NBER WORKING PAPER SERIES

THE ELITE ILLUSION:
ACHIEVEMENT EFFECTS AT BOSTON AND NEW YORK EXAM SCHOOLS

Atila Abdulkadiroglu
Joshua D. Angrist
Parag A. Pathak

Working Paper 17264
http://www.nber.org/papers/w17264

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2011

Our thanks to Kamal Chavda, Jack Yessayan, and the Boston Public Schools and to Jennifer Bell-Ellwanger, Thomas Gold, Jesse Margolis, and the New York City Department of Education for graciously sharing their data. The views expressed here are those of the of authors and do not reflect the views of the Boston Public Schools, the NYC Department of Education, or the National Bureau of Economic Research. We thank participants in the June 2010 Tel-Aviv Frontiers in the Economics of Education conference for comments and in particular our discussant Jonah Rockoff, who also contributed data on teacher tenure in NYC. We're also grateful to Daron Acemoglu, Gary Chamberlain, Glenn Ellison, and Guido Imbens for helpful discussions along the way. Weiwei Hu and Miikka Rokkanen provided superb research assistance. Pathak also thanks the Graduate School of Business at Stanford University, where parts of this work were completed, and the NSF for financial support.

The Elite Illusion: Achievement Effects at Boston and New York Exam Schools
Atila Abdulkadiroglu, Joshua D. Angrist, and Parag A. Pathak
NBER Working Paper No. 17264
July 2011
JEL No. I20

ABSTRACT

Talented students compete fiercely for seats at Boston and New York exam schools. These schools are characterized by high levels of peer achievement and a demanding curriculum tailored to each district's highest achievers. While exam school students clearly do very well in school, the question of whether an exam school education adds value relative to a regular public education remains open. We estimate the causal effect of exam school attendance using a regression-discontinuity design, reporting both parametric and non-parametric estimates. We also develop a procedure that addresses the potential for confounding in regression-discontinuity designs with multiple, closely-spaced admissions cutoffs. The outcomes studied here include scores on state standardized achievement tests, PSAT and SAT participation and scores, and AP scores. Our estimates show little effect of exam school offers on most students' achievement in most grades. We use two-stage least squares to convert reduced form estimates of the effects of exam school offers into estimates of peer and tracking effects, arguing that these appear to be unimportant in this context. On the other hand, a Boston exam school education seems to have a modest effect on high school English scores for minority applicants. A small group of 9th grade applicants also appears to do better on SAT Reasoning. These localized gains notwithstanding, the intense competition for exam school seats does not appear to be justified by improved learning for a broad set of students.

Atila Abdulkadiroglu
Duke University
Department of Economics
Durham, NC 27708
atila.abdulkadiroglu@duke.edu

Joshua D. Angrist
Department of Economics
MIT, E52-353
50 Memorial Drive
Cambridge, MA 02142-1347
and NBER
angrist@mit.edu

Parag A. Pathak
MIT Department of Economics
50 Memorial Drive
E52-391C
Cambridge, MA 02142
and NBER
ppathak@mit.edu

1 Introduction

The Boston and New York City public school systems include a handful of highly selective exam schools. Unlike most other American public schools, exam schools screen applicants on the basis of a competitive admissions test. Boston's exam school flagship, the Boston Latin School, is the oldest high school in the country; New York's venerable Bronx High School of Science and Stuyvesant High School also have storied histories. Just as many American high school seniors work and compete to gain admission to the country's most selective colleges and universities, younger students and parents in a few cities aspire to win coveted seats at top exam schools.[1]

Fewer than half of Boston applicants win a seat to one of three exam schools, and less than a sixth of exam school applicants are offered a seat at the three original exam schools in New York. Because exam school offers are test-based, exam school students have significantly higher test scores than do typical public school students. The pre-application Math and English scores of students offered a seat at one of the least competitive Boston and New York exam schools are on the order of 0.5-0.7 standard deviations (hereafter, σ) higher than the scores of those who apply but not offered.[2] Differences in baseline performance between applicants at the most competitive exam school and those in regular public schools are even more impressive, at over 1.5σ for Boston 7th graders and over 1.25σ for New York 9th graders.

At first blush, the intense competition for an exam school education is understandable. By any measure, exam school students are well ahead of virtually all other public school students. It is easy to see why many parents dream of placing their children in such a school. At the same time, it's also clear that at least some of the achievement advantage associated with exam school attendance reflects the schools' admissions policies and is not caused by attendance per se. After all, exam school students are a highly select group, a fact that must influence naive comparisons between exam school students and anyone else.

The purpose of this paper is to evaluate the causal effects of exam school attendance on applicant achievement as measured by standardized tests. We use a regression discontinuity (RD) research design that, if successful, eliminates the selection bias that contaminates naive comparisons. Our strategy in a nutshell is to compare the scores of exam school applicants who barely clear the admissions cutoff to the scores of those who fall just below. Those who

[1]Boston and New York exam schools claim a long list of distinguished alumni, including 11 Nobel laureates, five signers of the Declaration of Independence, and a few dozen distinguished economists, including Robert Fogel, Jerry Green, Jesse Shapiro (Stuyvesant); Claudia Goldin (Bronx Science); and Gary Chamberlain and Charles Manski (Boston Latin School). Other American cities with similarly selective public high schools include Chicago, San Francisco, and Washington.

[2]Pre-application scores come from 4th grade for 7th grade applicants, and from 8th grade Math and 7th grade English for 9th grade applicants. We refer to these pre-application scores as baseline scores.

clear the cutoff are more likely to attend an exam school (and attend for longer) than those who fall below the cutoff, though some in the latter group also eventually succeed in gaining admission. A dummy for clearing the admissions cutoff is therefore an instrument for exam school attendance in a fuzzy regression discontinuity (RD) setup.

RD estimation of the effects of an exam school offer is in principle straightforward, but implementation raises a number of practical challenges. Chief among these is the fact that applicants can apply to as many exam schools as they like. The cutoffs for the three traditional exam schools in each city are closely spaced and, in the case of Boston, involve distinct, though highly correlated running variables. Data-driven bandwidths for nonparametric estimation tend to straddle cutoffs, raising the possibility of confounding in a reasonably small neighborhood of the cutoff. Fully non-parametric procedures seem unlikely to have good finite-sample properties since they're predicated on a no-confounding assumption in an empirically relevant neighborhood of each admissions cutoffs. As a result, we implement a simple parametric correction to the Imbens and Kalyanaraman (2010) estimating equations at the heart of our empirical strategy. This approach appears to work well: a reasonably consistent picture of causal effects emerges from both fully parametric and non-parametric analyses.

Before turning to the details of the empirical analysis, it's worth asking what exam school attendance means for an admitted student. First, exam school students study with peers who have similarly high levels of ability. If peer effects are important, this alone should boost achievement. Second, the exam school curriculum is meant to challenge the highly able exam school population. Finally, some exam schools have resources and facilities typically unavailable at other public schools, such as modern science labs and well-equipped athletic facilities. The resource advantage is not entirely clear, however, since many exam schools operate with class sizes substantially larger than is typical of the schools in their host district. The exam school estimates reported here, therefore, seem most likely to be informative about a combination of peer and tracking effects on high-achieving public school students. This observation motivates us to construct instruments for the length of time exposed to an exam school curriculum and average levels of peer achievement, two potentially endogenous mediating variables for which there is a strong exam-school-offer first stage. In view of the fact that no single channel may satisfy an instrumental variables (IV) exclusion restriction, we report alternative two-stage least squares (2SLS) results using these instruments.

Our results offer little evidence of an achievement gain for those admitted to an exam school; most of the estimates can be interpreted as reasonably precise zeros, with a smattering of significant effects, both positive and negative. In other words, in spite of their exposure to much higher-achieving peers and a more challenging curriculum, marginal students admitted to exam schools generally do no better on a variety of standardized tests. One explanation for this finding is the local nature of RD: marginal applicants may be ill-positioned to benefit from an exam school education. We also show, however, that RD estimates for students who are in the upper half and upper quartile of the baseline test score distribution (a test that precedes the entrance examination) are broadly in line with the estimates for the entire population near admission cutoffs. Finally, we look at estimates for minorities, an inquiry motivated in part by litigation related to minority admissions. Here too we evidence of an exam school achievement boost on high school English tests for applicants in Boston. A small subset of admitted applicants also appears to score higher on the SAT than they otherwise might have.

The RD estimates reported here necessarily reflect the experiences of exam school applicants with entrance scores close to admissions cutoffs. This focus on marginal groups is an intrinsic feature of RD identification strategies. At the same time, estimation for these applicants can be seen as a practical virtue since estimates for marginal groups are of considerable scientific and policy interest. For one thing, applicants close to admissions cutoffs are still relatively high achievers, with measured ability far above that of most other urban public school students. Although education research often focuses on interventions meant to serve students in the lower tail of the ability distribution, the education production function for high achievers should also be of interest. Moreover, on the policy side, most commonly proposed innovations affecting exam school access, such as new campuses, earlier admissions grades, and strengthened minority or socioeconomic preferences, are likely to affect students near current admissions cutoffs. It's also worth noting that our applicant sample spans a wide range of ability. In 2009, Boston Latin students had SAT scores ranked among the top five schools in the state (634 in Math and 616 in Reading). At the other end, O'Bryant students have average SAT scores below the state mean (972 on the Reasoning test at O'Bryant vs. 1031 statewide), though still well above the BPS average (894 on the Reasoning test).

The next section describes Boston and New York exam schools in more detail and briefly reviews related literature. Section 3 discusses Boston data and descriptive statistics, while Section 4 lays out our discontinuity-based estimation framework. Section 5 presents the main analysis for Boston. We begin with Boston because Massachusetts state achievement scores - centrally and anonymously graded Math and English tests for multiple grades - appear to be more reliable than New York's Regents exams, which have a locally graded component and reflect the subject mix chosen by examinees. We also use the Boston data to look at effects on PSAT, SAT, and Advanced Placement exams. The Boston results include 2SLS estimates of peer and time-in-school effects and estimates for various subgroups Section 6 summarizes a parallel set of results for the effect of New York City's exam schools on Regents exams. The paper concludes in Section 7.

2 Boston and New York Exam Schools

Boston has three exam schools, each spanning grades 7-12. The best-known is the Boston Latin School, which enrolls about 2,400 students. Often described as the crown jewel of Boston's public school system, Boston Latin School was named a top 20 U.S. high school in the inaugural 2007 U.S. News & World Report school rankings. Founded in 1635, the Boston Latin School is America's first public school and the oldest still open (Goldin and Katz, 2008).[3] The Boston Latin School is a model for other exam schools. Imitators include the Brooklyn Latin School, recently opened in New York (Jan, 2006). The second oldest Boston exam school is Boston Latin Academy, formerly the Girls' Latin School. Opened in 1877, Latin Academy first admitted boys in 1972 and currently enrolls about 1,700 students. The John D. O'Bryant High School of Mathematics and Science (formerly Boston Technical High) is Boston's third exam school; O'Bryant opened in 1893 and currently enrolls about 1,200 students.

