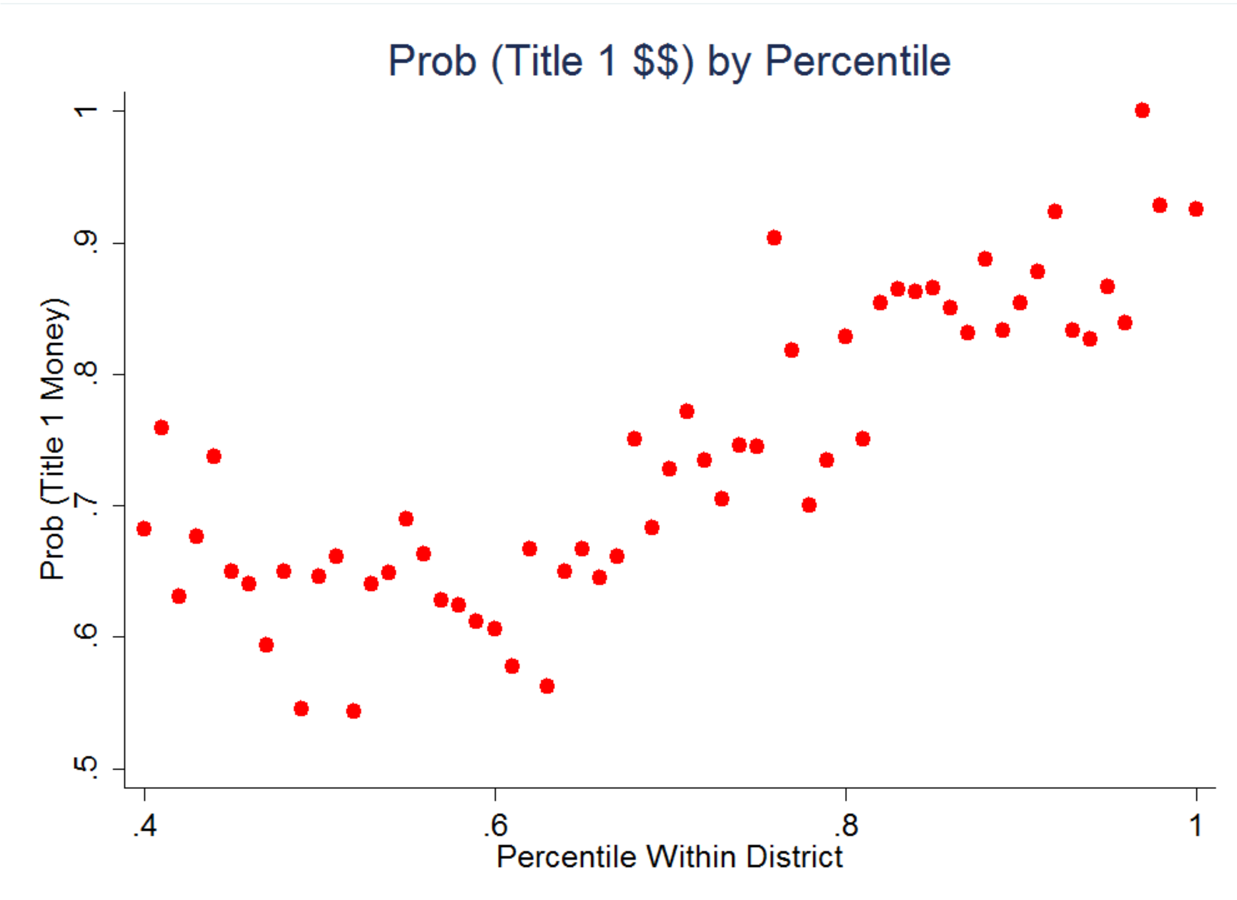
Notes: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the campus-by-year level.

Figure 3.1: Title I Probability by FRL Ratio



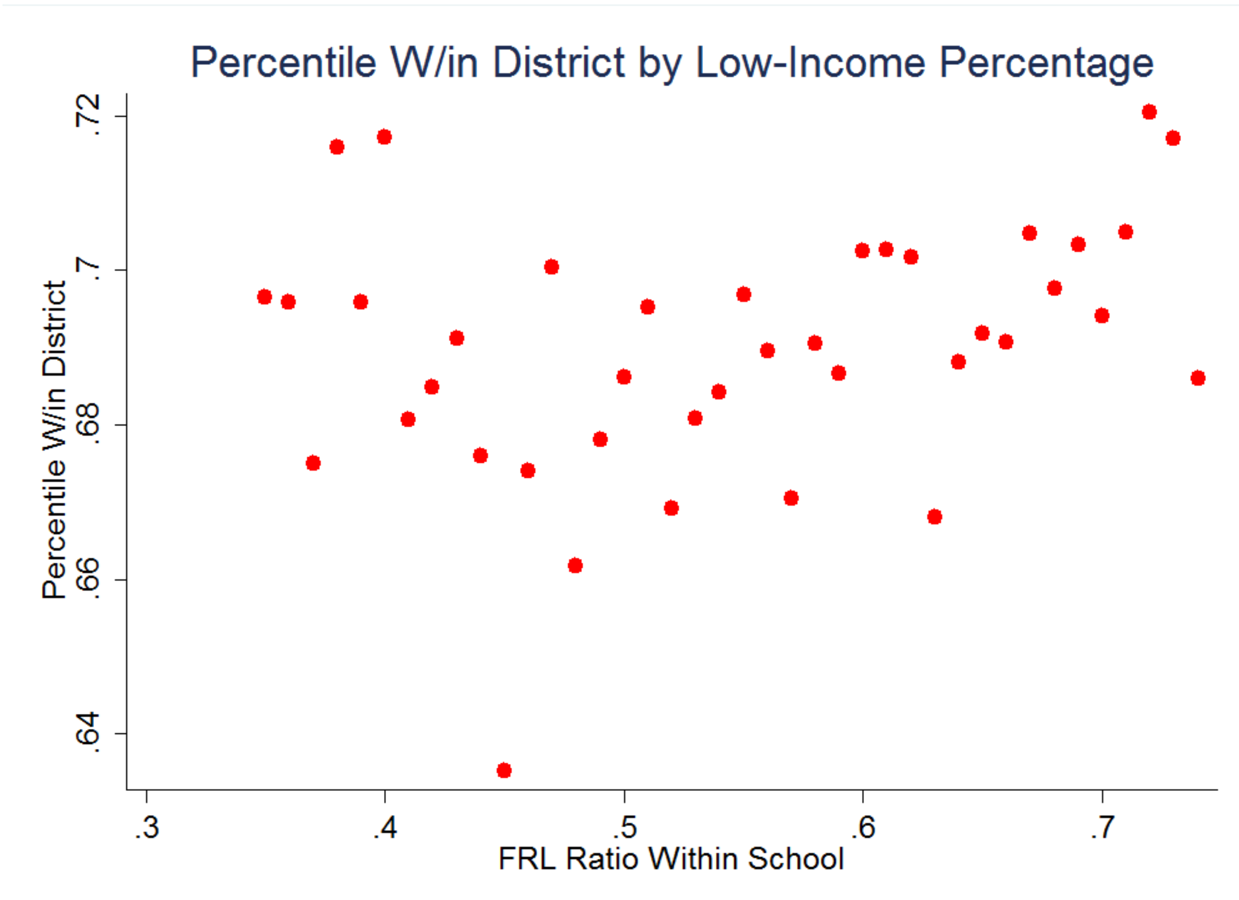
Notes: The figure shows the mean probability of Title I designation by the low-income ratio at the school.

