To Leave or Not to Leave?
A Regression Discontinuity Analysis of the Impact of Failing the High School Exit Exam

Forthcoming, *Economics of Education Review*

Dongshu Ou

Centre for Economic Performance
London School of Economics

*I thank New Jersey Department of Education for providing the data used in this paper. I am grateful to Dana Berry, Timothy Peters, Rob Akins and Don White for their help in answering my questions on the data as well as New Jersey's state educational assessment. I thank Henry Levin, Francisco Rivera-batiz, Randall Reback, and Miguel Urquiola for valuable discussions and comments. For their helpful suggestions, I also thank Linda Bell, Rocky Citro, Alex Eble, Yinghua He, Dong Wook Jeong, Ayako Kondo, Stephan Litschig, Sandra McNally, Guy Michaels, Steve Pischke, Emma Tominey, Jessica Wolff, anonymous referees, and participants at the Eastern Economic Association Annual Conference 2008, the American Education Finance Association Annual Conference 2008, the Campaign for Equity Student Research Conference, Institut d'Economia de Barcelona's summer school 2008, the Society of Economics and Education Seminar, and the CEP Labour Market Workshop at the London School of Economics. The research is supported by the Campaign for Educational Equity Student Research Grant, the Spencer Research Training Grant, and the Policy and Research Fellowship at Columbia University's Teachers College. Earlier drafts of this paper were titled "The Impact of High School Exit Exam on Student Dropouts: A Regression Discontinuity Analysis."

To Leave or Not to Leave? A Regression Discontinuity Analysis of the Impact of Failing the High School Exit Exam

Abstract

The high school exit exam (HSEE) is rapidly becoming a standardized assessment procedure for educational accountability in the United States. I use a unique, state-specific dataset to identify the effects of failing the HSEE on the likelihood of dropping out of high school based on a regression discontinuity design. The analysis shows that students who barely failed the exam were more likely to exit than those who barely passed, despite being offered retest opportunities. The discontinuity amounts to a large proportion of the dropout probability of barely failers, particularly for limited-English-proficiency, racial-minority, and low-income students, suggesting that the potential benefit of raising educational standards might come at the cost of increasing inequalities in the educational system.

JEL code: I21, I28, J24
Keywords: high school exit exam, student dropout, regression discontinuity

Dongshu Ou

Centre for Economic Performance
London School of Economics
Houghton Street,
London WC2A 2AE
Email: d.ou@lse.ac.uk.
Phone: +44-20-7955-6646
Fax: + 44-20-7955-7595

1. Introduction

Many states in the US require high school students to pass an exit exam as a graduation requirement. The exit exam is rapidly becoming a standardized assessment procedure for educational accountability. In 2006, high school students in 22 states were required to pass an exit exam to obtain their high school diplomas. Most states are phasing in exit exams and some have implemented more rigorous tests than the more traditional minimum-competency exams, adopting standards-based and end-of-course exams. These High School Exit Exams (HSEE) are more prevalent in states with higher percentages of economically disadvantaged and minority students.

HSEEs were introduced to verify that graduating students in high school had mastered the core curriculum (CEP, 2004). However, it has been suggested that such high-stakes exams can prompt minority and low-achieving students to leave high school early (CEP, 2006b; Garcia & Gopal, 2003; Jacob, 2001; Warren, Jenkins, & Kulick, 2006). Unfortunately, despite the vast and rapid adoption of HSEEs, there is very little causal research on their benefits, including whether exit exams effectively raise students' academic skills. Additionally, most of the studies that have looked HSEEs show mixed evidence on the association between state HSEE policy and state-level dropout and graduation rates. Martorell (2004) first explored the causal relationship between failing the HSEE and various student-level academic outcomes. He concluded that the HSEE in Texas in the 1990s did not discourage test failers to drop out *early*, but that failing the exam reduced post-secondary attainment.

This paper presents new empirical evidence on whether failing the HSEE increases the chance of leaving high school prior to high school completion. More importantly, I investigate the potentially different impacts of failing the HSEE on limited-English-proficiency (LEP),

racial-minority, and low-income students. I take advantage of a new longitudinal dataset from New Jersey that captures the most recent changes (i.e., higher standards) in the exit exam. Following Martorell (2004), I exploit the discontinuity in the likelihood of high school exit around the HSEE cutoff score and compare the exit probability of the students who barely pass or barely fail the test. Barely failers provide the counterfactual outcome for barely passers because the treatment status is "as good as randomly assigned" in the neighborhood of the treatment threshold (van der Klaauw, 2002). This paper also intends to quantify the magnitude to which different testing subjects affect student dropout, an issue that previous studies have not addressed.

The results demonstrate that students who barely failed the initial HSEE are more likely to exit high school early than students who barely passed. Furthermore, the difference in dropout probability among those who barely fail and those who barely pass is larger for racial- minority students, economically disadvantaged students, and for math tests relative to English tests. The impact of failing the first English test on dropout probability is also large for LEP students. The results are robust when using different functional forms to predict as well as to test discontinuity in a small neighborhood around the cutoff score. Although the average passing rates in both math and English increased year by year, the regression discontinuity estimates remained consistent, which might indicate that the impact of barely failing/passing the test was similar for students with different abilities.

My analyses investigate the propensity to drop out for students around the HSEE's pass/fail cutoff. The difference in dropout propensity between the two groups (a "random shock of failing the test") could be due to the psychological effects of failing[1], misunderstanding or not being

[1] Martorell (2004) calls this a "discouragement effect." It can be interpreted as a "net discouragement effect," as the discontinuity also exists when the students who barely pass boost their confidence and therefore increases the

informed of retest opportunities, or withdrawal because of the perceived high cost of studying for the retest. Given that there is no causal evidence demonstrating the benefits of raising educational standards with an HSEE, the present findings on the potential cost of high-stakes testing is even more of a reason to reanalyze current testing policy. One might be particularly worried if this cost is disproportionately incurred by disadvantaged students. Because remedial resources might be too expensive for some states or school districts (Jacob, 2001), schools and policymakers should consider providing these at-risk students with counseling services (McGray, 2006), inform them of retest opportunities, and reduce the stigma of failing the HSEE, especially for low-income, LEP, and racial-minority students (Cornell, Krosnick & Chang, 2006; Wilson, 2008). Despite some limitations, the present findings send a signal that more empirical (and preferably experimental) evidence is needed to test the validity, efficacy, and costs versus benefits of HSEEs.

This paper is organized as follows. Section 2 provides a background for the HSEE in the US, followed by a brief review of the related literature. Section 3 introduces the exit exam in New Jersey. Section 4 describes the data, and Section 5 discusses the empirical strategy used and its validity. Section 6 reports the empirical results, and Section 7 concludes the paper.

2. Overview of the HSEE in the US

HSEEs are "state-mandated tests that high school students must pass to receive a high school diploma" (CEP, 2006a). An exit exam is designed to test all subject matter learned during high school in a comprehensive fashion. The introduction of the HSEE as a graduation

---

probability of staying in high school. In principal, we are interested in the potential outcome differences between the barely passers and barely failers. See section 5 for a more detailed discussion. The "discouragement effect" resulting from failing the HSEE is also discussed by Cornell et al. (2006). Their qualitative study documents the potential psychological effects of failing the HSEE for students who were informed incorrectly that they had failed the exam. The authors argue that falsely identified failers might be discouraged from completing high school and lead to higher rates of early exit.

requirement can be traced to the 1980s,[2] when the report *A Nation at Risk* (National Commission on Excellence in Education, 1983) called for higher standards and expectations, including the use of standardized tests to improve the academic underachievement of U.S. students compared to other industrialized countries.

HSEEs can be categorized as one of three types: minimum-competency exams, standards-based exams, and end-of-course exams. A minimum-competency exam is a test that focuses on basic skills below the high school level. A standards-based exam is aligned with state standards, and an end-of-course exam is taken immediately after students complete the coursework in a given subject. States with exit exams allow students who do not pass the exam on the first try to retake it before the end of the 12th grade or even after. Strictly speaking, students are not able to graduate from high school if they fulfill other state or local requirements (such as coursework) but score lower than the required proficiency level on the HSEE.

The federal No Child Left Behind Act (NCLB) of 2001 set out to reform public education by introducing accountability measures, which required that each state administer annual, standards-based assessments to students in Grades 3 through 8 and at least once in high school. Most state exit exams are thus used to meet the NCLB high school assessment requirements (CEP, 2006b, 2007). Because the NCLB reinforced and expanded the requirements for high school students to demonstrate a certain level of competency in order to graduate, in recent years HSEEs have rapidly included more requirements and become more difficult. In addition, HSEEs have expanded quickly to different regions. Currently there are 20 states using either a comprehensive exam or an end-of-course exam as a high school graduation requirement (CEP, 2007).

[2] See, for example, Warren et al. (2006) and Dee and Jacob (in press) for a detailed discussion on the educational reform movement.

2.1 Previous Literature Related to HSEEs and Student Dropout

As more and more states have made the HSEE a graduation requirement and education stakeholders have expressed growing concern about the potential impacts on students, there has been an increasing interest in studying the effects of HSEEs at both the national and state levels. A large literature has focused on dropout rates, not only because the rates in HSEE states tend to be high but also because the cost of dropping out is substantial, both to the student and to society. For example, Belfield and Levin (2007) estimate that the social gains for an additional high school graduate will be up to $392,000 in present value for a person who was aged 20 in 2007.