New York's three original academic exam schools are Stuyvesant High School, Bronx High School of Science, and Brooklyn Technical High School, each spanning grades 9-12. The New York exam schools were established in the first half of the 20$^{\text{th}}$ century and share a number of features with Boston's exam schools. For example, Stuyvesant and Bronx Science are members of the Newsweek list of elite public high schools and all three have appeared in the U.S. News & World Report rankings. Stuyvesant enrolls just over 3,000 students, Bronx Science enrolls 2,600-2,800 students, and Brooklyn Technical has about 4,500 students. In 2002, three new exam schools opened in New York: the High School for Math, Science and Engineering at City College, the High School of American Studies at Lehman College, and Queens High School for the Sciences at York College. In 2005, Staten Island Technical High School converted to exam status, while the Brooklyn Latin School opened in 2006. The admissions process for these new schools is the same as for the three original exam schools, but we omit them from our study because they are not as well established as the traditional exam schools, and some have unusual characteristics (e.g., small enrollment). Finally, we've structured the New York analysis to parallel that for Boston.[4]

A defining feature of an exam school education is exposure to high-achieving peers. The difference between the average pre-application achievement of students enrolled at the Boston Latin School and those enrolled at a traditional Boston school, reported in Table 1, is over two standard deviations for Math and about 1.75σ for English. Although the other two Boston

[3]Boston Latin School was established one year before Harvard College. Local lore has it that Harvard was founded to give graduates of Latin a place to continue their studies.

[4]Estimates including New York's new exam schools are similar to those generated by the three-school sample. Other selective New York public schools include the Fiorello H. LaGuardia High School, which focuses on visual and performing arts and admits students by audition, and Hunter College High School, which uses a unique admissions procedure and is not operated by the New York Department of Education.

exam schools are not as selective as Boston Latin, peer achievement gaps at O'Bryant and Latin Academy are still substantial (more than 1.0σ for Math and English at O'Bryant, and over 1.25σ at Latin Academy). Students at the three New York City exam schools also have much higher pre-application scores than students at traditional public high schools. Students enrolled at Brooklyn Technical are roughly 1.5σ ahead of the New York average in both Math and English, while the Stuyvesant score advantage is more than two standard deviations (shown in Table 9).

The challenging nature of an exam school curriculum can be gauged by the number of advanced placement (AP) courses. Katnani (2010) reports that Stuyvesant offers thirty-seven AP courses, while Boston Latin School offers 23. Stuyvesant boasts the the highest number of AP test-takers in the country, as well as the most scoring at least 3 or higher on AP tests, typically the minimum required for college credit (Saulny, 2005). In addition to AP courses, exam schools offer other advanced courses and academic experiences. At Bronx Science, for example, students have the opportunity to do research with local scientists. Bronx Science and Stuyvesant send many finalists to the Intel (formerly, Westinghouse) Science Talent Search. Many exam school students compete in the American Mathematics Contest and similar achievement-driven face-offs.

Along with their rich menu of course offerings, exam schools typically impose high graduation standards. Boston Latin students take four years of Latin and give declamations in grades 7-10. O'Bryant students enroll in six years of Math. The New York exam schools offer advanced diplomas based on academic and extra-curricular work beyond that required for New York State Regents diploma.

Some exam schools have endowments and raise money for special projects. These extra resources are used for college scholarships, faculty training, and facilities. Each year, the Boston Latin School Association contributes about $700,000 to the school's annual budget from an endowment of about $15 million. The Brooklyn Technical Alumni Foundation completed a fundraising campaign of $10 million for the school in 2005. These funds went to a robotics laboratory, library improvements, and a gym, among other things (Steinberg, 1998a). The alumni associations of Stuyvesant and Bronx Science made similar large pledges (Steinberg, 1998b). Katnani (2010) reports that the Harry V. Keefe Library-Media Center at the Boston Latin School, named after a three-million-dollar alumni donor, is "the most advanced school library in the world."

Course offerings and facility upgrades notwithstanding, exam schools look somewhat worse than traditional public schools in comparisons of class size. The average student-to-teacher ratio at the Boston Latin School is 22, compared to a district-wide average of 12 for middle schools and 15 for high schools, a comparison also documented in Table 1. In New York, the exam student-to-teacher ratio is roughly 31, compared to about 27 district-wide.[5]

Like other public school teachers in New York and Boston, exam school teachers are members of the local bargaining unit, and exam school staffing decisions are proscribed by the union contract in force. Teaching and administrative assignments at New York's exam schools attract scrutiny because exam school jobs are considered highly desirable (see, e.g., Stern (2003) and Kugel (2005)). In practice, the exam school teaching staff is more senior and more likely to be defined as highly qualified according to state certification standards. At New York's exam schools, 44% of teachers are age 48 or older, compared to about 30% at other non-exam schools.[6] Likewise, the Boston exam school teaching staff is substantially more senior than that at other BPS schools.

The proportion of minority applicants admitted to exam schools has often been a lightning rod for controversy. Under Boston's court-mandated 1970s desegregation plan, Federal Judge Arthur Garrity ordered that "at least 35% of each of the entering classes at Boston Latin School, Boston Latin Academy and Boston Technical High in September 1975 shall be composed of black and Hispanic students." This policy maintained the proportion of black and Hispanic students at roughly 35% for many years. Racial preferences in Boston exam school admissions were first challenged in 1996. Following a series of court proceedings, Boston exam school admissions have been purely exam and GPA-based since 1999 (Boston Public Schools, 2007).[7]

In the 1960s, civil rights groups argued that New York's exam school admissions test is biased against black and Puerto Rican applicants. These challenges ultimately led to the 1972 Hecht-Calandra Act, a state law guaranteeing that exam school admissions be based solely on a competitive exam. To boost minority enrollment, the New York public school district runs the Specialized High School Institute (SHSI), a five-week summer training program for economically disadvantaged students enrolled in grades 6-8.

Related Work

As far as we know, ours is the first rigorous analysis of achievement effects at highly selective U.S. exam schools, but selective high schools have been studied elsewhere. Pop-Eleches and Urquiola (2010) estimate the effects of attending selective high schools in Romania, where the admissions process is similar to that used by Boston's exam schools. Selective Romanian high schools appear to boost scores on the high-stakes Romanian Baccalaureate test. Jackson (2010) similarly reports large score gains for those attending a selective school in Trinidad and Tobago.

[5] These numbers are computed using data from the 2007-08 New York State School Report Cards.

[6] These averages are weighted by number of teachers per school as of October 2008, tabulated from the file used by Rockoff and Herrmann (2010).

[7] Although preferences officially ended in 1999, the race of 7th grade applicants does not appear to influence school assignment by 1997, the year after the 1996 challenge.

On the other hand, despite the huge peer advantage enjoyed by selective school students in the UK, Clarke (2008) uses RD to show that these schools generate modest score gains at most. Likewise, using admissions lotteries to analyze the consequences of selective middle school attendance in China, Zhang (2010) finds no achievement gains for students randomly offered seats at a selective school.

A number of American studies overlap with ours as well. The closest is probably Bui, Craig, and Imberman (2011), who report RD estimates of the the impact of gifted and talented (GT) services on student outcomes in regular public schools in a large urban district, as well as lottery-based estimates of the effects of attendance at a GT magnet school in this district. They find little GT impact. Likewise, Cullen, Jacob, and Levitt (2006) use admissions lotteries to show that randomly assigned opportunities to transfer to higher-scoring high schools in Chicago do not appear to boost scores. Chicago magnet schools are not exam schools, though Chicago now has nine of these. In a related paper, Cullen and Jacob (2008) estimate the effects of attendance at Chicago GT programs in public elementary schools and similarly find no achievement effects.

Elite education is perhaps more pervasive in American higher education than at the secondary level. Dale and Krueger (2002) compare students who applied to and were rejected by comparable sets of colleges. Perhaps surprisingly, this comparison shows no earnings advantage for those who went to more selective schools, with the possible exceptions of minority and first-generation college applicants in more recent data (Dale and Krueger, 2011). In contrast with the Dale and Krueger results, Hoekstra (2009) reports that graduates of a state university's (relatively selective) flagship campus earn more later on than those who went elsewhere.

Finally, a large literature looks at peer effects in educational settings. Examples include Angrist and Lang (2004), Hoxby and Weingarth (2006), and Lavy, Silva, and Weinhardt (2009). Findings in the education peer effects literature are mixed and not easily summarized. It seems fair to say, however, that the potential for omitted variables bias in naive estimates motivates much of the econometric agenda in this context. Economists have also studied tracking. A recent randomized evaluation from Kenya looks at tracking as well as peer effects, finding gains from the former but little evidence of the latter (Duflo, Dupas, and Kremer, 2010).

3 Boston Data and Descriptive Statistics

We obtained registration and demographic information for Boston Public School (BPS) students from 1997-2009. BPS registration data is used to determine whether and for how many years a student was enrolled at a Boston exam school. Demographic information in the BPS file includes race, sex, subsidized lunch status, limited English proficiency status, and special education status.

BPS demographic and registration information were merged with Massachusetts Comprehensive Assessment System (MCAS) scores using the BPS student ID. MCAS test are administered each spring, typically in grades 3-8 and 10. The MCAS database contains raw scores for Math, English Language Arts (ELA), Writing, and Science. The current testing regime covers Math and English in grade 7, 8, and 10 (in earlier years, there were fewer tests.) Baseline (i.e., pre-application) scores for grade 7 applicants are from 4th grade MCAS exams. Baseline English scores for 9th grade applicants come from 8th grade Math and 7th grade English (the 8th grade English exam was introduced in 2006.) We lose some applicants with missing baseline scores. For the purposes of our analysis, scores were standardized by subject, grade, and year to have mean zero and unit variance in the BPS student population.

Our analysis file combines the student registration and MCAS files with the BPS exam school applicant file. This file contains applicants' BPS ID, grade, year, sending school, ranking of exam schools, Independent Schools Entrance Exam (ISEE) test schools, and each exam school's ranking of applicants.

The study sample includes BPS-enrolled students who applied for exam school seats in 7th grade from 1997-2008 or in 9th grade from 2001-2007. We focus on applicants enrolled in BPS at the time of application because we're interested in how an exam school education compares to a traditional BPS education. Private school applicants are much more likely to remain outside the BPS district and hence out of our sample if they fail to get an exam school offer (about 45% of Boston exam school applicants come from private schools). The 10% of applicants who apply to transfer from one exam school to another are also omitted. The data appendix gives a detailed explanation of our analysis file, along with more information on test and application timing.

3.1 Student Characteristics

Non-exam BPS students are mostly minority and poor enough to qualify for a subsidized lunch. Black and Hispanic students are somewhat under-represented among exam school applicants and students, but most exam school applicants are also poor. These statistics are reported in Table 2, which compares the demographic characteristics and baseline test scores of non-exam school BPS students with those of the exam school applicant sample.[8]

Not surprisingly, there are few special education students in an exam school, though many exam school applicants and students are classified as limited English proficient. Exam school applicants are clearly a self-selected group, with markedly higher baseline scores than other

[8]The sample here includes 6th and 8th graders who were enrolled in BPS and applying for admission in 7th and 9th grade. Data for grade 7 cover 1997-2008; data for grade 9 cover 2001-2007.

BPS students. For example, grade 7 applicants' 4th grade Math scores are almost 0.8σ higher than those of a typical BPS student. Offered students are even more positively selected, with a score gap of 1.4σ in Math and 1.3σ in English. Similarly large gaps emerge for 9th graders.

Finally, note that there are many more exam school seats in grade 7 than grade 9. As a result, the probability an applicant is offered a seat is much lower for 9th grade applicants.