Figure 3.2: Title I Probability by Percentile within District



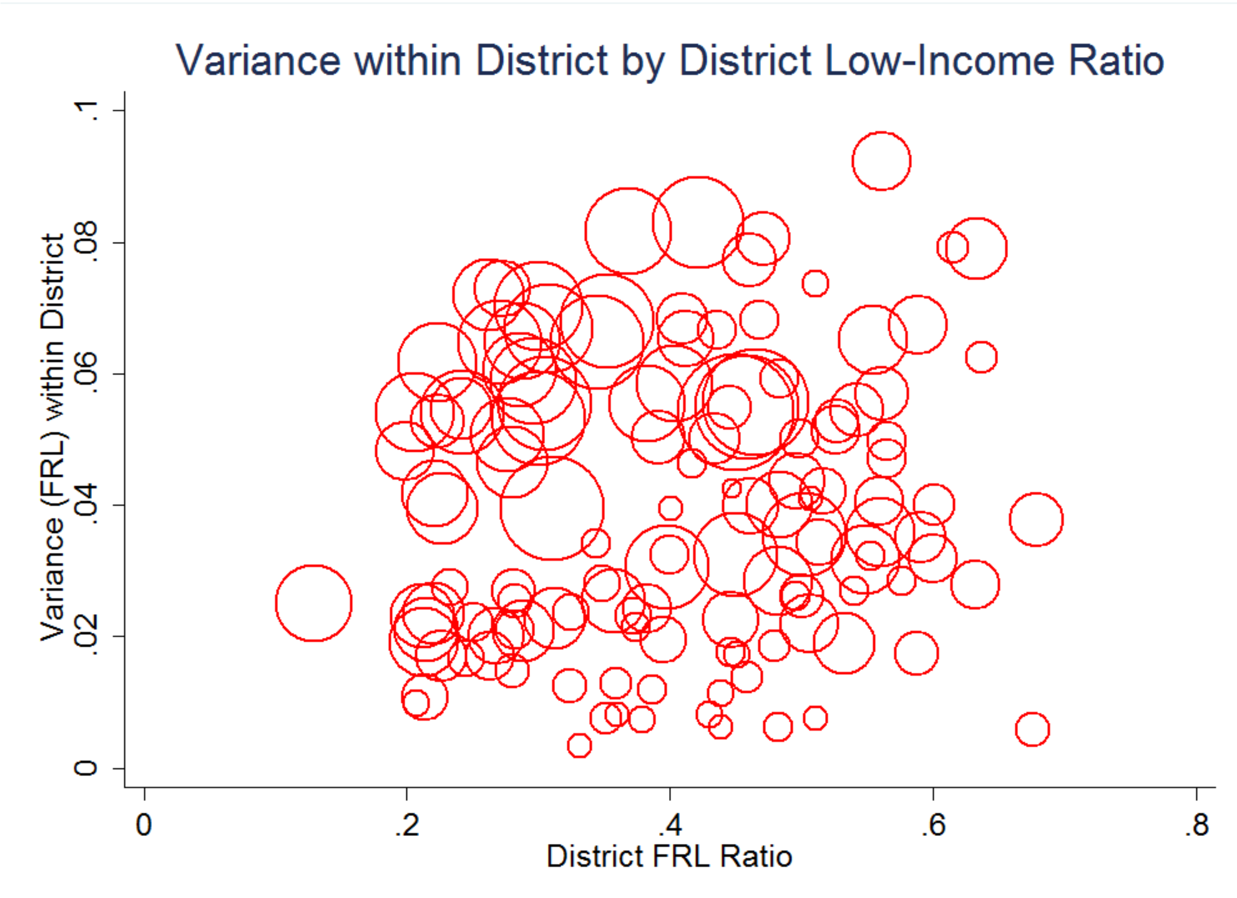
Notes: The figure shows the mean probability of Title I designation by the school's percentile within district with respect to FRL ratio.