Studies that have examined the effect of introducing an exit-exam policy have reached different conclusions. Some studies have found no evidence of a relationship between exit exams and dropping out (Greene & Winters, 2004; Muller, 1998; Muller & Schiller, 2000; Warren & Edwards, 2005). For instance, Greene and Winters (2004) concluded that the adoption of an HSEE had no effect on high school graduation rates for the classes of 1991 through 2001. They claim that it is possible that the increased dropout rate of those who failed the exit exam offset the increased graduation rate of those who were motivated by the exam. However, other researchers have found that exit exams encourage more students to drop out (Griffin & Heidorn, 1996; Jacob, 2001; Warren et al. 2006).[3] A more recent study by Dee and Jacob (in press) also reports that Minnesota's exit exam increased the dropout rates in urban, high-poverty, and high-minority school districts. Dee and Jacob (in press) further found that students in states with exit exams are more likely to drop out of high school than students of similar backgrounds in states with no exams. In their study, the effects were strongest among African-American males.

[3] Griffin and Heidorn (1996) find an increased likelihood of dropping out for higher GPA students associated with failure on Florida's competency test, but not for minority or low-achieving students. Using National Education Longitudinal Study data, Jacob (2001) found that low-ability students are more likely drop out in states with an exit-exam policy than those without an exam.

Using a regression discontinuity method on a longitudinal dataset, Martorell (2004) studied the impacts of the Texas exit exam (Texas Assessment of Academic Skills) during the 1990s on several educational outcomes, including dropout and graduation rates. Comparing the students who barely passed or barely failed the test, he concluded that failing the exam did not "discourage" students to drop out in early grades but reduced the amount of post-secondary attainment. However, the minute and statistically insignificant impact on student dropout could have been due to the multiple retesting opportunities (see also Greene and Winters, 2004 and CEP, 2004, for related discussions) and to the fact that the 1990s test measured more basic skills than Texas's current version (the Texas Assessment of Knowledge and Skills).

Although multiple academic subjects are tested in existing HSEEs, few studies have quantified the magnitude to which failing different subjects might affect student dropout rates, especially for minority and low-income students. Papay, Murnane, and Willett (2008) focused on the on-time graduation rate for barely failers relative to barely passers of Massachusetts's HSEE in urban schools in 2006. They found some negative effects of failing the initial math test for low-income urban students but no significant effects for failing the initial English test. They explained that the limited school and family resources faced by these students could be a barrier to high school graduation. A recent manuscript by Reardon et al. (2008) looked at the effects of failing California's HSEE on students' subsequent academic achievement, persistence to 12th grade, and graduation rates. They found that the initial exit-exam failure had very few effects on various outcomes measured in four large school districts in California. The authors claimed that it could be due to the multiple retest opportunities and variations of school funding on test preparation and remediation.

The present paper provides specific evidence from New Jersey to add to the pool of findings that will ultimately be useful to policymakers as they consider revisions to HSEE policy. My study compares results from four cohorts in an entire state instead of one cohort (Papay, Murnane & Willet, 2008) or a few districts (Reardon et al., 2008). The analysis also sheds light on how different subgroups including female, LEP, racial-minority, and economically disadvantaged students react differently, compared to their counterparts, to their performance on both the initial test and the retest of today's more rigorous exit exams.

3. The HSEE in New Jersey

New Jersey was one of the first states to adopt a statewide assessment test as a requirement for its high school diploma. Early in the 1981–1982 school year, 9th-grade students needed to pass the Minimum Basic Skills Test to obtain a high school diploma. The test was upgraded to a more challenging assessment, the Grade 9 High School Proficiency Test in 1983. Students who were first-time 9th graders during and after the fall semester of 1991 were required to pass the High School Proficiency Test to graduate. The state began to administer its new standards-based exam, the High School Proficiency Assessment (HSPA), in March 2002 for the first time to 11th graders. The class of 2003 was the first cohort to graduate under this new HSEE requirement.

The HSPA was intended to gauge students' knowledge and skills in the New Jersey Core Curriculum Content Standards and is mandatory for New Jersey high school juniors in public schools. The test contained two sections in 2002 to 2006: Mathematics and Language Arts Literacy (LAL). Both sections utilize multiple-choice as well as open-ended questions. The LAL section also includes essays. The multiple-choice questions are machine-scored by a company hired by the New Jersey Department of Education, and the open-ended questions and essays are scored by two trained independent raters whose scores are combined into a total score. Each

section of the test is scored separately, with scores on each section ranging from 100 to 300. The state defines a score under 200 as "partially proficient," 200 to 249 as "proficient," and 250 and above as "advanced proficient." A procedure called statistical equating is used to make sure that all future HSPA tests are at the same level of difficulty as the March 2002 test (New Jersey Department of Education, 2007). In order to pass the HSPA, a student must obtain a score of at least 200 on all sections.[4]

Eleventh graders are first tested on both math and LAL during the spring semester and, if they fail either section, again during the fall and/or spring semesters in 12th grade. Students only need to retake the test in the subject that they fail and have three opportunities to pass the exam before completing 12th grade. For students who left high school (for any reason including finishing 12th grade without a diploma), they can still take the exam as a "returning student" whenever the test is administered. Another option is to attend adult education schools and retake the test there. Students identified as LEP must meet the same graduation requirements as native, English-speaking students.

4. Data

The data used in the following analysis were obtained from the New Jersey Department of Education and include records of test scores on the HSPA for all students who enrolled in a high school in New Jersey from 2002 to 2006. In addition to math and LAL test scores and school enrollment information, the dataset contains information on the student's school, grade, gender, age, race/ethnicity, whether s/he is economically disadvantaged, special education status, LEP status, an indicator for less than one year spent in the school of testing, Individual Education

[4] Students who have met all the graduation requirements except passing the desired level (cutoff point) of the HSPA can undergo a Special Review Assessment (SRA). The SRA is aligned to the HSPA test specifications and is usually taken after school-district personnel receive the specific information about each student's proficiency results.

Program exemption, and an indicator of socioeconomic status of the citizens in the student's district.

My analysis is restricted to public-school, general-education students who initially took the HSPA in 11th grade. I traced these students to the end of 12th grade and constructed a longitudinal dataset with four cohorts. I excluded students with missing or inconsistent demographic information, migrant students, and students who had been in the school for less than one year when the test was administered. The final sample consisted of 299,948 observations for all cohorts.[5] A detailed discussion on how I constructed the data and selected the sample is provided in the Appendix.

Table 1 reports the descriptive statistics for the full sample and the retest sample. The test scores were adjusted by subtracting the cutoff score, 200, from each student's score. In general, Black, Hispanic, and economically disadvantaged students each comprised about 13% of the full sample.[6] The initial passing rate for the math test was 76%, which is lower than the rate for the LAL test (87%). Female and White students performed better and had a higher initial passing rate than Black, Hispanic, and economically disadvantaged students. For instance, on average, Black students received a math score under the cutoff (-9.5 with a standard deviation of 47.3), whereas White students scored 28.15 points above the cutoff (with a standard deviation of 37.38). Not surprisingly, the average exit probability after the initial test was higher for Black,

[5] Special education students with an IEP (Individual Education Program) exemption in any of the three test administrations were not included. The justification for this is that such students can graduate from high school without passing the exams as long as other graduation requirements are fulfilled. There are also cases where some special education students have been exempted from taking the test because of a personal situation. Their HSPA test scores do not indicate a "pass" or "fail" status. This would not give a perfect assignment to the cutoff and would violate the sharp regression discontinuity assumption. See the Appendix for further discussion.

[6] New Jersey has the lowest rate of students eligible for free or reduced-price meals (27%) in the country. My sample has a lower percentage of economically disadvantaged students because the analysis only includes students who took the HSPA for the first time in 11th grade. My sample also has a higher percentage of White and a lower proportion of Hispanic and Black students, whereas the state had 58% White, 18% Black, and 17% Hispanic students in the 2003-04 school year (Mackey, 2006).

Hispanic, and economically disadvantaged students than for White students. Compared to other groups, LEP students had an extremely low passing rate (less than 30% for both math and LAL) and a higher exit rate after failing one or both subjects. Panel C in Table 1 reveals the diverse composition of the students. 77% of students who failed the math test retook it, and 66% who failed the LAL test retook it. The percentages of economically disadvantaged students, racial-minority students, and LEP students were much higher in the retest than the initial test sample. Hispanic students and economically disadvantaged students did not have high passing rates in either subject, and the LEP students in the retest group had the lowest passing rate for both the math and LAL sections.

5. Empirical Strategy

There is a growing literature on educational and social program evaluations using the regression discontinuity approach (Angrist & Lavy, 1999; Chay, McEwan, & Urquiola, 2005; Jacob & Lefgren, 2004; Kane, 2003; van der Klaauw, 2002; Lemieux & Milligan, 2008). I employed a sharp regression discontinuity design (Trochim, 1984) to identify the effects of failing the HSPA on exiting high school. Individuals just above and below the cutoff can be assumed to be very similar because they tend to "have similar average outcomes in the absence of the [treatment] program as well as similar average outcomes when receiving treatment" (van der Klaauw, 2002), and thus we may assume that students close to the pass/fail cutoff score would have had the same dropout rate in the absence of "failing" status.