3.2 Descriptive Estimates

To set the stage for the RD estimates, we begin with a descriptive regression analysis of the relation between the standardized MCAS scores of student i tested in year t, denoted by y_{it}, and measures of exam school exposure. Specifically, we report ordinary least squares (OLS) estimates of equations like

$$y_{it} = \alpha_t + \sum_j \delta_j d_{ij} + \gamma' X_i + \rho M_{it} + \lambda' I_i + \epsilon_{it}, \tag{1}$$

fit to the sample of exam school applicants. Here, α_t is a test year effect, δ_j is a control for the student application cohort (interaction of year and grade), and X_i is a vector of demographic variables that includes gender, race, and free lunch status. (Students with missing demographics are omitted.) The exam school mediator, M_{it}, is captured either by a dummy for exam school enrollment following the application year; by the number of years a student was enrolled in an exam school from application date to test date; or, motivated by the fact that exam school enrollment is associated with exposure to high-achieving peers, by the average baseline score of peers in the year after application. Some specifications also include I_i, a vector of four ISEE scores from verbal, quantitative, reading, and math subtests. The estimates were computed in samples pooling 7th and 9th grade applicants and all available MCAS test outcomes for each applicant.[9]

Models without ISEE controls generate large positive coefficients for each measure of exam school exposure. For example, the enrollment estimate for Math reported in column (1) of Table 3 is nearly 1.0σ, while that reported in column (7) for ELA is 0.8σ. The corresponding per-year estimates are 0.36σ and 0.31σ. Models for peer means generate estimates of about 0.6σ for both Math and ELA. Not surprisingly given the nature of exam-school selection, the inclusion of ISEE score controls in equation (1) reduces the estimated exam school exposure coefficients considerably. Estimates for enrollment and exam years with ISEE controls fall to about a third of the size of estimates without ISEE controls. The decline in peer mean coefficients with ISEE controls is even larger, though even when estimated with controls, the estimated peer effects

[9]The standard errors here and elsewhere are clustered by enrollment school and test year (and by student when there are multiple test outcomes per student).

are still substantial and statistically significant at $0.1 - 0.15\sigma$.

The sensitivity of exam school mediator coefficients to the inclusion of ISEE controls highlights the fact that a good part of the apparent exam school advantage reflects positive selection bias. On the other hand, even conditional on ISEE controls, exam school exposure is highly positively correlated with student achievement. In the next section, we turn to an RD framework to determine whether this correlation is causal.

4 Boston RD Framework

4.1 The Boston Admissions Process

Boston residents interested in an exam school seat take the ISEE in the fall of the school year before they would like to transfer. Applicants also submit an authorized GPA report that winter, based on grades through the most recent fall term. Finally, exam school applicants are required to rank up to three exam schools on the application form. The exam school composite score is a weighted average of applicants' Math and English GPA, along with scores on the four parts of the ISEE (verbal, quantitative, reading, and math). Applicants are admitted by composite-score rank order until all seats are filled.

For the purposes of the analysis here, composite scores were standardized separately for each school using the sample of applicants to that school, generating ranking variable R_{ik} for student i, specific to school k. A smaller value of R_{ik} indicates that the student has a higher composite score and is more likely to gain admission. We focus on those applying for seats in the 7th and 9th grades (O'Bryant also accepts a handful of 10th graders).

Applicants are ranked only for schools to which they've applied, so applicants with the same GPA and ISEE scores might be ranked somewhat differently at different schools depending on where they fall in each school's applicant pool. Schools admit students based on a cutoff, C_k, the largest rank to obtain an offer at that school. Applicants who rank more than one school, as many do, are offered a seat at the school they most prefer among those for which their school-specific rank clears the relevant cutoff. For the purposes of our figures and empirical work, we scaled school-specific composite ranks according to:

$$r_{ik} = 100 \times \frac{R_{ik} - C_k}{\max_{j \in \mathcal{I}_k}\{R_{jk}\} - \min_{j \in \mathcal{I}_k}\{R_{jk}\}},$$

where $\mathcal{I}_k$ are students who ranked school k. Scaled school-specific ranks provide a running variable that equals zero at the cutoff for school k, with positive values indicating students who applied to and qualified for admission at that school.

4.2 Discontinuities in offers, enrollment, and peers

The exam school admissions process generates large discontinuities in the relation between r_{ik} and the probability of an exam school offer, with somewhat more modest though still substantial jumps in enrollment. This can be seen in Figures 1-3 for 7th grade applicants. Panels in the figures cover a scaled rank interval of [-20,+20] for each of the three Boston exam schools. Applicants outside the 20-unit band are either far below or well beyond the relevant cutoffs. Plotted points are conditional means for all applicants in a one-unit binwidth similar to the empirical conditional mean functions reported in Lee, Moretti, and Butler (2004).

The figures also show smoothed conditional mean functions allowing for jumps at each cutoff. Specifically, for school k, we construct local linear regression (LLR) estimates of $\hat{\mathrm{E}}[y_i|r_{ik}]$, where y_i is the dependent variable and r_{ik} is the running variable. The LLR smoother uses the edge kernel:

$$K(u_{ik}) = \mathbf{1}_{|u_{ik}|\leq 1}(1 - |u_{ik}|),$$

where $u_{ik} = \frac{r_{ik}}{h}$ and h is the bandwidth. In an RD context, LLR has been shown to produce estimates with good properties at boundary points (Hahn, Todd, and van der Klaauw (2001) and Porter (2003)). The bandwidth used here is a version of that proposed by Imbens and Kalyanaraman (2010) (hereafter, IK) who derive optimal bandwidths for sharp RD using a mean square-error loss function with a regularization adjustment. The jump in the conditional mean function at the cutoff is the IK sharp RD estimate of the effect of an offer at a particular school.[10]

Figure 1 captures key elements of the relation between running variables and school-specific enrollment rates. In each panel, the dark line plots the offer rate at the school for which the panel is labeled, while the dotted line is the probability of an offer at *other* exam schools. For example, the leftmost panel in Figure 1 shows that students who score just above the O'Bryant cutoff obtain an offer at the O'Bryant school with near certainty. But O'Bryant applicants who ranked another exam school ahead of O'Bryant may be offered a seat at this school instead. Hence, the O'Bryant panel also shows an increasing probability of admission to other exam schools as we move right from the O'Bryant cutoff. The center panel, for Latin Academy, shows very high probabilities of receiving an exam school offer (in this case, from O'Bryant) for those to the left of, but close to, the Latin Academy cutoff. Finally, almost everyone to the left of the Latin School cutoff gets an offer from another exam school, while to the right of the Latin School cutoff, offer rates at Latin School jump from 0 to about 0.75.

Figure 2 plots the relation between scaled ranks and exam school *enrollment* instead of offers. Applicants scoring just above admissions cutoffs are much more likely to enroll in a given school

[10] Our implementation of IK is discussed in detail in the next section.

than are those just below the cutoffs. On the other hand, enrollment rates at other schools also change around each school-specific cutoff. Figure 3 puts these pieces together by plotting jumps in the probability of enrollment in *any* exam school around each school-specific cutoff (this is the sum of the dark and dotted lines in Figure 2). Exam school enrollment jumps at the O'Bryant and Latin Academy cutoffs, but changes little at the Latin School cutoff because those to the left are very likely to enroll in either O'Bryant or Latin Academy.

Much like the relation for enrollment, Figure 4 shows that students with normalized composite-score ranks that clear school-specific cutoffs spend more time enrolled in exam schools. Although some applicants to the left of the O'Bryant cutoff eventually accumulate enrollment years by applying again in grade 9, those to the right of the O'Bryant cutoff spend about two years more in exam schools than those to the left. The discontinuity in exam school years is less pronounced at the Latin Academy cutoff, and there is no jump in years at the Latin School cutoff. This reflects the fact that students who come close to, but fail to clear, the Latin Academy and Latin School cutoffs, almost certainly get offers at the next school down in the Boston exam school hierarchy.[11]

An important component of the exam school experience is exposure to other high-achieving students. Figures 5 and 6 document this by plotting the average baseline score of peers for applicants on either side of admissions cutoffs.[12] Baseline peer means jump by about half a standard deviation at each admissions cutoff. This implies that (conditional on applying to an exam school) peers at Latin Academy are ahead of non-exam BPS peers by a full standard deviation, while peers at Latin School are ahead of non-exam BPS peers by about 1.5σ.

Although not shown here, offer and enrollment patterns for grade 9 applicants are similar to those shown here for grade 7 applicants. The grade 9 sample is much smaller, however, especially for Boston Latin Academy and Boston Latin School, which together account for only a quarter of 9th grade seats. Enrollment and peer discontinuities in the O'Bryant 9th grade sample look much like those for O'Bryant's 7th graders.

[11]Note that the total years variable plotted in this figure reflects the maximum exam school exposure available to the applicant cohorts in our data. For instance, 7th grade applicants who applied in 2006 will have spent at most two years in an exam school by the time we see them tested at the end of 8th grade, while our sampling window closes before we get a chance to see them tested in 10th.

[12]The peer mean score is the average baseline score of same-grade peers in the school in which an applicant enrolled in the year following the year of exam-school application.

5 Boston RD Estimates

5.1 Econometric Framework and Reduced Form Estimates

We constructed parametric and non-parametric RD estimates of the effect of an exam school offer using the normalized composite score as the running variable. We refer to this initial set of estimates as "reduced form" because these estimates capture the effect of an exam school offer, without adjustment for the relationship between offers and enrollment or other mediating variables. As in the plots, the Boston empirical work is limited to sets of applicants with school-specific running variables in the interval [-20,+20]. Applicants outside this window are well below or well above the relevant cutoffs. At the same time, the [-20,+20] window is wide enough to allow for reasonably precise inference using Boston applicant data.

The parametric estimating equation for applicants to school k is

$$\begin{aligned} y_{itk} = \alpha_{tk} + \sum_j \delta_{jk} d_{ij} &+ (1 - D_{ik}) f_{0kk}(r_{ik}) + D_{ik} f_{1kk}(r_{ik}) + \rho_k D_{ik} \\ &+ (1 - D_{il}) f_{0kl}(r_{ik}) + D_{il} f_{1kl}(r_{ik}) + \lambda_k D_{il} \\ &+ (1 - D_{im}) f_{0km}(r_{ik}) + D_{im} f_{1km}(r_{ik}) + \mu_k D_{im} + \eta_{itk}, \end{aligned} \tag{2}$$

where the variable D_{ik} is an indicator for $r_{ik} \geq 0$ and the coefficient of interest is ρ_k. The estimate of ρ_k captures the effect of an offer at school k, relative to the counterfactual of no offer at that school, for applicants in the Boston window.

Equation (2) controls for test year effects at school k, denoted α_{tk}, and for the year and grade of application, indicated by dummies, d_{ij}. Effects of the running variable at school k are controlled by the first of three pairs of third-order polynomials that differ on either side of the cutoff (a condition indicated by j), specifically

$$f_{jkk}(r_{ik}) = \pi_{jkk} r_{ik} + \xi_{jkk} r_{ik}^2 + \psi_{jkk} r_{ik}^3; \quad j = 0, 1.$$

The two additional running variable polynomials in equation (2) are introduced to control for the fact that applicants to school k typically apply to more than one school and may have (or lose) other exam school options. For example, highly qualified Latin Academy applicants also qualify for admission to Latin School. Looking in the other direction, while many and perhaps most applicants to Latin Academy also qualify for admission to O'Bryant, poorly qualified applicants to Latin Academy do not. In small neighborhoods around each cutoff, confounding from other offers disappears, but in the empirical Boston window nearby cutoffs determine counterfactual outcomes and are a potential source of nonlinearity in the relation between running variables and outcomes. The effects of other offers are captured in equation (2) by including dummies for non-k cutoffs in the model for school k; these other schools are denoted school l and school m, with offer-cutoff effects λ_k and μ_k in the model for school k, and running variable controls in the parametric model given by $f_{jkl}(r_{ik})$ and $f_{jkm}(r_{ik})$, where

$$f_{jkl}(r_{ik}) = \pi_{jkl} r_{ik} + \xi_{jkl} r_{ik}^2 + \psi_{jkl} r_{ik}^3;$$
$$f_{jkm}(r_{ik}) = \pi_{jkm} r_{ik} + \xi_{jkm} r_{ik}^2 + \psi_{jkm} r_{ik}^3; \quad j = 0, 1.$$

Note that the subscript j in these polynomials captures the fact that they interact with *other* school cutoff indicators, D_{il} and D_{im}. In this scheme, every combination of admissions offers is associated with a unique set of polynomial terms.