Figure 3.3: Percentile within District by FRL Ratio



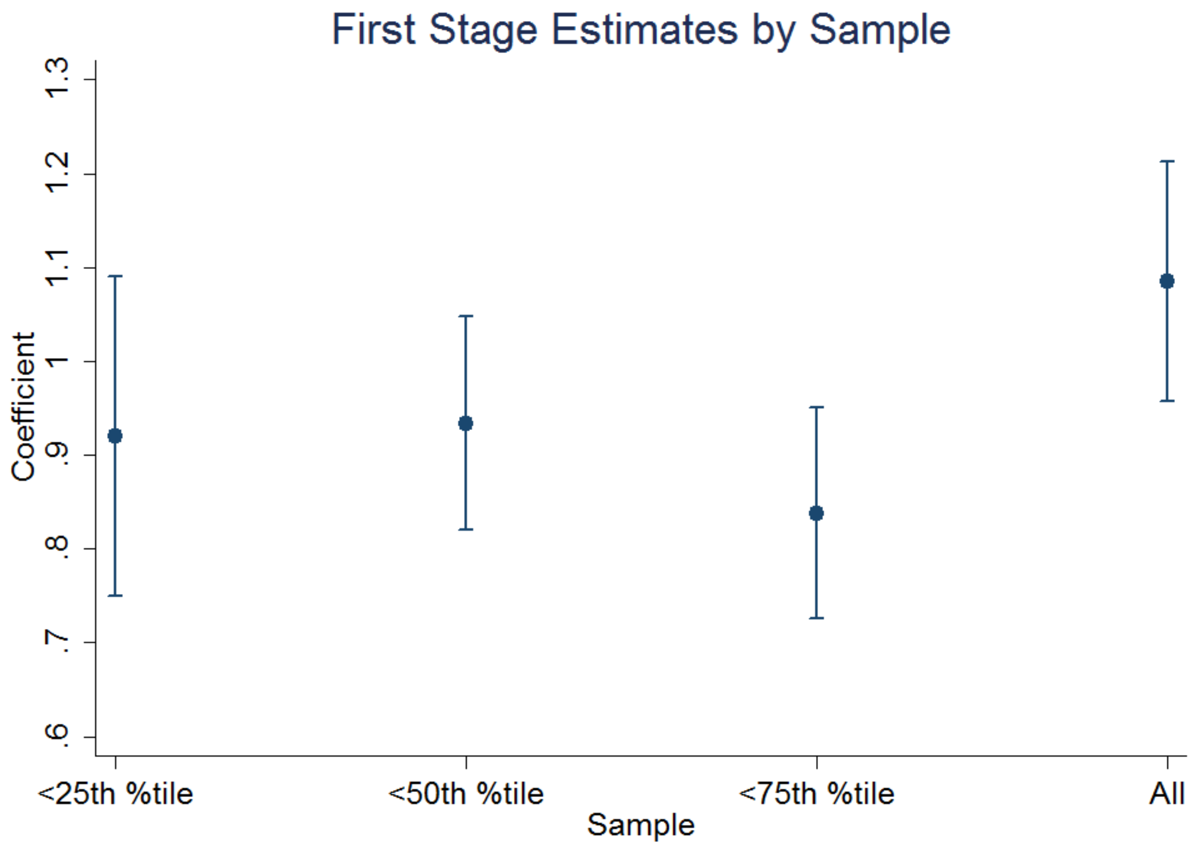
Notes: The figure shows the mean percentile of school FRL ratio within district by the low-income ratio at the school.

Figure 3.4: Within-District FRL Variance by District FRL Mean



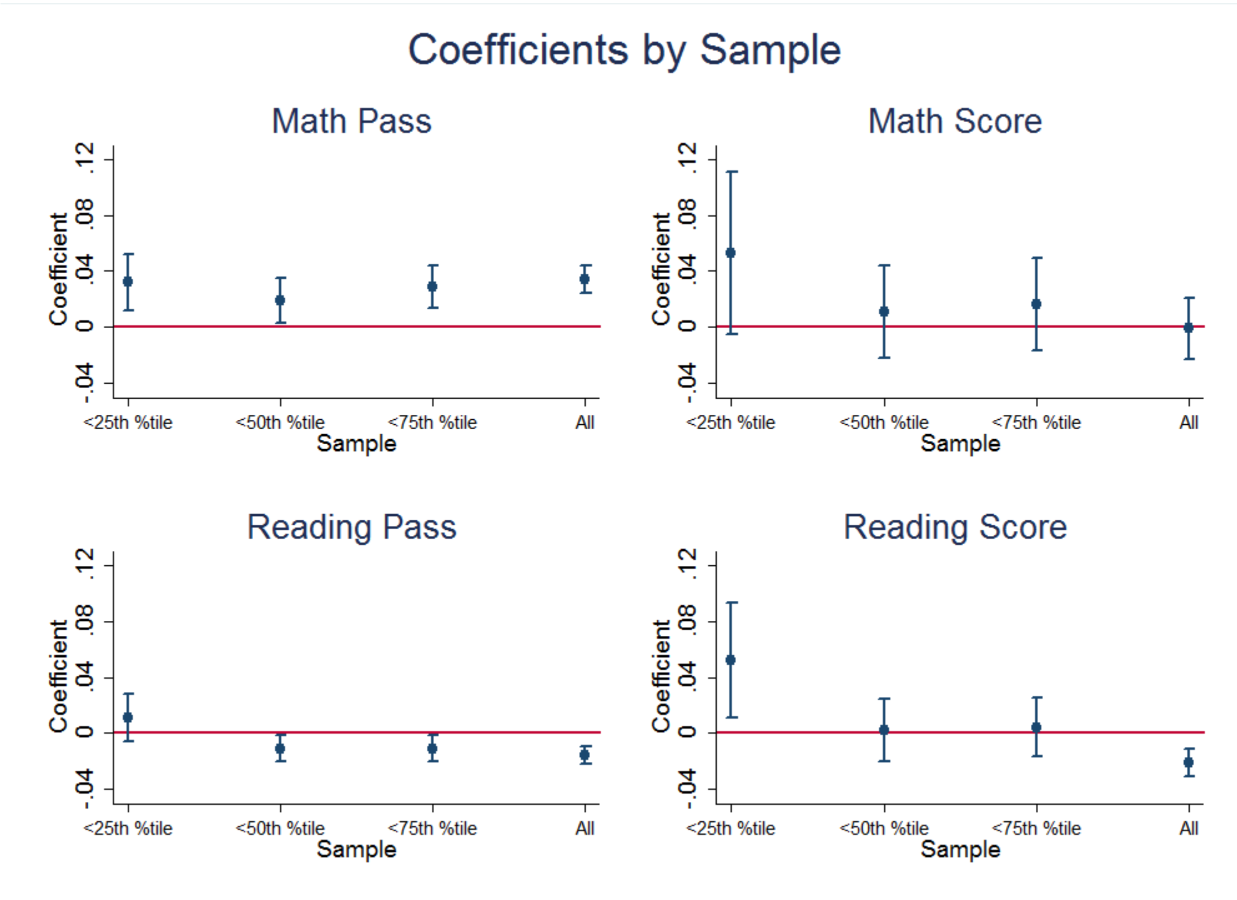
Notes: The figure shows the within-district variance in low-income percentage across schools over the mean low-income percentage within the district, weighted by the number of students in the district.