My main estimation equation is given by

$$Exit_i^t = \gamma_1 * 1(SCORE_{i,s}^t \geq 0) + f(SCORE_{i,s}^t) + \gamma_2 * 1(SCORE_{i,s}^t \geq 0) * f(SCORE_{i,s}^t) + \eta_i^t, \qquad (1)$$

where the $f(SCORE_{i,s}^{t})$ is an unknown smooth function of test score of subject *s* in period *t* for student *i*, and γ_1 is the coefficient of interest. The estimated discontinuity was the marginal effect of passing the HSPA as obtained from the probit regressions. $1(\bullet)$ is an indicator function that is equal to one if the enclosed statement is true, and η is the error term. I examined two situations in this paper. The first was how likely it was that students would exit after failing the initial attempt. The other was the likelihood of dropping out after the first retest.[7]

To reduce the sampling variation, covariates were added. This procedure was also done to test the credibility of the regression discontinuity approach because adding the covariates was not expected to change the estimates of the treatment effect if they were independent of the assignment variable (Imbens and Lemieux, 2008). Controls for t = 0 included student and school baseline characteristics and dummies for different cohorts. For period t = 1, I included an extra dummy variable for whether the student was retesting on only one subject, along with the controls for t = 0.

It is important to distinguish any effects that were caused by observed variables other than the treatment itself. The underlying assumption is that the only source of discontinuity in the probability of dropping out at the cutoff score is failing the exam. Graphically, there should be no jumps in other observed variables at the cutoff based on the test-score function. Appendix Table A-I presents the regression discontinuity estimates on a set of pre-determined characteristics for the students around the cutoff. Most of the results did not show any statistically significant differences between the barely passers and barely failers of the HSPA. A

[7] I did not study the dropout probability after the second retest because the data is not available. Also, I did not consider students who were retained in 11th grade in this study and the number of grade retainer was very few in the sample.

few do show that there was discontinuity around the boundaries; however, the plottings[8] reveal that these factors did not pose a threat to identification.

Furthermore, the correct specification of the smooth function $f(SCORE_{i,s}^{t})$ is the key to identifying the true treatment effect in the regression discontinuity method (Angrist & Lavy, 1999; McCrary & Royer, 2005; Trochim, 1984). To check the robustness, I compared the results from various specifications of the test score $f(SCORE_{i,s}^{t})$, and I also included the interaction terms to allow for any slope changes. The findings were not sensitive to these functional form differences.

Besides testing for alternate specifications for the $f(SCORE_{i,s}^{t})$, robustness checks include data restrictions to those close to the cutoff and changes in the set of control variables in the probit estimation. Given the standard deviations, and to ensure a comparable number of students on either side of the cutoff[9], I chose different bandwidths for the estimates of the math and LAL tests. These different ranges allowed me to compare the regression discontinuity estimates and sufficiently test their robustness. The results were similar when I controlled for observable individual covariates such as race/ethnicity, gender, cohort dummies, and so forth.

To examine the heterogeneity of the impact across the student population, I ran the regression discontinuity estimates for different subpopulations (e.g., failing one or both subjects, gender, race/ethnicity, cohort/year) and tested for statistical significance across groups and over time.

---

[8] Graphs on selected covariates are included in the working paper version (Ou, 2009). The only two variables that seem to discontinue around the cutoff are the dummy variable for gender and the middle SES group. However, my inclusion of these variables did not significantly change the treatment-effect estimates in the discontinuity sample, suggesting that neither of these variables was strongly correlated with the treatment status (van de Klaauw, 2008). Another possible exercise is to see if one finds more or fewer significant results, as in Table A-I, when using placebo cutoffs (e.g., -10, -20, -30, -40). The falsification tests failed just as often for the placebo cutoffs as for the real cutoffs; therefore, it was unlikely that the estimates presented a real problem.

[9] The test-score histograms (Appendix Figure A-I) did not show any sharp increase in the number of students just below or above the cutoff.

The difference between the exit probabilities for barely passers and barely failers is the "random shock of failing the test" that this paper attempts to investigate. It is possible that students who barely pass the HSEE are encouraged to graduate when they might have otherwise considered dropping out. It is also possible that students who barely fail might be discouraged from making the effort to graduate when otherwise graduation might have felt more attainable. In this study, I am interested in looking at whether the apparent effect on barely failers outweighs the apparent effect on barely passers, or vice versa. I also intend to discuss other nonpsychological reasons to explain students' decision to drop out, for example, perceived cost of staying in school due to the requirement of retesting for the failing subject(s), not being informed of the retest opportunities, updating their own information on their abilities, and so forth. Note that regression discontinuity provided a local estimate for the subgroup of individuals around the cutoff point. This indicates that any effects estimated in the model only applied to the students who were around the threshold. Because a retest for the failing subject(s) is required by the state as long as the student is enrolled, the retest sample includes students who failed any particular subjects initially but did not exit afterwards.

6. Results

To illustrate the results, Figure I shows the relationship between the math/LAL scores (relative to the cutoff) and the actual or estimated means of exiting high school using the pooled data of four cohorts. We can see that the probability of exiting from high school jumps up at the cutoff score for students who barely failed the initial test and the retest. The patterns are clearer in the graphs of the math test.

Even though the discontinuity of the exiting propensity between barely passers and barely failers was bigger for the math test, it is interesting to note that the average exit probability of

barely passers for the LAL test appears to have been higher than the exit probability of the barely failers for the Math tests. It is possible that students who passed at the margin were better students or more academically able than those who passed the LAL test at the margin. The barely passers in math might thus have felt a stronger pull to stay in high school because their likelihood of fulfilling other graduation requirements and going to college was relatively high.

In general, my regression discontinuity estimates are comparable to and consistent with the discontinuities illustrated in the graphs. My results were robust to various functional forms[10] of the test scores, to different bandwidths, and to the inclusion of a set of baseline characteristics of the students. All of the standard errors are clustered at the score level (Lee & Card, 2008). Accordingly, unless otherwise stated, the numbers used in the rest of the paper will be estimates based on the cubic form of $f(SCORE^0_{i,s})$ or the quadratic form of $f(SCORE^1_{i,s})$, with the slope of these functions varying on each side of the cutoff. For simplicity, only the estimates with covariates are discussed below.

6.1. Failing the Initial Test and Dropout

The statistically significant regression discontinuity estimates in Table 2 confirm the link between failing the HSEE and the increased probability of dropping out in early grades. The discontinuities based on the math score were larger than those based on the LAL scores. For example, the regression discontinuity estimates for the math score (-.011 with a standard error of .001) were twice as high as the regression discontinuity estimates for the LAL score (-.005 with a standard error of .002). This pattern held when looking at the observations in a smaller neighborhood closer to the cutoff score (Columns 5, 6, 7 in Table 2). We noticed that the average

[10] I have tested various specifications including linear, quadratic, cubic and fourth polynomial functional form of the math or LAL score, with and without controls of student baseline characteristics. For the ease of reporting, I do not include all the specifications in the tables. Results were robust when including students who had been in the school for less than a year.

passing rate in math was lower than the LAL rate for the HSPA exam (76% vs. 87%, as shown in Table 1). It is plausible that the HSPA math standards were set too high or that the students were poorly prepared for the math test. If either case is true today, students might become more discouraged when failing a more difficult test, or the perceived cost of preparing for (and passing) the retest might be higher when there is inadequate educational support for building up their math skills.

The cells labeled "exit after failure" also show the actual average exit rate for test failers within each bandwidth. For instance, the average exit rate for student math scores within the range of (-30,0) was 4.8% with a standard deviation of .214. Hence, the -.012 estimated discontinuity in students' exit probability was equal to a 33.3%[11] increase in the mean exit rate of barely failers without the random shock, accounting for 5.6% of the standard deviation of the raw exit probability of the barely failers. Similarly, the raw probability of dropping out after failing the LAL test within a bandwidth of (-25,25) was 6.1% (with a standard deviation of .239). The estimated discontinuity was -.007, which increased the raw dropout probability of the barely failers by 13% and the standard deviation by 2.9%.

Table 2 also presents regression discontinuity estimates individually for each cohort. Except for the results for Cohort 1, the statistically significant coefficients in three cohorts showed consistent evidence for the estimates obtained from the pooled sample. The difference between the regression discontinuity estimates for math and LAL scores was largest in Cohort 4. The estimated probability of exiting for students who barely failed the math exam was 1.7% (standard error .004) higher than the exit probability for those who barely passed the exam in math and 0.4% (standard error .002) higher for those who barely failed the LAL test compared to

[11] Similarly, the percentages of increased exit probability resulting from the shock in columns 6 and 7 were 22.2% and 18.4%, respectively.

those who barely passed it. It is worth noting that students in 2005 (i.e., Cohort 4) were given more time to take the math test, but the number of testing items and difficulty remained the same. If students received as much information on the retest opportunities as previous cohorts, it is possible that the psychological impact of discouragement could rise if students who believed they could achieve a higher score with the extended testing time, in fact, still failed.[12]

On another note, the purpose of the HSEE was to increase students' academic performance in school by encouraging them to develop their cognitive skills and learn more (not just to do better on the test). The fact that more students are indeed passing the first test today (Table 1), coupled with the assumption that HSEE difficulty has remained relatively constant over the years, suggests that students at the pass/fail cutoff today may actually have weaker academic abilities than their counterparts from a few years ago. Because there were no big changes (especially in the LAL test specifications) in the regression discontinuity estimates of failing the test on the exit rate from Cohort 1 to Cohort 4, it is possible that the effects of failing were similar for students with different ability levels.

6.2 Failing the Retest and Dropout

The discontinuities of exiting at the cutoff after the retest are shown in Table 3. The estimated magnitudes were similar to the results for the initial test. For instance, the discontinuity estimate of exiting among math retesters was -0.013, with a standard error of 0.001 for all cohorts. Barely failers of the math retest were more likely to drop out than barely failers of the LAL retest when compared to their barely passing counterparts. Some results based on the LAL retest scores in each cohort were not statistically significant, which may have been due to the reduced sample size.