Much of the appeal of RD comes from a parallel with randomized trials: near the relevant admissions cutoff, exam school offers can be taken to be as good as randomly assigned. Why, then, is our estimating equation so complicated? In practice, the need for extensive controls is generated by the fact that we must borrow information away from the cutoff if estimates near the cutoff are to be precise enough to be useful. As always, the inclusion of data away from the cutoff risks bias from omitted secular running variable effects. But in the Boston admissions problem, we must also worry about confounding effects of multiple cutoffs.

Figure 7 presents a stylized representation of the Boston admissions process that motivates equation (2). The bottom of the figure sketches own-school and other-school offers. At the cutoff for school k, offers at school k naturally jump, but as we move to the right, another offer is made, while moving to the left, the next offer down falls away. The changing pattern of offers is reflected in the reduced form relation between achievement and the school k running variable, introducing additional curvature and potentially even corners or jumps in the reduced form relation sketched at the top of the figure. Equation (2) allows us to distinguish this extra curvature and nonlinearity from the causal effects of admission at school k.

Non-parametric RD

Non-parametric estimates differ from parametric in three ways. First, they narrow the Boston window when the optimal data-driven IK bandwidth falls below 20.[13] Second, the non-parametric estimates use a tent-shaped edge kernel centered at admissions cutoffs instead of the uniform kernel implicit in parametric estimation. Finally, non-parametric models control for linear functions of the running variable only. We can write the non-parametric estimating

[13]The IK bandwidths for Table 4 range from about 8 to 26.

equation as

$$y_{itk} = \alpha_{tk} + \sum_j \delta_{jk} d_{ij} + \gamma_{0kk}(1 - D_{ik})r_{ik} + \gamma_{1kk} D_{ik} r_{ik} + \rho_k D_{ik} \tag{3}$$
$$+ \gamma_{0kl}(1 - D_{il})r_{ik} + \gamma_{1kl} D_{il} r_{ik} + \lambda_k D_{il}$$
$$+ \gamma_{0km}(1 - D_{im})r_{ik} + \gamma_{1km} D_{im} r_{ik} + \mu_k D_{im} + \eta_{itk},$$

for each of the three schools indexed by k. Non-parametric RD estimates come from a kernel-weighted least square fit of equation (3).

Figures 8-11 show non-parametric RD reduced forms for middle school (7th and 8th grade) and high school (10th grade) Math and English. Dots in the plots are averages in a one-unit binwidth, while lines are from the local linear smoother using IK bandwidth. Jumps in the smoothed scores at the admissions cutoff are the IK sharp regression discontinuity estimates of the effects of qualifying for an exam school offer on test scores. Except perhaps for 10th grade English, the plots offer little evidence of marked discontinuities in MCAS scores at any of the three admissions cutoffs.

Not surprisingly, the single-school reduced form score estimates, reported in Table 4, tell the same story as the figures. Few of these estimates are significantly different from zero and some of the significant effects at Latin School are negative (for example, Latin School effects on 10th grade Math and middle school English). Most of the estimates are small and some are precise enough to support a conclusion of no effect.

Stacking Schools

In an effort to increase precision, we also constructed estimates pooling applicants to all three of Boston's exam schools. The pooled estimating equations are essentially the same as equations (2) and (3), but with a single offer effect, ρ. Because the pooled model is saturated with a full set of main effects and interactions for school-specific subsamples, we can think of the estimate of ρ in this stack as a variance-of-treatment-weighted average of school-specific estimates.[14] Note that some students apply to more than one school and a given student may contribute up to three observations, even for a single outcome. Our inference framework takes account of this by clustering by student.[15]

Paralleling the pattern shown in the Boston reduced form figures, the estimated reduced-

[14] Variance-weighting is a property of models with saturated regression controls; see, e.g., Angrist (1998). Not quite literally in this case, however, since the model here is not fully non-parametric.

[15] An alternative stacking scheme partitions applicants according to the school they are most likely to get into. For most applicants, however, this is the O'Bryant school. As a result, the resulting stacked estimates look much like the O'Bryant estimates.

form offer effects from the stacked models, reported in columns labeled "All Schools" in Table 4, are mostly small, with few significantly different from zero. One substantial and significant positive effect, for 10th grade English scores, seems to stand out as it appears at individual schools, and in both parametric and non-parametric estimates. On the other hand, this positive finding is partly offset by a marginally significant negative effect on 7th and 8th grade English, so that when all scores are stacked and pooled the overall estimated impact is close to zero (scores are stacked in much the same way that schools are stacked). Two other marginally significant IK estimates for Math are also negative, as is the estimate for 7th grade ELA. Importantly, the combination of school- and score-pooling generates precise estimates, with standard errors on the order of 0.025 for both Math and ELA.

Appendix A reports results from an exploration of possible threats to a causal interpretation of the reduced form estimates in Table 4. Specifically we look for differential attrition (i.e., missing score data) to the right and left of exam school cutoffs and for discontinuities in covariates. There is some evidence that receipt of an exam school offer makes attrition somewhat less likely, but the gaps are small and unlikely to impart substantial selection bias in estimates that ignore them. A few covariate contrasts also pop up as significantly different from zero, but the spotty nature of these gaps, and the fact that the parametric and non-parametric findings are similar, support the notion that our controlled comparisons to the left and right of exam school admissions cutoffs are indeed a good experiment.

A related threat to validity comes from the possibility that marginal students switch out of exam schools at an unusually high rate. If school switching is harmful, excess switching might account for findings showing little in the way of score gains. As it turns out, however, exam school applicants who clear admissions cutoffs are more likely to stay at an assigned school through grade 12 than are traditional BPS students. This partly reflects the high rate of student turnover in Boston high schools – overall enrollment persistence in BPS first-choice high schools is only about 0.32 – a mobility pattern typical of American inner-city schools. The probability that a traditional BPS 7th grader in the Boston window enrolls in the same school as a senior is about 0.48; for 9th graders in the Boston window (mostly applying to O'Bryant), the re-enrollment rate falls to 0.35. Exam school offers *increase* enrollment persistence by 0.12 for 7th grade applicants and by 0.29 for 9th grade applicants.[16] This increase weighs against the view that unusually high exit rates from exam schools account for the findings reported here.

[16] These estimates come from a parametric reduced form analysis similar to that used to construct the covariate balance and attrition estimates in the appendix.

5.2 Reduced-Form Estimates for Subgroups

High Achievers

RD captures causal effects for students near exam school admissions cutoffs. It's worth emphasizing, however, that the three cutoffs in our sample cover a wide range of ability. Among 7th grade applicants, the cutoffs fall near the median of the ISEE distribution for O'Bryant to about the 75th quantile for Latin School, with the Latin Academy cutoff falling in the middle. Among 9th grade applicants, the O'Bryant cutoff falls near the 60th quantile, the Latin Academy cutoff is near the 87th quantile, and the Latin School cutoff is at about the 92nd quantile. It's impressive that the results are reasonably consistent across applicants to these three schools.

To further explore consistency across quantiles of the applicant ability distribution, we exploit the fact any single test is necessarily a noisy measure of ability. Although we can't construct RD estimates for, say, O'Bryant students with ISEE scores in the uppermost tail of the score distribution, we can look separately at subsamples of students with especially high baseline MCAS scores. Some in the high-baseline group are ultra-high achievers who landed in a marginal ISEE group by chance.

The average baseline score for students in the upper half of the baseline MCAS distribution hovers around $1.2 - 1.4\sigma$ in both Math and English. Importantly, MCAS scores remain informative even for these high achievers: among middle schoolers in the upper half of the applicant distribution, only 28% test at the Advanced Level for Math, while 12% are at the Advanced level for English. Among 10th graders, 76% test at Advanced in Math and 34% test at Advanced in English. A similar tabulation shows the MCAS to be informative even for applicants in the upper MCAS quartile. It's therefore of interest to see what RD estimates of exam school applicants look like for students with such high baseline MCAS scores, an inquiry made possible by the fact that some of these applicants have ISEE scores close to admissions cutoffs.

Perhaps surprisingly, RD estimates for applicants in the upper half and upper quartile of the baseline score distribution come out essentially similar to those for the full sample. These results, reported in Table 5, are mostly negative with few significantly different from zero with the exception being significant positive effects on 10th grade ELA. At the same time, the sample of high achievers generates a significant negative estimate of effects on middle school ELA – an effect of roughly the same magnitude as the positive ELA estimate for 10th graders. Thus, even in a sample of ultra high achievers, there is little evidence of a consistent exam school boost.

Minorities

Our interest in exam school effects on minority applicants is motivated in part by the contentious debate over minority representation in these schools. Is the fight over minority representation justified by evidence of achievement gains for minorities? In an investigation of the earnings consequences of attendance at selective colleges and universities, Dale and Krueger (2002, 2011) find no overall effect. At the same time, the Dale and Krueger estimates show some evidence of gains for minority applicants. The influential book-length analysis of minority admissions preferences at selective colleges and universities by Bowen and Bok (2000) also marshals a variety of evidence in support of the same point.

Our estimates for black and Hispanic applicants to exam schools, also reported in Table 5, are in line with the full-sample findings for Math and middle-school ELA scores. On the other hand, consistent with the full-sample results for 10th grade ELA, an exam school education seems especially likely to boost 10th grade English scores for minorities, with an estimated effect of 0.18σ. In fact, the full-sample ELA results appear to be driven primarily by the minority impact, since the (unreported) corresponding IK estimate for whites comes out at an insignificant 0.05σ (se=0.042). Blacks and Hispanics are a small share of exam school students, however, even in the heavily minority Boston and New York districts. We therefore return to the full sample, beginning with an investigation of causal mediators.

5.3 2SLS (Fuzzy RD) Estimates of Mediating Causal Effects

Exam school offers might affect achievement in a number of ways, most immediately through exam school enrollment. We can also think of exam school offer effects as being mediated by time spent attending an exam school, a measure of the intensity of educational tracking. Finally, we consider the possibility that the most important mediator for exam school offers is peer achievement.

Our investigation of exam school mediators uses exam school offers as instruments for mediating variables in a fuzzy RD analysis. This allows us to explore, for example, what the combination of a strong peer mean first stage and a small reduced form impact implies about the size of peer effects. Although we can't say for sure whether any single mediator satisfies an IV exclusion restriction, the bias from failures of the exclusion seems likely to be positive, so the resulting IV estimates can be thought of as providing an upper bound on one-at-a-time causal effects. Also relevant is the precision of the 2SLS estimates: among other things, this tells us whether we can reject positive peer effects of the size reported elsewhere.

Fuzzy RD is implemented here using two-stage least squares (2SLS). The 2SLS setup parallels that used for pooled reduced form estimation (pooling applicant grades and test years, as well as schools). Because the non-parametric analysis generates somewhat more precise estimates than the parametric, we focus here on IK estimates for the pooled sample. The second stage equation in this context is similar to the stacked reduced form based on equation (3), except that the three own-school cutoff dummies are excluded and used as instruments for mediating variables, M_{it}. To economize on notation, we write the 2SLS second stage by subsuming all controls, including year of test, grade, and application effects, and own- and other-school running variable controls, in a vector X_{itk} with conformable coefficient vector Γ_k. We can then write the second stage equation as

$$y_{itk} = \Gamma_k' X_{itk} + \theta M_{it} + \epsilon_{itk}, \tag{4}$$

where M_{it} is the endogenous variable to be instrumented and θ is the causal effect of interest. The corresponding first stage equations include these same controls using three own-school offer dummies as instruments, one for each set of applicants, stacked as when estimating equation (3). In principle, three instruments is enough to estimate the effects of three endogenous variables at the same time, but in practice this doesn't produce informative results. As a result, we estimate the effects of mediating variables one at a time. The mediators considered here are either an enrollment dummy ("in exam school the school year after application date"), years enrolled in an exam school between application and test date, and the applicant's baseline peer mean as experienced in the school year following exam school application. In an effort to increase precision, we also computed 2SLS estimates adding interactions between offer dummies and application cohort (year and grade).[17]

Applicants with a score above the O'Bryant cutoff are 73 percentage points more likely to enroll in an exam school and have spent about 1.6 more years at an exam school by the time they take an MCAS test. These and other first stage estimates are reported in Table 6. The first stage effects of the Latin Academy exam cutoff indicators on enrollment and years at an exam school are smaller than the O'Bryant effects because many who just miss a Latin Academy offer end up in O'Bryant. The corresponding first stage estimates for a Latin School offer are small and not significantly different from zero, a consequence of the fact that almost all near misses at the Latin School end up at Latin Academy.