Figure 3.5: First Stage Estimates by Sample



Notes: The figure shows the first stage estimate and 95% confidence interval across four different samples.

Figure 3.6: Main Estimates by Sample



Notes: The figure shows the principal results for the four outcomes and corresponding 95% confidence interval across four different samples.

Appendices

Appendix A

Targets of Opportunity: The Role of School-Specific Targeted Incentives on Student Achievement

Table A1: AYP Passing Standards by Year

Year	Math	Reading
2003	33	47
2004	33	47
2005	42	53
2006	42	53
2007	50	60
2008	50	60
2009	58	67
2010	67	73
2011	75	80

Notes: For each year and subject, the number indicated is the percentage of students that must pass the exam in order for the school to have met its standard. The passing score for each student varies by year, grade, and subject.

Table A2: NCLB AYP Sanctions

Consecutive years missed AYP in same indicator	Sanction
1	None
2	District must offer transfer option to other schools in the same district.
3	Students are eligible for supplemental educational services, paid for by the district.
4	School undergoes corrective action, which may include replacement of staff, new curriculum, a restructuring of the school, an extended school day or year, or appointment of outside experts.
5	District must prepare a restructuring plan for the school.
6	District must implement the restructuring plan. Must include alternative governance arrangements.

Table A3: Sample Restrictions

Restriction	N Campuses	N Race x Campus Groups	N Students
Have Math and Reading Scores	52,474	146,037	16,820,061
& Not AYP Exempt	50,689	141,585	16,446,768
& Grade $\leq$ 8	42,854	120,904	13,952,455
& Count $\geq$ 150	36,349	103,867	13,370,364
& Within Bandwidth of 5	915	4,821	239,464

Notes: Beginning sample includes students in grades 3-8 and 10 from public schools in Texas from 2004-2011 for whom there is enrollment and testing data.

Table A4: Results by Sample

	Math		Reading	
	Pass	Score	Pass	Score
All Grades, All Student Counts, Bandwidth=5				
Subgroup Qualifies	1.25***	0.033***	1.08***	0.032***
	(0.41)	(0.009)	(0.34)	(0.008)
	385,056	385,056	385,056	385,056
≤8th Grade, All Student Counts, Bandwidth=5				
Subgroup Qualifies	1.23***	0.036***	1.04***	0.034***
	(0.45)	(0.011)	(0.32)	(0.009)
	292,827	292,827	292,827	292,827
≤8th Grade, Student Count ≥150, Bandwidth=5				
Subgroup Qualifies	1.17**	0.032***	1.39***	0.034***
	(0.50)	(0.012)	(0.36)	(0.010)
	239,464	239,464	239,464	239,464
Student demographics	✓	✓	✓	✓
School demographics	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Notes: Each result is from a separate regression with a linear function of the count on each side of the threshold, allowing the slope to vary on either side. Regressions are done using different samples of students described in the left-most column, and include gender, race, FRL status, special education status, and LEP status at the student level, as well as gender, FRL, special education, LEP, and student count at the school level, as indicated. All regressions include school-by-race fixed effects.

*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table A5: Tests on Observables

Panel A: Demographics						
	Male	FRL	Special Ed	ESL	Gifted	At Risk
Student-level						
Subgroup Qualifies	0.46	0.93	-0.57	-0.99	-0.38	0.22
	(0.47)	(1.58)	(0.42)	(1.23)	(0.66)	(1.10)
Mean	50.31	57.27	8.22	9.61	8.24	38.16
N	239,464	239,464	239,464	239,464	239,464	239,464
# Clusters	4,821	4,821	4,821	4,821	4,821	4,821
School-level						
Subgroup Qualifies	0.02	-0.52	-0.20	0.23	-0.65	-0.56
	(0.19)	(1.73)	(0.26)	(0.47)	(0.54)	(1.16)
Mean	51.22	52.53	11.18	6.34	9.27	39.41
N	239,464	239,464	239,464	239,464	239,464	239,464
# Clusters	4,821	4,821	4,821	4,821	4,821	4,821

Panel B: Prior Year Scores				
	Math		Reading	
	Pass	Score	Pass	Score
Subgroup Qualifies	-0.05	0.016	0.23	0.01
	(0.56)	(0.014)	(0.43)	(0.013)
Mean	81.91	-0.801	88.51	-0.018
N	167,253	167,253	167,253	167,253
Clusters	4,795	4,795	4,795	4,795

Notes: Each result is from a separate regression with a linear function of the count on each side of the threshold, allowing the slope to vary on either side. The results presented are from regressions without control variables apart from the linear spline.
*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table A6: Placebo Threshold Effects