[12] Although this assumes that students were actually aware of the greater testing time. Maybe for students who had a real chance of failing, the discouragement effect might have been more salient.

It is noteworthy that the retesting sample was different than the initial-test sample. Specifically, the retesting sample excluded students who dropped out after failing the first test. Students who stayed in school for the second test were probably more persistent or motivated to pursue their academic training and thus were probably less likely to drop out even if they failed a second time. These students would have been informed of the retest opportunities as well. Therefore, the regression discontinuity estimates based on data taken from New Jersey could be evidence of the "discouragement effect" found by Martorell (2004). He explained that the effect of failing the very last retest was possibly due to the fact that students were unable to graduate from high school if they failed the test on their last chance to take it. In addition, given the positive average test-score gains from the initial test to the retest for the barely failers around the cutoff in my study[13], it is unlikely that the updated information on their own ability prevented them from staying in school. Therefore, it is plausible that the discouragement effect was large for barely failers of the initial test who put great effort into preparing for the retest but still failed.

At the same time, it might not be surprising that the discontinuity amounted to an even larger proportion of the actual exit probability of the barely failers of the retest. For example, the average exit rate for barely failers of math was 2%, and the regression discontinuity estimates indicated an increase of 1.5 percentage points of the exit rate compared to the barely passers of the math retest.

6.3. Heterogeneous Effects

Tables 4 and 5 illustrate the regression discontinuity estimates for the sub-samples on initial test and retest respectively. Apparently, barely failers in the racial-minority and economically

[13] For example, for math retesters who scored below the cutoff by less than 30 points, the average adjusted score on the first test was -31.22 (with a standard error of 44.48), and their average adjusted score on the retest was -17.34 (*SE* = 8.24). For LAL retesters who scored below the cutoff by less than 30 points, the average adjusted score on the first test was -7.93 (*SE* = 41.21), and their average adjusted score on the retest was -14.57 (*SE* = 8.68).

disadvantaged groups were more likely to drop out around the cutoff level, especially after the initial test. For Hispanic and economically disadvantaged groups, the regression discontinuity estimate was -.013, with a standard error of .002 on the initial math test. This is twice as large as the discontinuity found for the White students (the regression discontinuity estimate was around -.005 with a standard error of .001).

Interestingly, the dropout pattern for the retest was somewhat different. Female and White students who barely failed the retest were comparable to racial-minority and economically disadvantaged students. Given that the average exit rates for female and White students were lower than those for the other subgroups[14], the discontinuities found in these two groups could have amounted to a larger proportion of the actual dropout probabilities. However, one major concern is that racial-minority and economically disadvantaged students dropped out earlier because they were more likely to exit after failing the first test. As previously noted, the sorted retest sample for these relatively disadvantaged students, due to earlier dropouts after the initial test, could have underestimated the treatment effect.

Lastly, my results did not show a statistically significant impact for LEP students who barely failed the math exam, but did show an effect for those who barely failed the LAL exam. The regression discontinuity estimate is -.025 with a standard error of .012, which explained 14% of the raw probability of barely failers on the LAL test. The descriptive statistics in Panel B of Table 1 also provide some support for the claim that there was a potential negative impact of HSEE policy on students whose first language was not English (Garcia & Gopal, 2003). From 2002 to 2006, the average passing rate among LEP students was 28% for the initial math test and 22% for the initial English test. The average math test score was -20.06 (*SD* = 45.05) and the

[14] The average exit rate for female, LEP, White, Hispanic, Black, and low-income students were .022, .040,.020, .032,.027, and .031 respectively.

average LAL score was -33.37 (*SD* = 43.96). Further, 8% of LEP students who failed at least one subject in the HSPA dropped out, which was higher than the average. As Garcia and Gopal describe, the LEP students are very likely to exit because the exam makes them ineligible to participate in the rigorous standard curriculum.

6.4. Effects of Failing at Least One Subject

Are barely failers of one subject more likely to drop out if they fail the other subject too? The estimates in Table 4 were performed on a restricted sample who failed the subject not used in the test-score specification. Most of the results, especially the estimates for the initial test, were not statistically significant. The results on the retest for all cohorts were consistent with the estimated magnitudes in the unrestricted sample. This indicates that the barely failers were as likely as the barely passers to stay in or drop out of high school if they failed both subjects[15]. In other words, students who barely passed one subject and failed the other did not seem to be encouraged to stay in high school. Rather, these students might have felt that barely passing one subject was not giving a strong signal that they were doing well enough compared to their peers. Moreover, failing one subject still meant that they had not met the graduation requirement. The additional time to prepare for the retest may have been perceived as a high cost of staying in high school, which could have increased dropout propensity.

7. Conclusions and Policy Implications

The increasing movement towards school accountability is associated with the increased use of standardized tests. There is controversy over whether HSEEs stimulate students' motivation, enhance learning, or, conversely, prevent some high school students from graduating[16]. There is also concern that the potential gain of high-stakes testing comes at the cost of increasing

[15] Nevertheless, barely failers of either the math or LAL who failed the other subject and exited high school had lower scores on the other subject as well (results not shown but available upon request).
[16] See Ou (2009) for a related line of literature on the benefits of test-based accountability and the cost.

inequality between groups. This paper examines whether such a test can potentially affect a student's decision to complete high school by comparing students who were very close to the pass/fail cutoff and comparing the impacts on different subgroups, including racial minorities and economically disadvantaged students.

Using data from the New Jersey HSPA test, this paper finds statistically significant evidence that students who barely failed the exam, especially the math exam, were more likely to drop out than students who barely passed it. While the raw dropout rate of the students who barely failed the initial test was 5% to 6%, the regression discontinuity estimates indicated a 1 percentage point increase in the exit probability for barely failers compared to barely passers. In the retest the regression discontinuity estimates indicated a 1 to 1.5 percentage point increase of exit probability for barely failers, whose actual exit rate was 2% to 3%. The regression discontinuity estimates indicated a 14% increase in the raw exit probability for LEP students who barely failed the first LAL test. The regression discontinuity estimates in dropout propensity were larger and more significant for Black and Hispanic students as well as economically disadvantaged students.

My results are different from Martorell's (2004), which were based on Texas data. One possible reason for this difference is that the HSEEs examined in the present study are very different in design (both in content and the number of retakes allowed) and are set under a very different school environment than the HSEE in Texas. Additionally, the time periods we studied were different as well. Using recent data from Massachusetts, Papay et al. (2008) also discovered that failing the HSEE math test reduced the graduation rate of economically disadvantaged students. Though another recent study by Reardon and his colleagues (2008) did not find any evidence that failing California's HSEE impacted the graduation rates of low-income and

Hispanic students, their results might have been limited by the district-level data they used, which were not able to capture students' mobility within the state.

One caveat is that the regression discontinuity analysis provides local estimates and does not evaluate the overall effects of the exit exam policy on students' dropout behavior. Nevertheless, the results have broader implications for the US. New Jersey recently launched the "High School Graduation Campaign" as part of a national effort to reduce dropout rates (Hu, 2008). It is crucial for schools, parents, and policymakers to understand the factors affecting the choice to drop out. Students who barely fail might be more likely to drop out than those who barely pass if they try hard but still fail and become discouraged; alternatively, the perceived cost of preparing the retests is high when not many remedial resources are available or schools fail to provide sufficient information on their retake opportunities and alternative ways (Such as New Jersey's Special Review Assessment) to graduate.

It is fair to argue that any binding exit exams will cause some students to drop out, and also that barely passers are more likely to stay in high school than barely failers. This is no surprise given that the former group is on track to graduate whereas the other is not. Policymakers should, however, feel some concern if the exit exam is having a disproportionate effect on the traditionally disadvantaged students. The larger impact of failing the test for LEP, racial-minority, and economically disadvantaged students found in this research suggests that states should consider providing additional attention and academic support to reduce dropout among their ranks.

That said, if the students who barely fail the HSEE are not much different than those who barely pass in terms of academic ability, school counselors and teachers should encourage students to retake the test and reduce the stigma of retesting to offset any "discouragement"

effects (Cornell et al., 2006). Other guidance and remedial courses might reduce students' perceived cost of retaking the test and increase the probability of staying in high school if they fail (Reardon et al., 2008), although the cost of additional resources such as summer programs or after-school tutoring might be too high to afford (Jacob, 2001).

It is important to note that despite the rapid and vast adoption of HSEEs across the US, there is no sophisticated cost-benefit analysis or evaluation of the policy to provide a solid background for the reforms that are currently being proposed in various states. Without more data and further evidence, we cannot determine whether imposing a higher standard of testing increases or decreases students' academic learning, nor can we assert that the cost of causing marginal students to drop out outweighs the benefit of increased student achievement of the overall population. Future research can be done by linking HSPA test scores with information on post-high-school indicators to study the long-term impacts of the exam. Furthermore, it is important to know whether students who drop out early will drop out permanently, especially because the cost of permanent dropouts is clearly higher than that of temporary dropouts (Belfield & Levin, 2007).

References:

Angrist, J, & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics, 114*(2), 533-575.

Belfield, C., & Levin, H.M. (2007). The economic losses from high school dropouts in California. California Dropout Research Project Report 1.

Chay, K. Y., McEwan, P. J. & Urquiola, M. (2005). The central role of noise in evaluating interventions that use test scores to rank schools. *American Economic Review, 95*(4), 1237–1258.