2SLS estimates of the effect of exam school enrollment or years of attendance are small, with none significantly different from zero (estimates for Math are negative). The addition of cohort interactions to the instrument list generates only slight precision gains, but the estimates are reasonably precise either way. It's especially noteworthy that these estimates are precise enough to be statistically distinguishable from the corresponding OLS estimates in Table 3, whether

[17]Paralleling the reduced form setup, the own-school cutoff is an instrument for mediators in the sample of applicants from that school, while other school cutoffs are included as controls.

the latter are estimated in models with or without ISEE controls. Compare, for example, the 2SLS estimates of the effect of exam years on Math (-0.017 with standard error of 0.030), to the OLS estimate of 0.088 with standard error 0.016 in a model with ISEE controls.

The first stage estimates reveal large and precisely estimated impacts of exam school offers on applicants' peer achievement. O'Bryant offers increase average baseline peer scores by over two-thirds of a standard deviation, while the gain is about 0.4σ at the Latin Academy cutoff, and $0.53 - 0.62\sigma$ at the Latin School cutoff. Consistent with the reduced form estimates, however, the 2SLS estimates show no significant effects of peer achievement on applicant achievement. An important piece of information in this context is the precision of the 2SLS estimates of peer effects, which come out significantly different from the large positive OLS estimates in Table 3. The estimated peer-effect zeros in Table 6 are also significantly different from many of the positive education peer effects reported elsewhere as well; see, e.g., Sacerdote (2001), who estimates college freshman GPA peer effects on the order of 0.12.

The Wrong Pond

Our investigation of peer effects is motivated by econometric research predicated on the hypothesis that better peers boost achievement. At the same time, a parallel literature originating in educational psychology explores the apparently contradictory hypothesis that high-achieving peers are demoralizing and reduce achievement, at least for those not as strong. Marsh, Chessor, Craven, and Roche (1995) and Bui, Craig, and Imberman (2011) reference this "Big Fish Little Pond Effect" (BFLPE) as a possible explanation for the failure to find achievement gains in gifted and talented programs. Here, BFLPE might explain the mostly weak effects of an exam school education since marginal admitted applicants, though positively selected relative to where they're coming from, will typically not be at the top of an exam school class.

The flip side of the peer first stage documented in Table 6 is indeed a decline in students' percentile rank among peers. This is documented in Figures 12 and 13, which plot applicants position in the baseline Math and English score distributions. The plots show sharp drops at admissions cutoffs, essentially the mirror image of the peer first stage reported in Table 6. On the other hand, while applicants just above the cutoff at the O'Bryant school necessarily have the lowest ISEE/GPA composite score among all those offered an O'Bryant seat, their baseline scores place them in the middle of the baseline distribution among those offered a seat. This is a decline from about the 75th percentile of the baseline test score among non-offered peers. The baseline score ranking of marginal Latin Academy and Latin School applicants fall similarly though somewhat less sharply over a range in the middle of the relevant distributions.

We investigate BFLPE more formally by allowing for interactions between exam school enrollment and the difference between applicant achievement and those of peers at the targeted exam school. The idea here is to check the BFLPE prediction that students who enroll in exam schools where they can expect to be substantially weaker than classmates gain less or lose more than those with baseline achievement at the peer mean or better. To formalize this, let b_i denote applicant i's baseline score and $\bar{b}_{(i)k}$ be the the peer mean at an applicant's target exam school. The potential peer gap at the applicant's target school is

$$g_{ik} = (b_i - \bar{b}_{(i)k}).$$

Adding peer gap interactions to a second-stage equation that captures causal effects of exam school enrollment, E_{it}, we have

$$y_{itk} = \Gamma_k' X_{itk} + \theta_0 E_{it} + \theta_1 E_{it} g_{ik} + \epsilon_{itk}, \tag{5}$$

with two endogenous variables: E_{it} (exam school enrollment) and $E_{it}g_{ik}$ (exam school enrollment interacted with the potential peer gap). The instruments in this case are offer cutoff indicators, D_{ik}, and the cutoff indicator times the potential peer gap, $D_{ik}g_{ik}$.

We estimated equation (5) in the pooled sample of 7th and 9th grade applicants with stacked schools and pooled MCAS test outcomes. This generates 2SLS estimates of θ_1 of about -0.12 (se = 0.045) for Math and -0.15 (se = 0.044) for ELA.[18] The fact that the impact of exam school enrollment seems to be decreasing in the gap between an applicant's baseline ability and that of his peers at the targeted schools suggests the BFLPE mechanism plays little or no role in mediating the overall exam school impacts reported here.

5.4 Other Boston Outcomes

PSAT and SAT Scores

MCAS scores are measures of achievement that, with the exception of the 10th grade test that also serves an as an exit exam, are only indirectly linked to ultimate educational attainment. We also look at additional test outcomes that are highly correlated with MCAS scores, but may be more important.[19] The first of these is the Preliminary SAT/National Merit Scholarship Qualifying Test (PSAT/NMSQT), which serves as a warmup for the SAT and is used in the National Merit scholarship program; the second is the SAT Reasoning Test itself (formerly, the Scholastic Aptitude Test).

[18]These are IK estimates in models that control for a potential peer gap main effect scores. Models swapping the potential peer gap with baseline scores generate virtually identical results since the leave-one-out peer mean is essentially a linear combination of the complete peer mean and the individual baseline score.

[19]The correlation between 10th grade MCAS Math and PSAT or SAT is about 0.7; the correlation for English is similar. These estimates come from models with the same controls as in Table 3 (without ISEE scores).

SAT and PSAT tests are usually taken towards the end of high school, so scores are unavailable for the youngest applicant cohorts in our sample (appendix Table C2 lists included cohorts). In March 2005, the College Board added a writing section to the SAT. Since the writing section does not appear in earlier years, we focus on the sum of Critical Reading (Verbal) and Mathematics scores, also known as the SAT Reasoning score. As with MCAS outcomes, SAT and PSAT are standardized to have mean zero and unit variance among all test-takers in a given year.[20] The average PSAT score for applicants in the Boston window (Critical Reading and Math) is 91.3, while the average SAT score is 1019. These can be compared with 2010 national average PSAT and SAT scores of 94 and 1017.

Consistent with the MCAS results discussed above, Figures 14 and 15 show no apparent discontinuity in either PSAT or SAT scores for students near admissions cutoffs. The points plotted in these two figures are necessarily for students who took the PSAT or SAT. We report reduced form offer effects for the probability of taking tests as well as for scores conditional on taking. These results are reported in Table 7 for both tests.

Exam school offers increase the likelihood of taking the PSAT by about five percentage points overall, with the boost coming mostly near the O'Bryant cutoff. PSAT scores among takers appear to be mostly unaffected by exam school offers. For a subsample of black and Hispanic applicants, the impact on PSAT participation is 0.08 (se = 0.03). Increased PSAT-taking does not seem to have had much of an effect on SAT-taking, where the estimates are almost all small and insignificant. Selection bias in the sample of test takers seems likely to be second order, especially for SATs.

As with the MCAS, the SAT results here offer little in the evidence of consistent score gains. The estimated SAT score effects point to a gain for 9th grade applicants SAT scores, emerging most strongly at the Latin Academy and Latin School. 9th grade applicants are a minority of the exam school population; when pooled with the sample of 7th grade applicants, this effect fades considerably. Likewise, estimates for black and Hispanic students suggest an SAT boost in 9th grade at two schools (0.30 with se=0.12 at Latin Academy and 0.23 with se=0.12 at Latin School), but the overall SAT effect for minority students is zero, with se=0.04.

AP Tests

Motivated by the prevalence of AP courses in the Boston exam school curriculum, we estimated exam school effects on AP taking and AP scores. As with the PSAT/SAT analysis, younger cohorts are excluded since these tests are usually taken in grades 11-12 (again, appendix Table C2 gives details).

AP tests are scored on a scale of 1-5 and some colleges grant credit for some AP subjects

[20] We use the first score recorded in the Boston Public Schools SAT and PSAT files.

in which an applicant scores at least 3 or 4. At the high end, Latin School students take an average of three to four AP exams. At the same time, Figure 16 shows little evidence of a jump in the number of tests taken at any exam school cutoff. Figure 17 looks at the sum of AP scores, awarding zeros for tests not taken. This figure also offers no evidence that exam schools increase AP success rates.

The estimates that go with Figures 16 and 17 appear in Table 8. Here we look at sums and scores for all AP exams, as well as for a subset of the most popular exams, defined as those taken by at least 500 students in our BPS score file. This restriction narrows the set of exams to include subjects like math, science, english, history, and economics, but omits music and art.[21] Overall there is little impact on the number of AP exams taken. In the analysis of effects on popular tests, a positive effect on the sum of scores for O'Bryant 7th grade applicants is offset by negative effects at the other two schools. On balance, therefore, it seems unlikely that exam school enrollment improves AP-related outcomes. Results for black and Hispanic applicants support a similar conclusion in this case.

6 New York Estimates

Results for New York are presented here in a format much like that used for Boston, though more briefly. We focus on the three oldest academic New York exam schools: Brooklyn Technical, Bronx Science, and Stuyvesant.

Data from New York City comes from three sources: enrollment and registration files containing demographic information and attendance records; application and assignment files; and the Regents exam file. Our analysis covers four 9th grade applicant cohorts (from 2004-2007), with follow up test score information through 2009. The data appendix explains how these files were processed in detail.

6.1 New York Admissions

The New York exam school admissions process is simpler than the Boston process because selection is based solely on performance on the Specialized High School Achievement Test (SHSAT), whereas Boston schools rely on school-specific composites. New York 8th graders interested in an exam school seat take the SHSAT and submit an application listing school preferences (we omit a handful of 9th grade applicants). Students are ordered by SHSAT scores. Seats are then allocated down this ranking, with the top scorer getting his first choice,

[21] Tests with at least 500 takers are Calculus AB, Statistics, Biology, Chemistry, Physics B, English Language and Composition, English Literature and Composition, European History, US Government and Politics, US History, Microeconomics, Macroeconomics, and Spanish Language.

the second highest scorer get his most preferred choice among schools with remaining seats, and so on.[22]

As in Boston, we standardized and centered the running variable for each school. Let C_k denote the minimum rank score needed for a seat at school k and let R_i denote the rank of student i. Stuyvesant is the most competitive exam school, so the minimum score needed to obtain an offer exceeds the minimum at Bronx Science and Brooklyn Technical. We construct a school-specific running variable as

$$r_{ik} = 100 \times \frac{R_i - C_k}{\max_{j \in \mathcal{I}_k}\{R_j\} - \min_{j \in \mathcal{I}_k}\{R_j\}},$$

where $\mathcal{I}_k$ are students who ranked school k. These normalized running variables equal zero at each cutoff, with positive values indicating those who obtain an offer. Also as in Boston, applicants can qualify for placement at one school, but rank a less competitive school first and get an offer at that school instead. Note that New York admissions are based on a single underlying running variable, while school-specific running variables in Boston are correlated but distinct. New York cutoffs are typically separated by six standardized rank units using the formula above.