	Math		Reading	
	Pass	Score	Pass	Score
Threshold at 35	-0.45	-0.015	-0.32	-0.004
	(0.49)	(0.012)	(0.37)	(0.011)
N	211,444	211,444	211,444	211,444
Threshold at 40	-0.09	-0.004	-0.47	0.001
	(0.51)	(0.011)	(0.36)	(0.010)
N	220,347	220,347	220,347	220,347
Threshold at 45	0.35	0.018	0.49	0.023**
	(0.46)	(0.011)	(0.33)	(0.009)
N	233,837	233,837	233,837	233,837
Threshold at 50	1.17**	0.032***	1.34***	0.034***
	(0.50)	(0.012)	(0.36)	(0.010)
N	239,464	239,464	239,464	239,464
Threshold at 55	0.12	-0.001	0.09	-0.007
	(0.51)	(0.012)	(0.36)	(0.010)
N	242,995	242,995	242,995	242,995
Threshold at 60	0.21	0.017	-0.10	0.004
	(0.53)	(0.013)	(0.37)	(0.010)
N	243,755	243,755	243,755	243,755
Threshold at 65	-0.01	0.010	0.579	0.023**
	(0.53)	(0.013)	(0.37)	(0.010)
N	247,580	247,580	247,580	247,580

Notes: Each result is from a separate regression with a linear function of the count on each side of the threshold indicated, allowing the slope to vary on either side. All regressions include year, grade, and school-by-race fixed effects, gender, race, FRL status, special education status, and LEP status at the student level, as well as gender, FRL, special education, LEP, and student count at the school level.
*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table A7: Results by Student Racial Group Performance Prior Year

	Math				Reading			
	Pass	Pass	Score	Score	Pass	Pass	Score	Score
Subgroup Qualifies	1.08**		0.030**		1.29***		0.034***	
	(0.52)		(0.012)		(0.38)		(0.010)	
Subgroup Qualifies * 1st Quartile		-0.13		0.020		1.99*		0.034
		(1.26)		(0.023)		(1.13)		(0.025)
Subgroup Qualifies * 2nd Quartile		0.13		0.006		1.01		0.039**
		(0.88)		(0.018)		(0.71)		(0.017)
Subgroup Qualifies * 3rd Quartile		1.29**		0.003		0.85*		0.019
		(0.59)		(0.016)		(0.46)		(0.014)
Subgroup Qualifies * 4th Quartile		1.96***		0.055***		1.38***		0.040***
		(0.64)		(0.017)		(0.48)		(0.015)
Student demographics	✓	✓	✓	✓	✓	✓	✓	✓
School demographics	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
p-value for equality of coefficients		0.301		0.099		0.701		0.608
Mean	81.94	81.94	-0.093	-0.093	88.87	88.87	-0.027	-0.027
N	232,109	232,109	232,109	232,109	232,109	232,109	232,109	232,109
Clusters	4,657	4,657	4,657	4,657	4,657	4,657	4,657	4,657

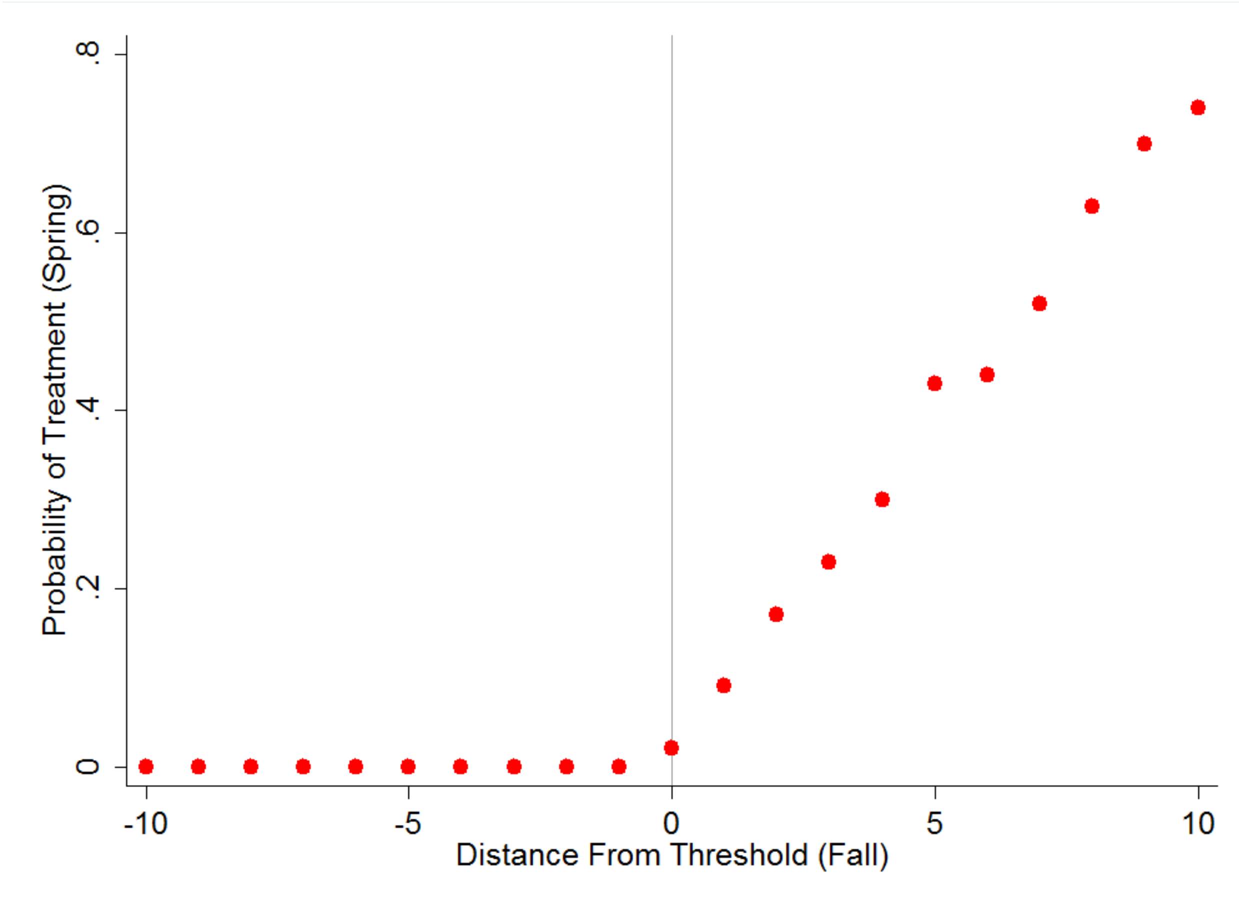
Notes: Each column is from a separate regression with a linear function of the count on each side of the threshold, allowing the slope to vary on either side. All regressions include year, grade, and school-by-race fixed effects, gender, race, FRL status, special education status, and LEP status at the student level, as well as gender, FRL, special education, LEP, and student count at the school level. For columns with quartiles, each independent variable is interacted with each of the quartile dummies in one regression, allowing for the coefficients to vary by quartile.
*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table A8: Results by Performance of Other Students from Prior Year