Center on Education Policy. (2004). *State High School Exit Exams: A Measuring Reform.* Washington, DC: Author.

Center on Education Policy. (2006a). Retrieved on December 19, 2008 from http://www.cep-dc.org.

Center on Education Policy. (2006b). *State High School Exit Exams: A Challenging Year.* Washington, DC: Author.

Center on Education Policy. (2007). *State High School Exit Exams: Working to Raise Test Scores.* Washington, DC: Author.

Cornell, D. G., Krosnick, J. A. & Chang, L.C. (2006). Student reactions to being wrongly informed of failing a high-stakes test. *Educational Policy, 20*(5). 718-751.

Dee, T. S., & Jacob, B.A. (In Press). Do high school exit exam influence education attainment or labor market performance? In Gamoran, A., (Eds.), *Will No Child Left Behind help close the poverty gap?* Washington, D.C.: Brookings University Press.

Fine, M., Pappas, L., Karp, S., Hirsch, L., Sadovnik, A., Keeton, A. & Bennett, M. (2007). New Jersey's special assessment: Loophole or lifeline? A policy brief. Education Law Center.

Garcia, P., & Gopal, M. (2003). The relationship to achievement on the California high school exit exam for language minority students. *NABE Journal of Research and Practice,1*(1), 123-137.

Griffin, B. W., & Heidorn, M.H. (1996). An examination of the relationship between minimum competency test performance and dropping out of high school. *Educational Evaluation and Policy Analysis, 18*(3), 243-252.

Greene, J. P., & Winters, M.A. (2004). Pushed out or pushed up? Exit exams and dropout rates in public high schools. Manhattan Institute Education Working Paper 5.

Hu, W. (2008). A plan to cut the high school dropout rate. *New York Times.* October 24. Retrieved on March 30, 2009 from http://www.nytimes.com/2008/10/26/nyregion/new-jersey/26educnj.html?_r=1&ref=education&oref=slogin

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

Jacob, B. A. (2001). Getting tough? The impact of high school graduation exams. *Educational Evaluation and Policy Analysis, 23*(2), 99-122.

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

Kane, T. J. (2003). A quasi-experimental estimate of the impact of financial aid on college-going. National Bureau of Economic Research Working Paper 9703.

Lee, D. & Card, D. (2008). Regression discontinuity inference with specification error, *Journal of Econometrics, 142*(2), 655-74.

Lemieux, T., & Milligan, K. (2008). Incentive effects of social assistance: A regression Discontinuity Approach. *Journal of Econometrics, 127*(2), 807-828.

Mackey, P. E. (2006). *New Jersey's public schools: A biennial report for the people of New Jersey, 2006-07 edition.* New Brunswick, NJ: Public Education Institute.

Martorell, F. (2004). Do high school graduation exams matter? Evaluating the effects of exit exam performance on student outcomes. PhD Dissertation Chapter, University of California, Berkeley.

McCrary, J., & Royer, H. (2005). The effect of maternal education on fertility and infant health: Evidence from school entry policies using exact date of birth. Retrieved on March 30, 2009 from http://sitemaker.umich.edu/hroyer/files/jmp.pdf.

McGray, D. (2006). Counseling kids to graduation and beyond. *Los Angeles Times.* September 6. Retrieved on March 30, 2009 from http://www.newamerica.net/publications/articles/2006/counseling_kids_to_graduation_and_beyond_4011.

Muller, C. (1998). The minimum competency exam requirement, teachers' and students' expectations and academic performance. *Social Psychology of Education, 2*(2), 199-216.

Muller, C., & Schiller, K.S. (2000). Leveling the paying field? Students' educational attainment and states' performance testing. *Sociology of Education, 73*(3), 196-218.

National Commission on Excellence in Education (1983). A nation at risk: The imperative for educational reform. Open letter to the American people. Retrieved on March 30, 2009 from http://www.ed.gov/pubs/NatAtRisk/title.html.

New Jersey Department of Education. (2007). *March 2007 and October 2007 High School Proficiency Assessment (HSPA) cycle I and cycle II score interpretation manual.* NJDOE PTM 1506.11.

Ou, D. (2009). To leave or not to leave? A regression discontinuity analysis of the impact of failing the high school exit exam. CEP Discussion Paper No. 907. Centre for Economic Performance, London School of Economics.

Papay, J., Murnane, R.J., & Willett, J.B. (2008). The consequences of high school exit examinations for struggling low-income urban students: Evidence from Massachusetts. National Bureau of Economic Research Working Paper 14186.

Reardon, S. F., Arshan, N., Atteberry, A. & Kurlaender, M. (2008). High stakes, no effects: Effects of failing the California High School Exit Exam. Paper presented at the Annual Meeting of the Association for Pubic Policy Analysis and Management, Los Angeles.

Trochim, W. M. K. (1984). *Research design for program evaluation: The regression discontinuity approach.* Beverly Hills, CA: Sage Publications.

van der Klaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment: A regression-discontinuity approach. *International Economic Review, 43*(4), 1249-1287.

van der Klaauw, W. (2008). Regression-discontinuity analysis: A survey of recent developments in Economics. *LABOUR, 22*(2), 219-245.

Warren, J. R., Jenkins, K.N. & Kulick, R.B. (2006). High school exit examinations and state-level completion and GED pass rates: 1975-2002. *Educational Evaluation and Policy Analysis, 28*(2), 131-152.

Warren, J. R., & Edwards, M.R. (2005). High school exit examinations and high school completion: Evidence from the early 1990s. *Educational Evaluation and Policy Analysis, 27*(1), 53-74.

Wilson, L. (2008). Exit strategies: Confronting faulty grad tests. *WireTap Magazine.* July 3. Retrieved on March 30, 2009 from http://www.wiretapmag.org/education/43620/.

TABLE 1: Descriptive Statistics

Panel A. Student Characteristics by Cohort

Variables	Cohort1	Cohort 2	Cohort 3	Cohort 4	All cohorts
Female	.51	.51	.51	.51	.51
White	.66	.66	.66	.64	.66
Black	.13	.13	.13	.13	.13
Hispanic	.12	.12	.12	.13	.12
Economically Disadvantaged	.12	.12	.13	.13	.13
Limited English Proficiency	.03	.03	.03	.03	.03
Special Education Students	.07	.07	.06	.07	.07
Age	17.02	17.02	17.01	17.00	17.01
	[.56]	[.56]	[.55]	[.53]	[.56]
Average Math	16.10	16.50	21.30	25.45	19.98
Score (adjusted)	[43.83]	[42.08]	[41.32]	[42.80]	[42.69]
Average LAL	21.86	22.77	25.46	27.65	24.52
Score (adjusted)	[4.07]	[38.79]	[35.52]	[34.82]	[37.36]
Initial Passing Rate: Math	.73	.71	.76	.81	.76
Initial Passing Rate: LAL	.85	.85	.88	.89	.87
Exit (11th grade)	.00	.03	.02	.03	.02
Observations	72,561	72,955	74,769	79,663	299,948
Exit if failing at least one subject initially	.00	.09	.09	.12	.07
Observations	21,720	22,780	19,567	17,380	81,447

Panel B. Student Characteristics by Subgroup

Variables	Female	White	Black	Hispanic	Economically Disadvantaged	Limited English Proficiency
Average Math Score (adjusted)	17.42	28.15	-9.52	-2.39	-5.19	-20.06
	[41.33]	[37.38]	[47.29]	[44.68]	[45.12]	[45.05]
Average LAL Score (adjusted)	28.24	31.19	5.63	5.22	3.36	-33.37
	[35.83]	[31.44]	[45.50]	[44.78]	[44.10]	[43.96]
Initial Passing Rate: Math	.73	.85	.44	.52	.49	.28
Initial Passing Rate: LAL	.89	.93	.71	.70	.68	.22
Exit (11th grade)	.02	.01	.05	.04	.04	.07
Observations	152,722	197,203	38,982	36,863	39,514	9,000
Exit if failing at least one subject initially	.07	.06	.09	.08	.08	.08
Observations	43,215	34,138	22,910	19,166	22,027	7,665

Panel C. Indicators for Retest

Variables	Cohort1		Cohort 2		Cohort 3		Cohort 4		All cohorts	
	Math	LAL	Math	LAL	Math	LAL	Math	LAL	Math	LAL
Passing rate	.30	.43	.31	.59	.22	.45	.22	.37	.27	.46
Observations	14,853	6,865	16,098	6,951	13,474	5,901	10,810	5,977	55,235	25,694
Exit if fail the retest	.01	.02	.04	.08	.04	.08	.05	.07	.04	.06
Observations	10,456	3,895	11,087	2,866	10,558	3,270	8348	3,782	40,539	13,813

Variables	Female		White		Black		Hispanic		Economically Disadvantaged		Limited English Proficiency	
	Math	LAL	Math	LAL	Math	LAL	Math	LAL	Math	LAL	Math	LAL
Passing rate	.30	.44	.30	.68	.31	.42	.22	.34	.22	.35	.12	.16
Observations	32,810	11,604	20,773	7,244	17,067	7,623	14,147	8,376	10,810	9,575	5,339	5,736
Exit if fail the retest	.03	.06	.04	.07	.04	.06	.04	.06	.04	.05	.04	.04
Observations	24,420	6,456	12,481	2,336	14,375	4,408	11,454	5,554	13,571	6,191	4,720	4,790

Note: Panel A. Columns 1,2,3,4 contain 11th graders in the general education program in a public high school who took the HSPA the first time in 2002, 2003, 2004, and 2005 respectively. The sample in Column 5 is the sum of students from Columns 1 to 4. Panel B include subsets of students from the sample in Column 5, Panel A. Panel C includes students who didn't pass either subject initially and but signed up for the initial retest in the fall of 12^{th} grade. The average math and LAL scores were adjusted by subtracting the cutoff level, 200, from the actual scale score. Standard deviations are presented in square brackets.