The descriptive statistics in Table 9 show that New York exam school applicants are positively selected relative to the population of New York 8th graders. Applicants' baseline scores exceed those of other 8th graders by about $0.7 - 0.8\sigma$, while the score gap for offered students is $1.7 - 1.8\sigma$. Exam school applicants reflect the New York public school population in that a substantial fraction are eligible for a subsidized lunch. In contrast to Boston, however, only about 15% of New York's offered students are black or Hispanic.

In (unreported) OLS models paralleling those used to construct the estimates in Table 3, exam school exposure in New York is associated with a large achievement advantage, whether exposure is measured by enrollment, exam years, or peer means. Most of these estimates remain substantial even after controlling for SHSAT scores. For example, in models without SHSAT controls, the peer mean coefficient for Advanced Math is 0.47σ. This falls to 0.17σ with a linear control for SHSAT scores, still a large and precisely estimated effect.

6.2 RD Plots and Estimates

The estimation window for each of the New York schools is set at $[+5, -5]$. The New York window is narrower than the Boston window of +/-20 because there are many more New York applicants, so the sample is much larger than Boston's even in a window one quarter the

[22] The NYC exam school assignment mechanism is a *serial dictatorship* with students ordered by the admissions test score. See Abdulkadiroğlu, Pathak, and Roth (2009) for a detailed description.

size, and because trimming at five greatly mitigates the problem of confounding from nearby admissions cutoffs (as described in Figure 7). As noted above, cutoffs are separated by about six standardized units - the separation in this case is clearer than for Boston because New York admissions rely on a single underlying running variable.

Figure 18 shows how New York offers are related to the running variable. Although the estimation sample is five units wide in each direction, the plot window runs for 10 units to show behavior outside the window. The dots in Figure 18 plot averages in half-unit bins, while the fitted lines in the plots are IK estimates using the bandwidth generated in the estimation sample. Own-school offers jump at each cutoff, but five or six points to the right of the Brooklyn Tech and Bronx Science cutoffs, offers at the next school up replace those at the target schools. (Offers at other schools remain positive just to the right of the Bronx Science cutoff because some Bronx Science applicants who qualify for admission there ranked Brooklyn Tech first, perhaps because they live in the neighborhood.)

Offers at each exam school lead to enrollment at that school, though the offer-to-enrollment conversion rate differs across schools. This pattern is documented in Figure 19. Enrollment jumps at the Brooklyn Tech and Bronx Science cutoffs are lower than the corresponding offer jumps, though both enrollment jumps remain substantial. The Stuyvesant enrollment jump is about as large as the offer jump, implying that nearly all offered a seat at Stuyvesant enroll there. This pattern is mirrored in a plot of average years of exam school exposure against the New York running variables, shown in Figure 20.[23]

New York has considerable school choice, with other selective schools outside the set of traditional exam schools. Admission to one of these schools is nevertheless associated with a sharp jump in peer achievement, as can be seen in Figures 21-22. The average baseline Math and English score of peers increases by about 0.5σ for Math and 0.4σ for English near the Brooklyn Tech cutoff. The jump is smaller for Bronx Science and Stuyvesant, though still substantial at about 0.2σ.

Although New York exam school exposure jumps at admissions cutoffs, there is little evidence of a corresponding discontinuity in achievement. This is apparent in Figures 23 and 24, which plot performance on the Advanced Math and English components of the New York Regents exam against the standardized New York running variables. These results are confirmed in Table 10, which reports parametric and IK reduced-form estimates of offer effects on Advanced Math and English as well as for other Regents test outcomes. The estimated equations are like equations (2) and (3) for applicants in the $[-5, +5]$ interval.[24]

The New York estimates are precise enough to rule out even modest score gains. For

[23]Exposure is capped at two or three years for the last two applicant cohorts since our registration files end before these students finish high school.

[24]The IK bandwidth in the Table 10 estimates ranges from 3-11.

example, the IK estimate of the effect on English in the stacked sample is 0.01σ, with a standard error also around 0.01. The few significant pooled estimates in Table 10 are negative.

Using models and estimation procedure similar to those used to construct 2SLS estimates of mediating effects for Boston, we computed 2SLS estimates of the effects of enrollment, exam years, and baseline peer means on student achievement for New York. These results, along with the associated first stage estimates, are reported in Table 11. Just as in Boston, offers at the first two NYC exam schools increase enrollment and time spent at an exam school, though first-stage enrollment effects are not significantly different from zero at Stuyvesant. Admission to any of these three schools generates a substantial jump in peer achievement, as much as half a standard deviation at Brooklyn Tech. As in Boston, however, the 2SLS estimates of peer and other mediating effects come out close to zero.

Estimates for subgroups of high achievers and minorities in New York appear in Appendix Table B3. Few of the estimates for high achievers are significantly different from zero and most are negative, in line with the Boston findings. On the other hand, unlike Boston, New York's exam schools do not appear to boost Regents achievement for blacks and Hispanics.

6.3 Pooling Boston and New York

To maximize precision, we used the combined Boston and New York samples to construct pooled 2SLS estimates of mediating effects. The models here parallel those used to construct the single-city stacked estimates, with a full set of covariate interactions for each city included as controls. The multi-city results appear in Table 12, again reported for two sets of instruments. Pooling indeed generates a precision gain relative to the single-city 2SLS estimates, and reinforces the main findings showing no overall peer or other mediating effects. The most precisely estimated effects of exam years come with standard errors of $0.019 - 0.023$, while the corresponding peer effects have estimated standard errors on the order of 0.034.

7 Summary and Conclusions

The results reported here show only scattered test score gains due to an exam school education, even for students with relatively high baseline scores. Because the exam school experience is associated with exposure to high-achieving peers and a decline in the relative standing of successful applicants in comparison to peers, these results weigh against the importance of peer effects in the education production function. The outcome that appears to be most strengthened by exam school attendance is the 10th grade ELA score, a result that appears to be driven by gains for minorities. We also find evidence of SAT score gains for a subset of 9th grade applicants, but not enough to boost SAT scores significantly overall. The high achievers in our samples clearly have good outcomes, but most of these students would have done well without the benefit of an exam school education.

It's interesting to contrast the results reported here with those from recent studies of Boston and New York charter schools using quasi-experimental research designs. Abdulkadiroğlu, Angrist, Dynarski, Kane, and Pathak (2011) and Dobbie and Fryer (2011) show substantial gains from attendance at charter schools that embrace the *No Excuses* pedagogical model. Many of these schools serve exceptionally low achievers. Moreover, the relationship between baseline ability and treatment effects *within* the urban charter population appears to be negative (Angrist, Dynarski, Kane, Pathak, and Walters, 2010; Angrist, Pathak, and Walters, 2011). The results reported here, showing evidence of achievement gains for minorities, are therefore broadly consistent with the charter findings. The comparison between *No Excuses* charters and exam schools also suggests that the scope for improvement in learning may be wider at the low end of the ability distribution than at the top. Together, these findings weigh against the view expressed recently by Cunha and Heckman (2007), among others, that "... returns to adolescent education for the most disadvantaged and less able are lower than the returns for the more advantaged" (page 33).

Of course, test scores are only part of the picture. It seems likely, for example, that the Boston Latin School improves students' knowledge of Latin. The many clubs and activities at some exam schools may, perhaps, expose students to ideas and concepts not easily captured by achievement tests. The certification that comes with an exam school education might open doors at elite colleges and universities. If these effects exist, however, our estimates suggest they operate through channels other than increased cognitive achievement.

Finally, our results are relevant to the broader debate on the impacts of school choice as expressed in analyses by Hoxby (2003), Rothstein (2006), among others. The heavy rates of oversubscription for exam schools together with the lack of broad achievement effects suggests that parents either mistakenly equate good peers with high value added, or that they value exam schools for reasons other than their impact on learning. Both of these scenarios reduce the likelihood that parental choice has strong demand-side effects on the production of human capital in schools.

Table 1. Boston School Characteristics

	Traditional Boston Schools		Exam Schools				
	Middle School	*High School*	*Middle School*	*High School*	O'Bryant	Latin Academy	Latin School
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Baseline Peer Mean in Math	-0.251	-0.346	1.508	1.345	0.850	1.159	1.864
Baseline Peer Mean in English	-0.252	-0.274	1.371	1.096	0.731	1.050	1.565
Student/Teacher ratio	12.4	15.2	21.3	21.1	19.6	21.2	22.0
Teachers licensed to teach assignment	87.7%	89.2%	96.3%	96.6%	97.4%	95.9%	96.4%
Core academic teachers identified as highly qualified	84.7%	85.0%	94.0%	93.9%	92.7%	93.8%	94.7%
Teachers above age 40	46.6%	47.4%	54.4%	55.3%	63.4%	51.7%	52.9%
Teachers above age 48	31.9%	35.3%	42.0%	43.0%	51.3%	38.3%	41.2%
Teachers above age 56	11.8%	13.8%	21.4%	22.1%	27.1%	18.7%	21.3%
Number of teachers	46.1	63.1	91.5	89.0	64.5	79.0	110.3
Total number of teachers in core academic areas	37.9	51.7	77.4	76.1	55.7	64.7	95.2
Number of schools	47	40	3	3	1	1	1

Notes: This table shows student weighted average characteristics of teachers and schools using data posted on the Mass DOE website at http://profiles.doe.mass.edu/state_report/teacherdata.aspx. Peer baseline means are enrollment-weighted scores on 4th grade MCAS Math and English for middle school covering Fall 2000 to Fall 2008 for middle school and middle school MCAS scores covering years Fall 2002 to Fall 2008 for high school. Teachers licensed in teaching assignment is the percent of teachers who are licensed with Provisional, Initial, or Professional licensure to teach in the area(s) in which they are teaching. Core classes taught by highly qualified teachers is the percent of core academic classes (defined as English, reading or language arts, mathematics, science, foreign languages, civics and government, economics, arts, history, and geography) taught by highly qualified teachers (defined as teachers not only holding a Massachusetts teaching license, but also demonstrating subject matter competency in the areas they teach). All teacher characteristics are from Fall 2003 to Fall 2008, except information on core academic teachers which is from Fall 2003-2006 and teacher age which is only available from Fall 2007-2008. Middle schools include all schools with positive enrollment in at least one of grades 6, 7, and 8. High schools include all schools with positive enrollment in at least one of grades 9, 10, 11, and 12.

Table 2. Boston Descriptive Statistics

	7th Grade				9th Grade			
	All Boston	Exam Applicants	Offered Students	Enrolled Students	All Boston	Exam Applicants	Offered Students	Enrolled Students
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				A. Demographics				
Female	0.479	0.536	0.559	0.562	0.476	0.540	0.614	0.602
Black	0.478	0.386	0.245	0.239	0.505	0.493	0.361	0.367
Hispanic	0.301	0.199	0.158	0.149	0.331	0.243	0.233	0.215
Free Lunch	0.725	0.717	0.630	0.626	0.762	0.805	0.783	0.799
LEP‡	0.201	0.139	0.110	0.110	0.181	0.130	0.117	0.133
SPED¥	0.232	0.045	0.009	0.009	0.250	0.079	0.019	0.015
N	61,161	13,730	6,418	5,652	30,484	5,540	1,461	1,095
				*B. Baseline Scores**				
Math	-0.017	0.758	1.399	1.436	-0.313	0.227	1.036	1.058
English	-0.020	0.725	1.286	1.315	-0.246	0.275	0.835	0.824
N	37,780	9,423	4,577	4,055	27,505	5,461	1,436	1,081

Notes: This table reports sample means for 1997-2008. The All Boston sample includes 6th and 8th grade students in Boston public schools who had not previously enrolled in any exam school. Exam Applicants are students with a valid application; offered students are applicants who receive an offer at any exam school; enrolled students are applicants who enrolled at any exam school in the following school year. Baseline Math and English scores for 7th grade applicants are from 4th grade. Baseline scores for 9th grade applicants are from middle school. N is the number of observations with at least one non-missing value for the variable listed.