	Math				Reading			
	Pass	Pass	Score	Score	Pass	Pass	Score	Score
Subgroup Qualifies	1.15**		0.033***		1.30***		0.034***	
	(0.50)		(0.012)		(0.36)		(0.010)	
Subgroup Qualifies * 1st Quartile		-0.022		0.020		1.85*		0.029
		(1.25)		(0.023)		(1.12)		(0.025)
Subgroup Qualifies * 2nd Quartile		0.21		0.008		1.25*		0.040**
		(0.86)		(0.19)		(0.70)		(0.017)
Subgroup Qualifies * 3rd Quartile		1.44**		0.005		0.92**		0.019
		(0.58)		(0.016)		(0.45)		(0.014)
Subgroup Qualifies * 4th Quartile		1.95***		0.057***		1.32***		0.040***
		(0.62)		(0.017)		(0.46)		(0.014)
Student demographics	✓	✓	✓	✓	✓	✓	✓	✓
School demographics	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
p-value for equality of coefficients		0.322		0.092		0.819		0.607
Mean	82.16	82.16	-0.084	-0.084	89.00	89.00	-0.020	-0.020
N	239,375	239,375	239,375	239,375	239,375	239,375	239,375	239,375
Clusters	4,819	4,819	4,819	4,819	4,819	4,819	4,819	4,819

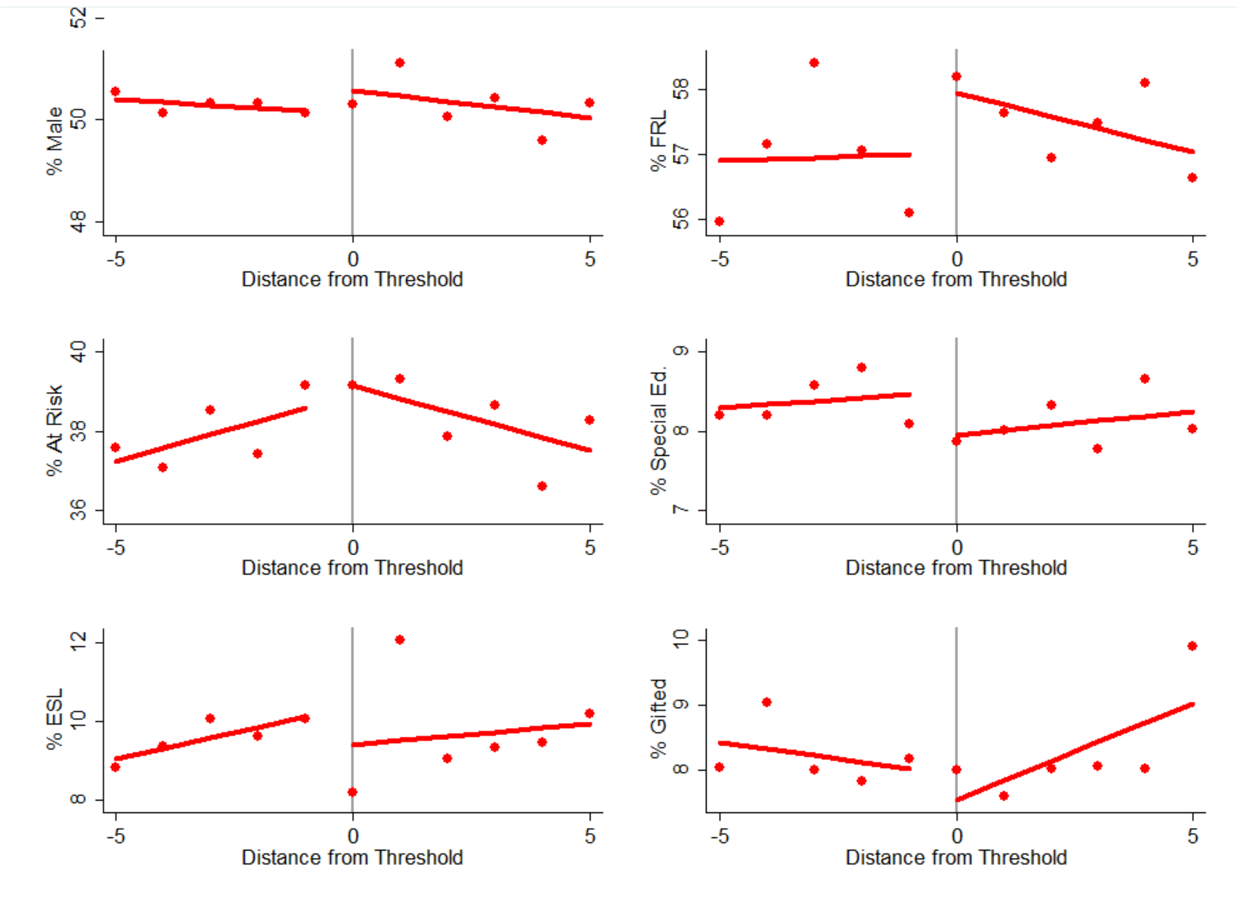
Notes: Each column is from a separate regression with a linear function of the count on each side of the threshold, allowing the slope to vary on either side. All regressions include year, grade, and school-by-race fixed effects, gender, race, FRL status, special education status, and LEP status at the student level, as well as gender, FRL, special education, LEP, and student count at the school level. For columns with quartiles, each independent variable is interacted with each of the quartile dummies in one regression, allowing for the coefficients to vary by quartile.
*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Figure A1: Spring Treatment as a Function of Fall Counts