TABLE 2 Regression Discontinuity Estimates of Exiting from High School after Initial Attempt to Pass the HSPA

					Bandwidth			Cohort			
	All		All		+/-30	+/-40	+/-50	1	2	3	4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Exit after failure†					.048	.055	.058				
Specification	cubic	cubic	quartic	quartic	linear	linear	linear	cubic	cubic	cubic	cubic
Initial Math Test	-.013***	-.011***	-.013***	-.010***	-.012***	-.010***	-.009***	-.000	-.015***	-.013***	-.017***
	[.002]	[.001]	[.002]	[.001]	[.001]	[.001]	[.001]	[.000]	[.003]	[.002]	[.004]
Observations	299,948	299,948	299,948	299,948	138,633	180,748	222,672	72,561	72,955	74,769	79,663
					Bandwidth			Cohort			
	All		All		+/-25	+30/-40	+/-50	1	2	3	4
Exit after failure†					.061	.068	.070				
Specification	cubic	cubic	quartic	quartic	linear	linear	linear	cubic	cubic	cubic	cubic
Initial LAL Test	-.006***	-.005***	-.006***	-.005***	-.007***	-.005***	-.003***	-.000	-.007***	-.010***	-.004***
	[.002]	[.002]	[.002]	[.002]	[.001]	[.001]	[.001]	[.000]	[.003]	[.002]	[.002]
Observations	299,948	299,948	299,948	299,948	94,478	128,760	242,510	72,561	72,955	74,769	79,663
Controls	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y

Note: † Each cell is the average exit rate for students who scored under the passing level within each bandwidth. Regressions were conducted on students who took the HSPA tests for the first time in the spring of 11th grade and who did not have an Individual Education Program exemption. Students who had been in the school for less than one year when that test was administered were excluded. The dependent variable is a binary variable indicating whether the student exited from high school the semester following the initial test. Each cell represents a separate estimate of the discontinuity in failing the subject of testing listed in the first column and the specific functional form. Test scores are adjusted by subtracting the passing score, 200, from the original scale score. Robust standard errors clustered on the adjusted score level are in parentheses. All models contain a linear interaction term for test score and the passing status of the subject, allowing the slopes to differ on both sides of the HSPA cutoff. Controls include dummies for female gender, special education students, economically disadvantaged students, LEP students, age, White, Black, Hispanic, and socioeconomic status of citizens in the district.
* $p < .1$ ** $p < .05$ *** $p < .01$.

TABLE 3 Regression Discontinuity Estimates of Exiting after Second Attempt to Pass the HSPA

	All Cohorts				Bandwidth		Cohort			
					-/+25	-/+30	1	2	3	4
Exit after failure†					.019	.021				
Specification	quadratic	quadratic	cubic	cubic	quadratic	cubic	quadratic	quadratic	quadratic	quadratic
Math 1st Retest	-.015***	-.013***	-.016***	-.015***	-.015***	-.015***	-.004**	-.011***	-.018***	-.024***
	[.002]	[.001]	[.003]	[.003]	[.002]	[.004]	[.002]	[.003]	[.004]	[.006]
Observations	55,235	55,235	55,235	55,235	30,496	41,499	14,853	16,089	13,474	10,810
Exit after failure†					.030(.171)	.031(.174)				
LAL 1st Retest	-.007	-.007**	-.012*	-.011**	-.016***	-.011***	-.005*	-.005	-.005	-.017**
	[.005]	[.003]	[.007]	[.005]	[.005]	[.004]	[.003]	[.007]	[.007]	[.004]
Observations	25,694	25,694	25,694	25,694	13,980	18,188	6,865	6,951	5,901	5,977
Controls	N	Y	N	Y	Y	Y	Y	Y	Y	Y

Note: † Each cell is the average exit rate for students who scored under the passing level within each bandwidth. Standard errors are in brackets and are robust to heteroskedasticity and correlation within clusters. The dependent variable is a binary variable indicating whether the student exited from high school the semester following the second test. Each cell represents a separate estimate of the discontinuity in failing the subject of testing listed in the left-hand column. For more detail see the note under Table 2.
* $p < .1$ ** $p < .05$ *** $p < .01$.

TABLE 4 Regression Discontinuity Estimates of Exiting from High School for Subgroups (Initial Attempt)

	Male		Female		Limited English Proficiency (LEP)		Non-LEP		White	
Specification	cubic	quartic	cubic	quartic	cubic	quartic	cubic	quartic	cubic	quartic
Math	-.005***	-.004***	-.007***	-.007***	-.007	-.012	-.005***	-.005***	-.005***	-.004***
Initial Test	[.001]	[.001]	[.001]	[.001]	[.006]	[.009]	[.001]	[.001]	[.001]	[.001]
LAL	-.003***	-.003**	-.001	-.002**	-.025**	-.018*	-.002***	-.003***	-.005***	-.004**
Initial Test	[.001]	[.001]	[.001]	[.001]	[.012]	[.011]	[.001]	[.001]	[.001]	[.001]
Observations	152,722	152,722	152,722	152,722	9,000	9,000	290,948	290,948	197,203	197,203
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

	Hispanic		Black		Economically Disadvantaged		Not Economically Disadvantaged	
Specification	cubic	quartic	cubic	quartic	cubic	quartic	cubic	quartic
Math	-.013***	-.015***	-.014***	-.015***	-.013***	-.016***	-.006***	-.006***
Initial Test	[.002]	[.002]	[.003]	[.003]	[.002]	[.003]	[.001]	[.001]
LAL	-.008***	-.010***	.000	-.003	-.006***	-.006***	-.003***	-.003***
Initial Test	[.002]	[.002]	[.002]	[.002]	[.001]	[.002]	[.001]	[.001]
Observations	36,863	36,863	38,982	38,982	39,514	39,514	260,434	260,434
Controls	Y	Y	Y	Y	Y	Y	Y	Y

Note: Standard errors are in brackets and are robust to heteroskedasticity and correlation within clusters. Regressions on the specified subgroups of students who took the HSPA test for the first time in the spring of 11th grade (Table 4) and who did not have an Individual Education Program exemption. Students who had been in the school for less than one year when that test was administered were excluded. The dependent variable is a binary variable indicating whether the student exited from high school the semester following the first test. Each cell represents a separate estimate of the discontinuity in failing the subject of testing listed in the left-hand column. For more detail see the note under Table 2.

TABLE 5 Regression Discontinuity Estimates of Exiting from High School for Subgroups (Retest)

Specification	Male		Female		Limited English Proficiency (LEP)		Non LEP		White	
	quadratic	cubic	quadratic	cubic	quadratic	cubic	quadratic	cubic	quadratic	cubic
Math	-.015***	-.018***	-.012***	-.012***	-.006	-.008	-.013***	-.015***	-.014***	-.014***
1st Retest	[.003]	[.005]	[.002]	[.003]	[.006]	[.010]	[.002]	[.003]	[.003]	[.003]
	22,425	22,425	32,810	32,810	5,339	5,339	49,896	49,896	20,773	20,773
LAL	-.005	-.009	-.010**	-.014**	-.004	-.014	-.008**	-.011**	-.008	-.009*
1st Retest	[.003]	[.005]	[.004]	[.006]	[.005]	[.014]	[.004]	[.005]	[.005]	[.005]
Observations	14,090	14,090	11,604	11,604	5,736	5,736	19,958	19,958	7,244	7,244
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Specification	Hispanic		Black		Economically Disadvantaged		Not Economically Disadvantaged	
	quadratic	cubic	quadratic	cubic	quadratic	cubic	quadratic	cubic
Math	-.013***	-.018**	-.009***	-.009	-.012***	-.014**	-.014***	-.015**
1st Retest	[.003]	[.008]	[.002]	[.005]	[.003]	[.009]	[.002]	[.003]
	14,147	14,147	17,067	17,067	16,305	16,305	38,930	38,930
LAL	-.002	-.027	-.012*	-.015	-.000	-.007	-.012***	-.014***
1st Retest	[.002]	[.016]	[.007]	[.011]	[.004]	[.007]	[.004]	[.006]
Observations	8,376	8,376	7,623	7,623	9,575	9,575	16,119	16,119
Controls	Y	Y	Y	Y	Y	Y	Y	Y

Note: Standard errors are in brackets and are robust to heteroskedasticity and correlation within clusters. Regressions on the specified subgroups of students who took the HSPA retest in the fall of 12th grade and who did not have an Individual Education Program exemption. Students who had been in the school for less than one year when that test was administered were excluded. The dependent variable is a binary variable indicating whether the student exited from high school the semester following second test. Each cell represents a separate estimate of the discontinuity in failing the subject of testing listed in the left-hand column. For more detail see the note under Table 2.

* $p < .1$ ** $p < .05$ *** $p < .01$.