‡ Limited English Proficient (LEP) only available beginning in year 1998.

¥ Special Education (SPED) status only available for years 1998-2004.

* Baseline scores available from 2000 onward for 6th grade and from 2002 onward for grade 8.

Table 3. Boston OLS Estimates for Enrollment, Years in Exam School, and Peer Means

	Math						English					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Enrollment	0.990***			0.272***			0.788***			0.247***		
	(0.049)			(0.032)			(0.043)			(0.028)		
Exam Years		0.361***			0.088***			0.314***			0.121***	
		(0.012)			(0.016)			(0.012)			(0.010)	
Peer Mean			0.610***			0.110***			0.558***			0.154***
			(0.015)			(0.021)			(0.014)			(0.023)
N	24349	24368	20650	24349	24368	20650	22737	22750	21453	22737	22750	21453
ISEE Controls	NO	NO	NO	YES	YES	YES	NO	NO	NO	YES	YES	YES

Notes: This table reports first-stage estimates for three measures of exam school exposure. Enrollment is an indicator of enrollment at an exam school in the year following application; exam years is the number of years enrolled at an exam school prior to test; peer mean is the average baseline score of peers in the year following application. Models control for application cohort and grade, test year, and demographics (race, gender, free lunch). ISEE controls are raw scores from the verbal, quantitative, reading, and math sections of the test. Robust standard errors, clustered on year and school at the time of testing,
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 4. Boston Reduced Form Estimates - MCAS Math and English

		Parametric (Discontinuity Sample)				Optimal Bandwidth (IK)			
Application Grade	Test Grade	O'Bryant (1)	Latin Academy (2)	Latin School (3)	All Schools (4)	O'Bryant (5)	Latin Academy (6)	Latin School (7)	All Schools (8)
					A. Math				
7th	7th and 8th	-0.120	0.007	-0.047	-0.052	-0.065	-0.074	-0.031	-0.056*
		(0.092)	(0.075)	(0.092)	(0.045)	(0.062)	(0.061)	(0.050)	(0.031)
		5562	5572	5308	16442	5261	5503	5102	15866
7th and 9th	10th	0.075	-0.046	-0.042	0.002	0.060	-0.035	-0.080***	-0.014
		(0.070)	(0.076)	(0.061)	(0.033)	(0.047)	(0.042)	(0.027)	(0.025)
		3732	3393	3123	10248	3335	2714	2760	8809
7th and 9th	7th, 8th, and 10th	-0.037	-0.012	-0.046	-0.032	-0.013	-0.061	-0.048	-0.041*
		(0.066)	(0.060)	(0.072)	(0.035)	(0.047)	(0.041)	(0.038)	(0.024)
		9294	8965	8431	26690	8596	8217	7862	24675
					B. English				
7th	7th and 8th	-0.119	-0.024	-0.144**	-0.094***	-0.066	0.020	-0.122***	-0.059**
		(0.081)	(0.068)	(0.059)	(0.035)	(0.046)	(0.044)	(0.033)	(0.028)
		5003	4992	4649	14644	3835	4719	4442	12996
7th and 9th	10th	0.120	0.204**	0.008	0.113**	0.156***	0.194***	0.030	0.130***
		(0.081)	(0.094)	(0.078)	(0.045)	(0.052)	(0.057)	(0.064)	(0.035)
		3741	3401	3127	10269	2658	2038	1981	6677
7th and 9th	7th, 8th, and 10th	-0.006	0.063	-0.087	-0.009	0.029	0.076*	-0.078**	0.006
		(0.058)	(0.062)	(0.057)	(0.035)	(0.036)	(0.041)	(0.034)	(0.025)
		8744	8393	7776	24913	6493	6757	6423	19673

Notes: This table reports estimates of the effects of exam school offers on MCAS scores. The discontinuity sample covers students within 20 standardized units of offer cutoffs. Parametric models include a cubic function of the running variable, allowed to differ on either side of offer cutoffs. IK estimates use the edge kernel, with bandwidth computed following Imbens and Kalyanaraman (2010). Controls include other-school offers. Running variable controls are interacted with all offer dummies and allowed to differ on either side of offer cutoffs. Optimal bandwidths were computed separately for each school. Robust standard errors, clustered on year and school, are shown in parentheses. Standard errors for all schools estimates and for those pooling outcomes also cluster on student. The number of observations is reported below standard errors.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 5. Boston Reduced Form Estimates for Subgroups

		Baseline in Upper Half			Baseline in Upper Quartile			Blacks and Hispanics	
		Baseline	Proportion	All	Baseline	Proportion	All	Test score	All
Application	Test	Mean	Advanced	Schools	Mean	Advanced	Schools	Mean	Schools
Grade	Grade	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
					A. Math				
7th	7th and 8th	1.534	0.277	-0.116***	2.128	0.388	-0.088	0.772	-0.057
				(0.040)			(0.069)		(0.042)
		5254	4684	7853	2620	2361	3490	4763	8239
7th and 9th	10th	1.377	0.765	-0.024	1.854	0.848	-0.019	0.733	-0.019
				(0.025)			(0.031)		(0.040)
		4625	3608	5290	2606	2096	3123	3180	4724
					B. English				
7th	7th and 8th	1.418	0.123	-0.093***	1.847	0.174	-0.098**	0.823	-0.042
				(0.032)			(0.046)		(0.039)
		5841	5222	7316	2971	2686	3489	4221	7375
7th and 9th	10th	1.249	0.336	0.081**	1.575	0.412	0.088**	0.761	0.182***
				(0.037)			(0.042)		(0.040)
		4346	3326	4855	2474	1913	2622	3191	3718

Notes: This table reports reduced form estimates for students with high baseline scores and for minorities. Baseline means and the proportion of applicants at an advanced level are computed for those who belong to at least one discontinuity sample. The table shows IK estimates with bandwidth computed as in the all schools model in Table 4.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6. Boston 2SLS Estimates for Enrollment, Years in Exam School, and Peer Means

	Instrument: Offer Indicators						Instrument: Offer Indicators x Application Cohort					
	Math			English			Math			English		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	2SLS Estimates						*2SLS Estimates*					
Enrollment	-0.036			0.055			-0.025			0.042		
	(0.063)			(0.049)			(0.061)			(0.046)		
Exam Years		-0.017			0.028			-0.002			0.026	
		(0.030)			(0.026)			(0.029)			(0.025)	
Peer Mean			-0.098**			-0.017			-0.062			-0.011
			(0.045)			(0.045)			(0.040)			(0.041)
	First Stage Estimates											
O'Bryant	0.732***	1.559***	0.762***	0.740***	1.352***	0.696***						
	(0.063)	(0.125)	(0.069)	(0.067)	(0.125)	(0.067)						
Latin Academy	0.156***	0.299**	0.421***	0.157**	0.227**	0.384***						
	(0.059)	(0.132)	(0.082)	(0.063)	(0.114)	(0.075)						
Latin School	0.027	0.068	0.618***	0.024	0.041	0.525***						
	(0.018)	(0.051)	(0.092)	(0.018)	(0.041)	(0.078)						
N	24659	24675	20702	19670	19673	18743	24659	24675	20702	19670	19673	18743

Notes: This table reports two-stage least squares (2SLS) estimates of the effects of exam school enrollment, years spent in exam school, and mean baseline peer achievement on MCAS scores. The table shows IK estimates using the reduced form bandwidths computed for Table 4. Instruments for columns 1-6 are three offer dummies. Columns 7-12 show the results of adding cohort interactions to the instrument list.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7. Boston Reduced Form Estimates - PSAT and SAT Scores

	Probability Tested				Test Score for Takers			
Application Grade	O'Bryant (1)	Latin Academy (2)	Latin School (3)	All Schools (4)	O'Bryant (5)	Latin Academy (6)	Latin School (7)	All Schools (8)
				A. PSAT				
7th	0.077*	0.061	-0.024	0.039	0.047	-0.069	0.017	-0.002
	(0.039)	(0.039)	(0.036)	(0.025)	(0.059)	(0.055)	(0.054)	(0.033)
	2712	2679	2410	7801	1797	1631	1113	4541
9th	0.085**	-0.020	0.134*	0.067**	0.025	0.223***	0.050	0.070*
	(0.041)	(0.057)	(0.074)	(0.032)	(0.061)	(0.055)	(0.118)	(0.042)
	1745	886	442	3073	1306	447	372	2125
7th and 9th	0.080***	0.044	-0.005	0.046**	0.037	-0.003	0.025	0.022
	(0.028)	(0.033)	(0.033)	(0.020)	(0.043)	(0.050)	(0.054)	(0.028)
	4457	3565	2852	10874	3103	2078	1485	6666
				B. SAT				
7th	0.024	0.077*	0.030	0.043	0.044	-0.069	0.075	0.014
	(0.042)	(0.042)	(0.038)	(0.026)	(0.060)	(0.052)	(0.060)	(0.037)
	2694	2686	2443	7823	1425	1701	1454	4580
9th	0.040	-0.083	0.015	0.011	0.036	0.318***	0.183**	0.118***
	(0.047)	(0.075)	(0.083)	(0.038)	(0.056)	(0.096)	(0.077)	(0.046)
	1687	832	544	3063	1075	428	335	1838
7th and 9th	0.031	0.044	0.027	0.034	0.040	0.008	0.094*	0.045
	(0.031)	(0.037)	(0.035)	(0.022)	(0.044)	(0.041)	(0.052)	(0.029)
	4381	3518	2987	10886	2500	2129	1789	6418

Notes: This table reports estimates of the effects of exam school offers on PSAT and SAT test taking and scores. Outcome-specific IK estimates, bandwidths, and standard errors were computed as for Table 4.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 8. Boston Reduced Form Estimates - AP Exams

	Number of Exams				Sum of Scores			
Application Grade	O'Bryant (1)	Latin Academy (2)	Latin School (3)	All Schools (4)	O'Bryant (5)	Latin Academy (6)	Latin School (7)	All Schools (8)
				A. All Exams				
7th	0.245**	-0.182	0.031	0.037	1.164***	-0.410	0.068	0.250
	(0.118)	(0.170)	(0.139)	(0.070)	(0.310)	(0.438)	(0.510)	(0.242)
	1750	1871	1886	5507	1750	1871	1471	5092
9th	-0.115	0.088	-0.178	-0.076	0.170	0.817	-0.264	0.235
	(0.167)	(0.289)	(0.348)	(0.112)	(0.325)	(0.855)	(1.048)	(0.247)
	1244	668	499	2411	1244	544	447	2235
7th and 9th	0.096	-0.116	-0.003	0.004	0.716**	-0.166	-0.001	0.245
	(0.120)	(0.171)	(0.136)	(0.058)	(0.281)	(0.419)	(0.467)	(0.183)
	2994	2539	2385	7918	2994	2415	1918	7327
				B. Exams with 500+ Takers				
7th	0.103	-0.191	-0.199	-0.089	0.609***	-0.662*	-0.929**	-0.333
	(0.096)	(0.141)	(0.125)	(0.064)	(0.214)	(0.396)	(0.375)	(0.221)
	1750	1707	1886	5343	1750	1475	1886	5111
9th	-0.220	0.118	-0.403	-0.175*	-0.109	0.895	-0.602	0.008
	(0.139)	(0.249)	(0.321)	(0.105)	(0.268)	(0.804)	(0.921)	(0.232)
	1244	641	570	2455	1244	539	475	2258
7th and 9th	-0.036	-0.110	-0.238**	-0.116**	0.263	-0.272	-0.871***	-0.221
	(0.097)	(0.141)	(0.117)	(0.057)	(0.200)	(0.399)	(0.312)	(0.164)
	2994	2348	2456	7798	2994	2014	2361	7369