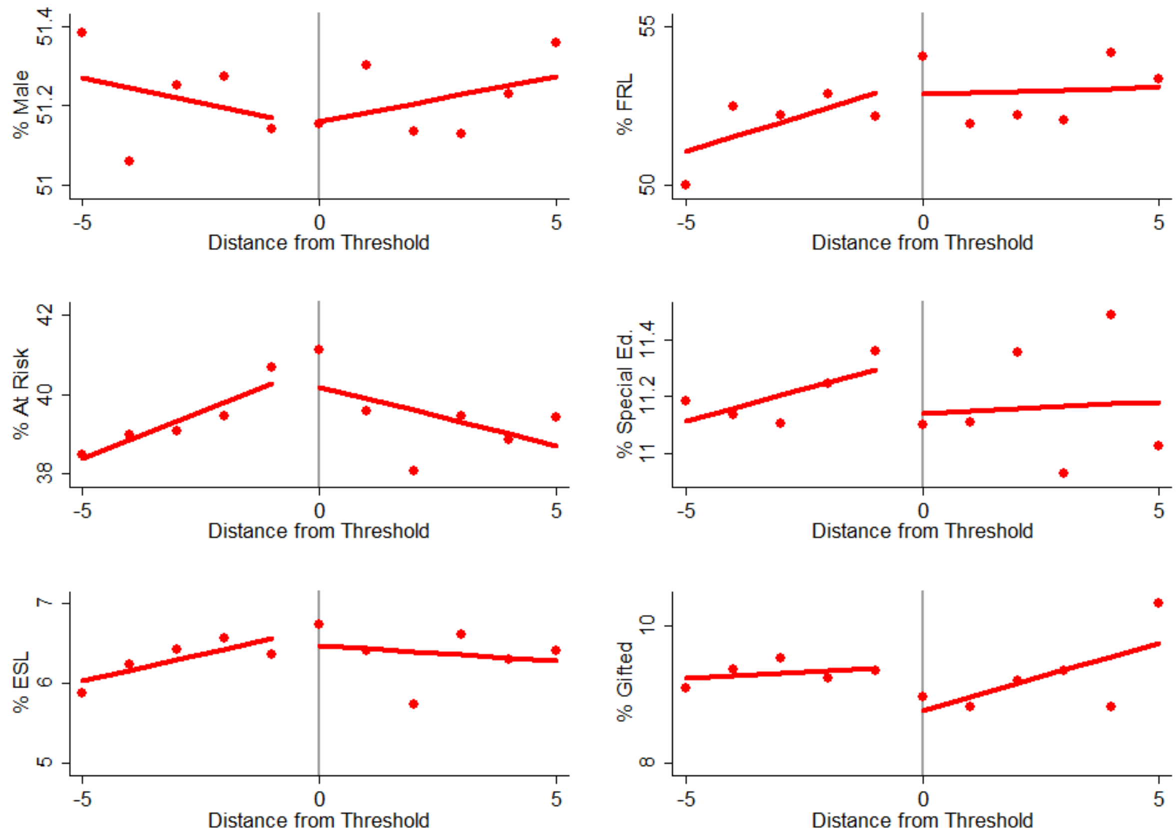
Notes: The figure shows the probability of official spring treatment as a function of the fall count of students by subgroup. Only students enrolled in the fall qualify as a member of the subgroup in the spring.

Figure A2: Individual Covariates as a Function of Subgroup Count



Notes: Each panel shows the mean value of the given covariate by the distance from the threshold, as well as the predicted values generated by the default regression specification described in the text.

Figure A3: School Covariates as a Function of Subgroup Count



Notes: Each panel shows the mean value of the given covariate by the distance from the threshold, as well as the predicted values generated by the default regression specification described in the text.

Bibliography

Thomas Ahn and Jacob Vigdor. The impact of no child left behind's accountability sanctions on school performance: Regression discontinuity evidence from north carolina. Working Paper 20511, National Bureau of Economic Research, September 2014. URL http://www.nber.org/papers/w20511.

Patricia M Anderson, Kristin F Butcher, and Diane Whitmore Schanzenbach. Adequate (or adipose?) yearly progress: Assessing the effect of" no child left behind" on children's obesity. Technical report, National Bureau of Economic Research, 2011.

Marigee Bacolod, John DiNardo, and Mireille Jacobson. Beyond incentives: Do schools use accountability rewards productively? Technical report, National Bureau of Economic Research, 2009.

Martha J Bailey and Susan M Dynarski. Gains and gaps: Changing inequality in us college entry and completion. Technical report, National Bureau of Economic Research, 2011.

Dale Ballou and Matthew G Springer. Achievement trade-offs and no child left behind.

G.S. Becker. *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago Press, 1994.

Trish Boland, Alyssa Pearson, Nazanin Mohajeri-Nelson, and Mariah Aldinger. Title i, part a dissemination report.

Geoffrey D Borman and Jerome V DAgostino. I and student achievement: a meta-analysis of federal evaluation results. *Educational Evaluation and Policy Analysis*, 18(4):309–326, 1996.

D. Card and A.A. Payne. School finance reform, the distribution of school spending, and the distribution of student test scores. *Journal of Public Economics*, 83(1):49–82, 2002.

Martin Carnoy and Susanna Loeb. Does external accountability affect student outcomes? a cross-state analysis. *Educational evaluation and policy analysis*, 24(4):305–331, 2002.