TABLE 6 Effects of Failing At Least One Subject in the HSPA

	All		Cohort 1		Cohort 2		Cohort 3		Cohort 4	
	1st Test	Retest	1st Test	Retest	1st Test	Retest	1st Test	Retest	1st Test	Retest
Specification	quadratic	linear	quadratic	linear	quadratic	linear	quadratic	linear	quadratic	linear
Math Test	-.007	-.010**	-.000	-.002	.003	-.013	-.026*	-.015	-.002	-.007
	[.007]	[.004]	[.002]	[.006]	[.014]	[.014]	[.015]	[.011]	[.009]	[.018]
Observations	39,325	11,921	10,649	3,439	10,589	2,590	8,940	2,854	9,147	3,038
LAL Test	-.003	-.007*	-.000	-.007*	-.011*	-.006	-.008	-.003	.013	-.012
	[.006]	[.003]	[.001]	[.003]	[.007]	[.008]	[.005]	[.009]	[.009]	[.008]
Observations	73,459	16,894	19,729	4,486	20,874	4,621	17,671	4,011	15,185	3,776
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Note: Standard errors are in brackets and are robust to heteroskedasticity and correlation within clusters. The dependent variable is a binary variable indicating whether the student exited from high school the semester following the initial or second test. Each cell represents a separate estimate of the discontinuity in failing the subject of testing listed in the left-hand column conditional on failing the other subject. The functional form of the test score is specified in row 3. For more detail see the note to Table 2.
* $p < .1$ ** $p < .05$ *** $p < .01$.

[FIGURE I ABOUT HERE]

Appendix

a. Sample construction

The raw data are provided by the NJDOE, which contains student characteristics and test scores in math and Language Art Literacy for all test takers in each test administration from March 2002 to March 2006.

HSPA Administration by Academic Cohort

Cohort	Spring 2002	Fall 2002	Spring 2003	Fall 2003	Spring 2004	Fall 2004	Spring 2005	Fall 2005	Spring 2006
2003	X	X	X						
2004			X	X	X				
2005					X	X	X		
2006							X	X	X

As indicated from the table above, I traced students in each cohort who took the HSEE from 11th grade to the end of 12th grade. Information on all three tests (including two retests) will be combined if applicable. Datasets of three test administrations were merged using a unique student HSPA test ID and school ID. Students who transferred to a different school within the state during the test administration were not included in my sample. In order to ensure that I correctly matched students over three testing administrations, I also compared the students' dates of birth. I then compared the sample size with the official publicized statistics on the NJDOE website, which indicates that the numbers were identical before I conducted further selections (or deletions) and merging.

I cleaned the data by cohort, and then merged the clean datasets of the four cohorts together into a pooled sample. For each cohort, I deleted observations that did not have consistent date of birth (DOB) information in different test administrations. Students who were older than 21 years and younger than 15 years were also excluded. I took out all migrant students (27 in total from 2002 to 2006). Moreover, I excluded from the sample any students with missing or incorrect coding on gender, grade, and ethnicity. The cleaned sample sizes for the four cohorts were 84,325, 103,838, 108,704, and 118,078. The pooled data therefore contained 414,945 students for all four cohorts.

In my analysis, I only included 356,359 observed cases who initially took the HSPA in the spring semester of 11th grade. I included the students who attended some professional courses but enrolled in a general education program in a public school. However, I did not consider students who are from vocational schools, private schools, and rural districts (18,270 students). I further excluded 12,270 students who had been in their school for less than one year when the test was administered, as well as 24,613 special education students who got the Individual Education Program exemption for the first test and 1,258 special education students who got the Individual Education Program exemption for the retest. This left a final sample of 299,948 observations.

The retest sample includes all students who failed at least one subject initially and were enrolled in the fall semester. I excluded all special education students (N = 9,741) because the special education students did not seem to be randomly distributed around the cutoff score from the regression discontinuity validity test.

b. *Exiting from High School After Initial Attempt & Second Test*

The "exit" status is indicated in the retest records. This information was taken from the "Exit Student Roster" in the NJ HSPA database, which was reported and updated by the school in which the student was enrolled at the beginning of every semester. Every student will appear in a subsequent test file as a 12th grader, retained 11th grader, exit, return student, or being enrolled in an adult school. If a student passed the exam on the first try and left the school without registering for the next semester/term in the school or other high schools in New Jersey, the school enrollment would be shown as "exit." So more precisely, the "exit" represents a student's dropout episode that occurs between the two test administrations, for example, exiting from high school after the first exam means student dropped out before starting the 12th grade fall semester. Similarly, "exit" from high school after the second test means students did not enroll in school for the spring semester of 12th grade.

Appendix Table A-I
Regression Discontinuity Estimates on Pre-Determined Student Characteristics

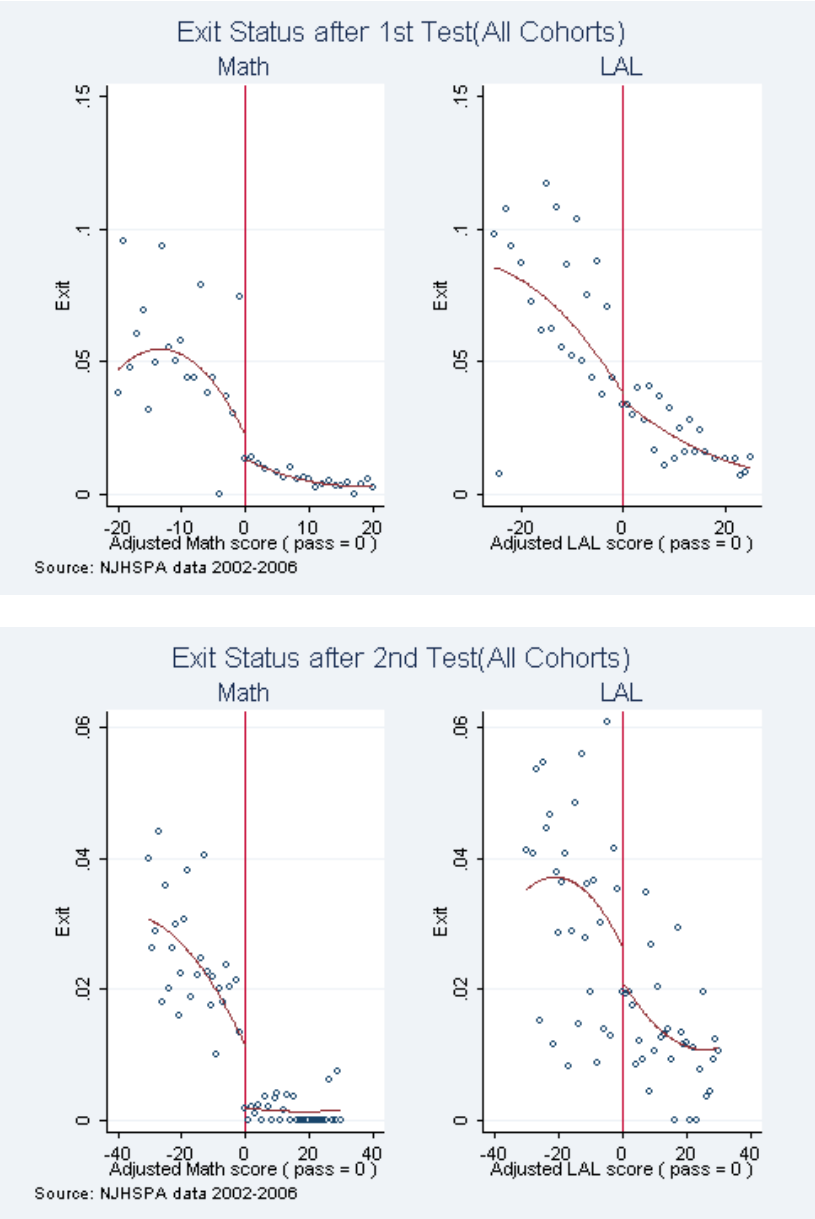
Variables	Math		LAL	
	1st Test	**Retest**	**1st Test**	**Retest**
Female	-.004	.001	-.025***	-.047***
	(.008)	(.008)	(.008)	(.011)
White	-.010	-.049***	-.000	.015
	(.011)	(.009)	(.011)	(.017)
Black	-.007	.006	-.008*	.000
	(.006)	(.005)	(.004)	(.015)
Hispanic	.005	.018***	.012*	.023***
	(.009)	(.006)	(.007)	(.009)
Economically Disadvantaged	.009*	.012**	.002	.005
	(.005)	(.006)	(.006)	(.011)
Limited English Proficiency	.003	.012***	.006**	.014**
	(.003)	(.003)	(.003)	(.007)
Older Students	.001	.000	-.002	-.000
	(.001)	(.000)	(.002)	(.000)
Special Education Student	.011**	-	.006*	-
	(.004)	-	(.003)	-
Lowest SES quartile	.015	.038***	-.009	.006
	(.010)	(.008)	(.007)	(.014)
Second-Lowest SES quartile	-.010	-.026***	.032***	.040***
	(.007)	(.009)	(.007)	(.008)
Second-Highest SES quartile	-.005	-.017**	.012	-.001
	(.004)	(.007)	(.007)	(.011)
Cohort 1	-.029	.004	-.012	-.032
	(.046)	(.069)	(.067)	(.111)
Cohort 2	-.031	-.020	-.014	.047
	(.072)	(.079)	(.073)	(.083)
Cohort 3	.063	.028	.062	.002
	(.084)	(.067)	(.123)	(.078)
Number of Observations	299,948	55,235	299,948	25,694

Data source: New Jersey High School Proficiency Assessment (HSPA) Database 2002–2006.