J.S. Coleman, EQ Campbell, CJ Hobson, J. McPartland, AM Mood, FD Weinfeld, and RL York. *Equality of educational opportunity [summary report]*. US Department of Health, Education, and Welfare, Office of Education, 1966.

Steven G Craig, Scott A Imberman, and Adam Perdue. Does it pay to get an a? school resource allocations in response to accountability ratings. *Journal of urban Economics*, 73(1):30–42, 2013.

Thomas S Dee and Brian Jacob. The impact of no child left behind on student achievement. *Journal of Policy Analysis and Management*, 30(3):418–446, 2011.

David N Figlio and Cecilia Elena Rouse. Do accountability and voucher threats improve low-performing schools? *Journal of Public Economics*, 90(1):239–255, 2006.

David N Figlio, Cecelia E Rouse, and Analia Schlosser. Leaving no child behind: Two paths to school accountability. *The Urban Institute*, 2009.

N. Gordon. Do federal grants boost school spending? evidence from title i. *Journal of Public Economics*, 88(9):1771–1792, 2004.

J. Hahn, P. Todd, and W. Van der Klaauw. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209, 2008.

Eric A Hanushek. Assessing the effects of school resources on student performance: An update. *Educational evaluation and policy analysis*, 19(2):141–164, 1997.

Eric A Hanushek. Conclusions and controversies about the effectiveness of school resources. *Economic Policy Review*, 4(1), 1998.

Eric A Hanushek and Margaret E Raymond. Does school accountability lead to improved student performance? *Journal of policy analysis and management*, 24(2):297–327, 2005.

Eric A. Hanushek and Steven G. Rivkin. Chapter 18 teacher quality. volume 2 of *Handbook of the Economics of Education*, pages 1051 – 1078. Elsevier, 2006. doi: 10.1016/S1574-0692(06)02018-6. URL http://www.sciencedirect.com/science/article/pii/S1574069206020186.

Justine S Hastings and Jeffrey M Weinstein. No child left behind: Estimating the impact on choices and student outcomes. Technical report, National Bureau of Economic Research, 2007.

Steven Wkan Hemelt. Performance effects of failure to make adequate yearly progress (ayp): Evidence from a regression discontinuity framework. *Economics of Education Review*, 30(4): 702–723, 2011.

F Cadelle Hemphill and Alan Vanneman. Achievement gaps: How hispanic and white students in public schools perform in mathematics and reading on the national assessment of educational progress. statistical analysis report. nces 2011-459. *National Center for Education Statistics*, 2011.

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

Brian Jacob and Jens Ludwig. Improving educational outcomes for poor children. Technical report, National Bureau of Economic Research, 2008.

Thomas J Kane and Douglas O Staiger. The promise and pitfalls of using imprecise school accountability measures. *The Journal of Economic Perspectives*, 16(4):91–114, 2002a.

Thomas J Kane and Douglas O Staiger. Racial subgroup rules in school accountability systems. 2002b.

John M Krieg. Which students are left behind? the racial impacts of the no child left behind act. *Economics of Education Review*, 30(4):654–664, 2011.

E. Leuven, M. Lindahl, H. Oosterbeek, and D. Webbink. The effect of extra funding for disadvantaged pupils on achievement. *The Review of Economics and Statistics*, 89(4):721–736, 2007.

Susanna Loeb and Miguel Socias. Federal contributions to high-income school districts: the use of tax deductions for funding k-12 education. *Economics of Education Review*, 23(1):85–94, 2004.

J.D. Matsudaira, A. Hosek, and E. Walsh. An integrated assessment of the effects of title i on school behavior, resources, and student achievement. *Economics of Education Review*, 31(3): 1–14, 2012.

Derek Neal and Diane Whitmore Schanzenbach. Left behind by design: Proficiency counts and test-based accountability. *The Review of Economics and Statistics*, 92(2):263–283, 2010.

Aviv Nevo and Adam M Rosen. Identification with imperfect instruments. *Review of Economics and Statistics*, 94(3):659–671, 2012.

Erwin Ooghe. The impact of equal educational opportunity funds: A regression discontinuity design. Technical report, Discussion paper series//Forschungsinstitut zur Zukunft der Arbeit, 2011.

Sean F Reardon. The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. *Whither opportunity*, pages 91–116, 2011.

Randall Reback. Teaching to the rating: School accountability and the distribution of student achievement. *Journal of Public Economics*, 92(5):1394–1415, 2008.

Randall Reback, Jonah Rockoff, and Heather L Schwartz. Under pressure: Job security, resource allocation, and productivity in schools under nclb. Technical report, National Bureau of Economic Research, 2011.

Jonah Rockoff and Lesley J Turner. Short-run impacts of accountability on school quality. *American Economic Journal: Economic Policy*, 2(4):119–147, 2010.

Cecilia Elena Rouse, Jane Hannaway, Dan Goldhaber, and David Figlio. Feeling the florida heat? how low-performing schools respond to voucher and accountability pressure. Technical report, National Bureau of Economic Research, 2007.

David P Sims. Strategic responses to school accountability measures: It's all in the timing. *Economics of Education Review*, 27(1):58–68, 2008.

David P Sims. Can failure succeed? using racial subgroup rules to analyze the effect of school accountability failure on student performance. *Economics of Education Review*, 32:262–274, 2013.

Matthew G Springer. The influence of an nclb accountability plan on the distribution of student test score gains. *Economics of Education Review*, 27(5):556–563, 2008.

Gail L Sunderman. Accountability mandates and the implementation of title i schoolwide programs: A comparison of three urban districts. *Educational Administration Quarterly*, 37(4): 503–532, 2001.

W. Van der Klaauw. Breaking the link between poverty and low student achievement: An evaluation of title i. *Journal of Econometrics*, 142(2):731–756, 2008.

Kenneth K Wong and Stephen J Meyer. I schoolwide programs: a synthesis of findings from recent evaluation. *Educational Evaluation and Policy Analysis*, 20(2):115–136, 1998.