Notes:
All scores are adjusted relative to the passing cutoff of the HSPA test. All estimates for the 1st test are marginal effects of the dependent variables specified in the first column at the passing cutoff in the probit regression based on the fourth polynomial form of the specific test score using the selected samples. All estimates for the retest are marginal effects of the dependent variables specified in the first column at the passing cutoff in the probit regression based on the quadratic form of the specific test score, with the dummy indicating passing the test as well as other students. Standard errors in parentheses are clustered at the score level. Special education students exempted from the HSPA are excluded from the analysis of the initial test; all special education students are excluded in the analysis of the retest.
* $p < .1$ ** $p < .05$ *** $p < .01$.

[FIGURE A-I ABOUT HERE]

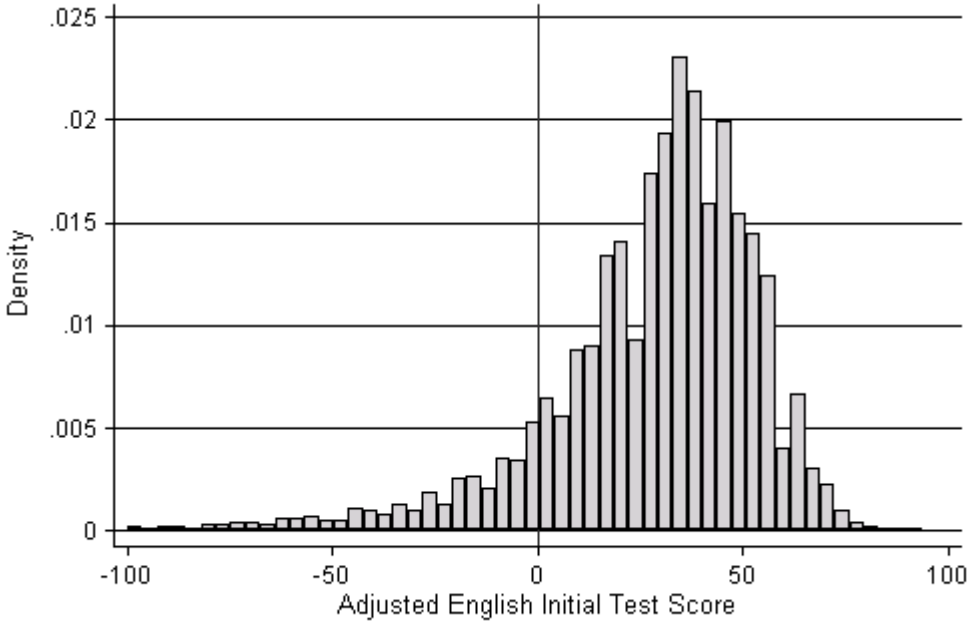
Figure I. The Relationship between 1st and 2nd HSPA Test Scores and Exiting High School by Testing Subject for Four Cohorts (2002-2005)

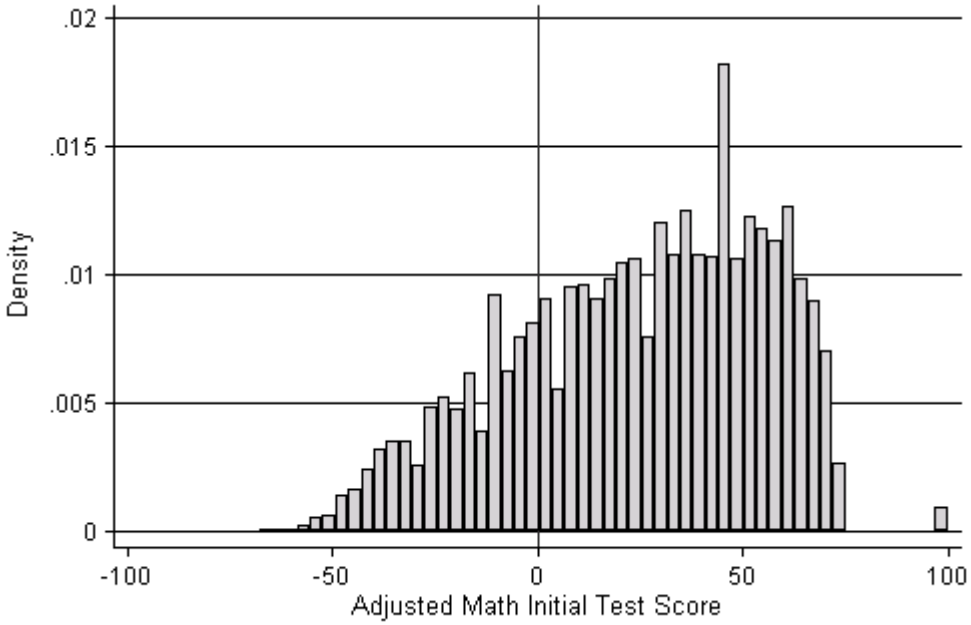


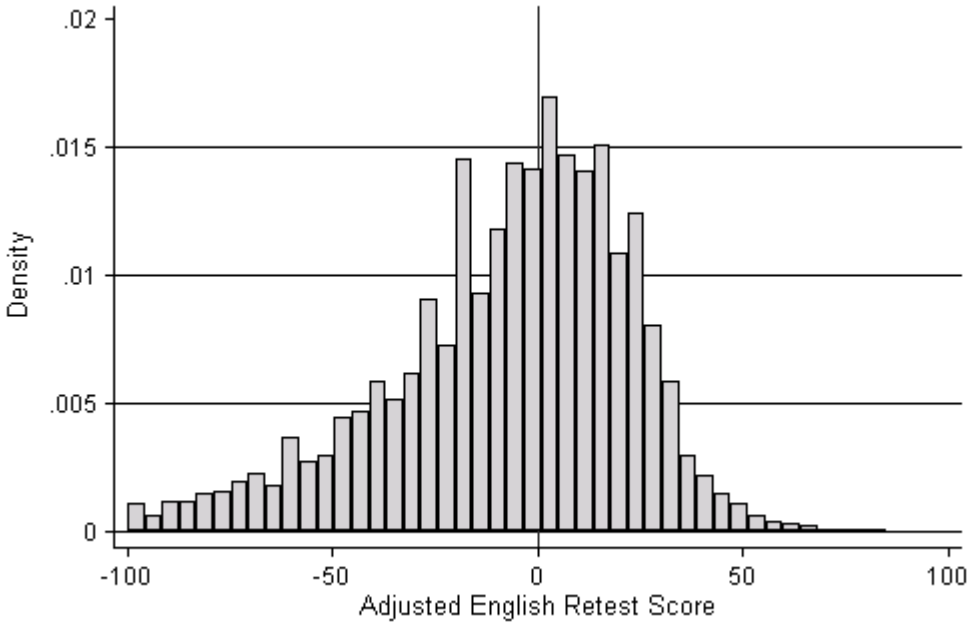
Note: The left panel shows the probability of exiting (Y-axis) from high school along the adjusted math score (X-axis); the right panel shows the probability of exiting (Y-axis) from high school along the adjusted LAL score (X-axis). The circles reflect the actual average probability of exiting at each score point. The solid line represents the predicted exit probability from a fourth-order polynomial function of the initial test score (upper graph) or a cubic form for the retest score (lower graph), along with an interaction between the passing dummy and a linear term for the test score.

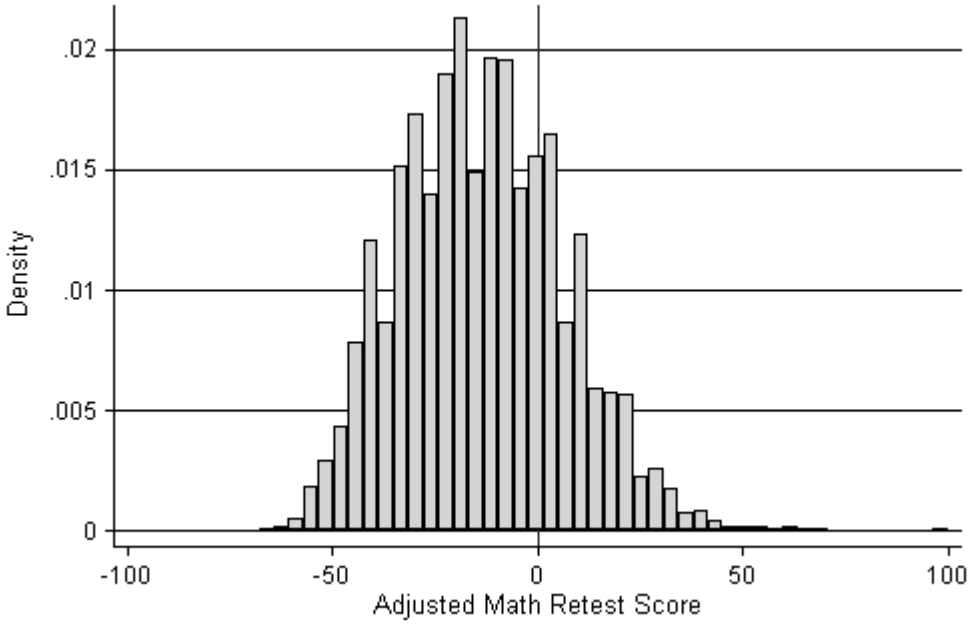
Appendix

Figure A-I. Histograms of English and Math Scores for Initial and First Retest (2002-2006)









Note: Test scores are adjusted by subtracting the actual score from the pass/fail cut-off. Data come from the New Jersey Department of Education High School Proficiency Assessment Test Score Data (2002-2006).