Notes: This table reports estimates of effects of exam school offers on AP test taking and scores. Tests with 500+ or more takers are Calculus AB, Statistics, Biology, Chemistry, Physics B, English Language and Composition, English Literature and Composition, European History, US Government and Politics, US History, Microeconomics, Macroeconomics, and Spanish Language. Outcome-specific IK estimates, bandwidths, and standard errors were computed as for Table 4.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 9. Descriptive Statistics of NYC Exam School Applicants

		Any Exam			Enrolled in		
	All NYC (1)	Exam Applicants (2)	Offered Students (3)	Enrolled Students (4)	Brooklyn Tech (5)	Bronx Science (6)	Stuyvesant (7)
				A. Demographics			
Female	0.487	0.503	0.456	0.426	0.415	0.443	0.429
Black	0.336	0.299	0.078	0.076	0.133	0.040	0.019
Hispanic	0.377	0.248	0.073	0.067	0.089	0.070	0.030
Free Lunch#	0.667	0.685	0.671	0.681	0.664	0.682	0.706
LEP	0.125	0.039	0.004	0.005	0.007	0.003	0.003
SPED	0.089	0.006	0.000	0.000	0.000	0.000	0.000
N	453233	84539	11914	9364	4255	2405	2704
				B. Baseline Scores			
Math	-0.004	0.779	1.780	1.802	1.619	1.771	2.119
English	-0.005	0.709	1.714	1.667	1.426	1.666	2.047
N	349817	82527	11841	9312	4231	2397	2684

Notes: This table reports sample means for 2004-2007. The All NYC sample includes 8th graders in NYC public schools. Exam applicants are students who applied to Brooklyn Tech, Bronx Science, or Stuyvesant. Offered students are applicants offered a seat at any of these schools. Enrolled students are applicants who register at one of these schools in the year following application. Baseline scores are from 8th grade NYSED Math and Reading.

For applicants in 2004 and 2005, free lunch status is from school year 2004-2005 (after assignment), while for applicants in 2006 and 2007, free lunch status is from school year 2004-2005 and 2005-2006 (before assignment).

Table 10. NYC Reduced Form Estimates - Regents Exams

	Parametric (Discontinuity Sample)				Optimal Bandwidth (IK)			
	Brooklyn Tech	Bronx Science	Stuyvesant	All Schools	Brooklyn Tech	Bronx Science	Stuyvesant	All Schools
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Math	0.084	-0.097*	-0.056	-0.021	0.013	-0.130***	-0.037	-0.055***
	(0.065)	(0.056)	(0.039)	(0.029)	(0.040)	(0.033)	(0.038)	(0.018)
	4264	3746	3800	11810	3743	3746	3417	10906
Advanced Math	-0.030	-0.022	-0.030	-0.027	0.004	-0.057	-0.022	-0.024
	(0.080)	(0.072)	(0.050)	(0.044)	(0.044)	(0.040)	(0.027)	(0.021)
	5619	5524	6584	17727	5619	5524	6584	17727
English	0.028	-0.035	-0.028	-0.013	0.051	-0.018	-0.005	0.013
	(0.057)	(0.042)	(0.030)	(0.025)	(0.039)	(0.022)	(0.022)	(0.014)
	4950	4581	5150	14681	4950	4217	5150	14317
Global History	-0.085	-0.024	-0.008	-0.036	-0.072**	-0.017	0.009	-0.025
	(0.053)	(0.042)	(0.038)	(0.025)	(0.035)	(0.028)	(0.025)	(0.015)
	6277	5925	6863	19065	4757	4699	5222	14678
US History	-0.071*	-0.012	0.038	-0.011	-0.053**	-0.017	0.038	-0.007
	(0.038)	(0.032)	(0.036)	(0.023)	(0.024)	(0.026)	(0.023)	(0.016)
	4440	4281	4987	13708	2808	4148	3797	10753
Living Environment	-0.059	0.092**	-0.072**	-0.015	-0.080***	0.056**	-0.031	-0.025**
	(0.041)	(0.039)	(0.033)	(0.020)	(0.022)	(0.024)	(0.020)	(0.012)
	5801	5508	6276	17585	5801	5508	6276	17585

Notes: This table reports estimates of the effect of exam school offers on New York Regents scores. The discontinuity sample includes applicants five standardized units from the cutoff. Model parameterizations and estimation procedures are the same as for Boston. Math scores are from Regents Math A (Elementary Algebra and Planar Geometry) or Integrated Algebra I. Advanced Math scores are from Regents Math B (Intermediate Algebra and Trigonometry) or Geometry. The table reports robust standard errors, clustered on year and school of test, in parentheses. Standard errors are also clustered on student when schools are stacked. Sample sizes for each outcome are reported below the standard errors.

*significant at 10%; **significant at 5%; ***significant at 1%.

Table 11. New York 2SLS Estimates for Enrollment, Years in Exam School, and Peer Means

	Advanced Math			English		
	(1)	(2)	(3)	(4)	(5)	(6)
				2SLS		
Enrollment	-0.065			0.070		
	(0.083)			(0.069)		
Exam Years		-0.024			0.028	
		(0.038)			(0.029)	
Peer Mean			-0.032			0.066
			(0.062)			(0.052)
				First Stage		
Brooklyn Tech	0.403***	0.972***	0.561***	0.407***	0.903***	0.523***
	(0.075)	(0.155)	(0.064)	(0.089)	(0.220)	(0.057)
Bronx Science	0.328***	0.637***	0.178***	0.307***	0.790***	0.155**
	(0.112)	(0.213)	(0.068)	(0.113)	(0.293)	(0.073)
Stuyvesant	0.087	0.104	0.272***	0.067	0.118	0.258***
	(0.074)	(0.131)	(0.076)	(0.081)	(0.197)	(0.095)
N		17727			14317	

Notes: This table reports 2SLS estimates of the effect of exam school enrollment, years in exam school, and peer achievement on Regents scores. Advanced Math scores are from Regents Math B (Intermediate Algebra and Trigonometry) or Geometry. Peer means are from 8th grade NYSED tests. The table shows IK estimates using the same bandwidths as for the reduced form estimates in Table 10. Robust standard errors clustered by year and school of test, and by student, are shown in parenthesis.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 12. Boston and New York: 2SLS Estimates for Enrollment, Years in Exam School, and Peer Means

	Instrument: Offer Indicators						Instrument: Offer Indicators x Application Cohort Indicators					
	Math			English			Math			English		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	2SLS Estimates											
Enrollment	-0.045			0.059			-0.027			0.043		
	(0.051)			(0.040)			(0.049)			(0.037)		
Exam Years		-0.019			0.028			-0.007			0.021	
		(0.024)			(0.020)			(0.023)			(0.019)	
Peer Mean			-0.080**			0.007			-0.048			0.011
			(0.037)			(0.035)			(0.034)			(0.033)
	First Stage Estimates											
O'Bryant	0.732***	1.559***	0.762***	0.740***	1.352***	0.696***						
	(0.063)	(0.125)	(0.069)	(0.067)	(0.125)	(0.067)						
Latin Academy	0.156***	0.299**	0.421***	0.157**	0.227**	0.384***						
	(0.059)	(0.132)	(0.082)	(0.063)	(0.114)	(0.075)						
Latin School	0.027	0.068	0.618***	0.024	0.041	0.525***						
	(0.018)	(0.051)	(0.092)	(0.018)	(0.041)	(0.078)						
Brooklyn Tech	0.403***	0.972***	0.561***	0.407***	0.903***	0.523***						
	(0.075)	(0.155)	(0.064)	(0.089)	(0.220)	(0.057)						
Bronx Science	0.328***	0.637***	0.178***	0.307***	0.790***	0.155**						
	(0.112)	(0.213)	(0.068)	(0.113)	(0.293)	(0.073)						
Stuyvesant	0.087	0.104	0.272***	0.067	0.118	0.258***						
	(0.074)	(0.131)	(0.076)	(0.081)	(0.197)	(0.095)						
N	42386	42402	38429	33987	33990	33060	42386	42402	38429	33987	33990	33060

Notes: This table reports two-stage least squares (2SLS) estimates of the effects of exam school enrollment, years spent in exam school, and mean baseline peer achievement on MCAS scores in a sample combining Boston and New York. Boston scores are from MCAS Math and English tests for all grades tested; NYC scores are Advanced Math (Regents Math B or Geometry) and Regents English. The table shows IK estimates using bandwidths computed one city at a time. Instruments for columns 1-6 are three offer dummies. Columns 7-12 show the results of adding cohort interactions to the instrument list.
* significant at 10%; ** significant at 5%; *** significant at 1%

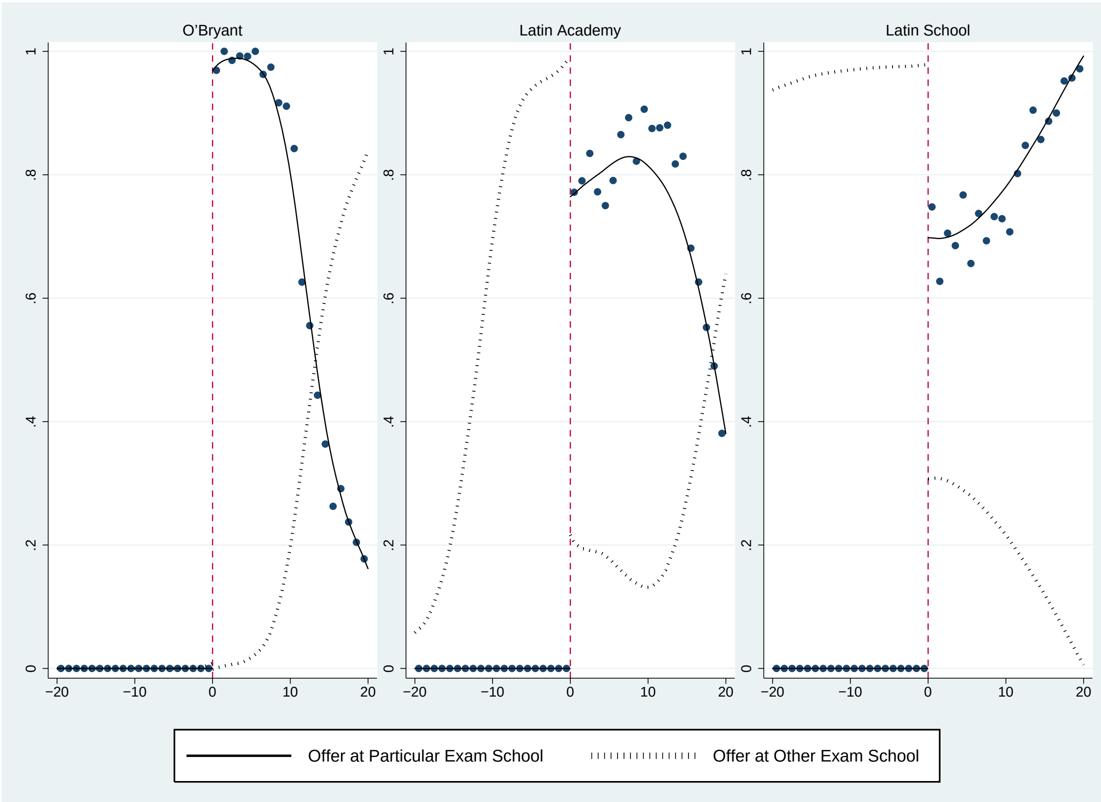

Figure 1. Offers at Each Boston Exam School for 7th Grade Applicants (1997-2008)

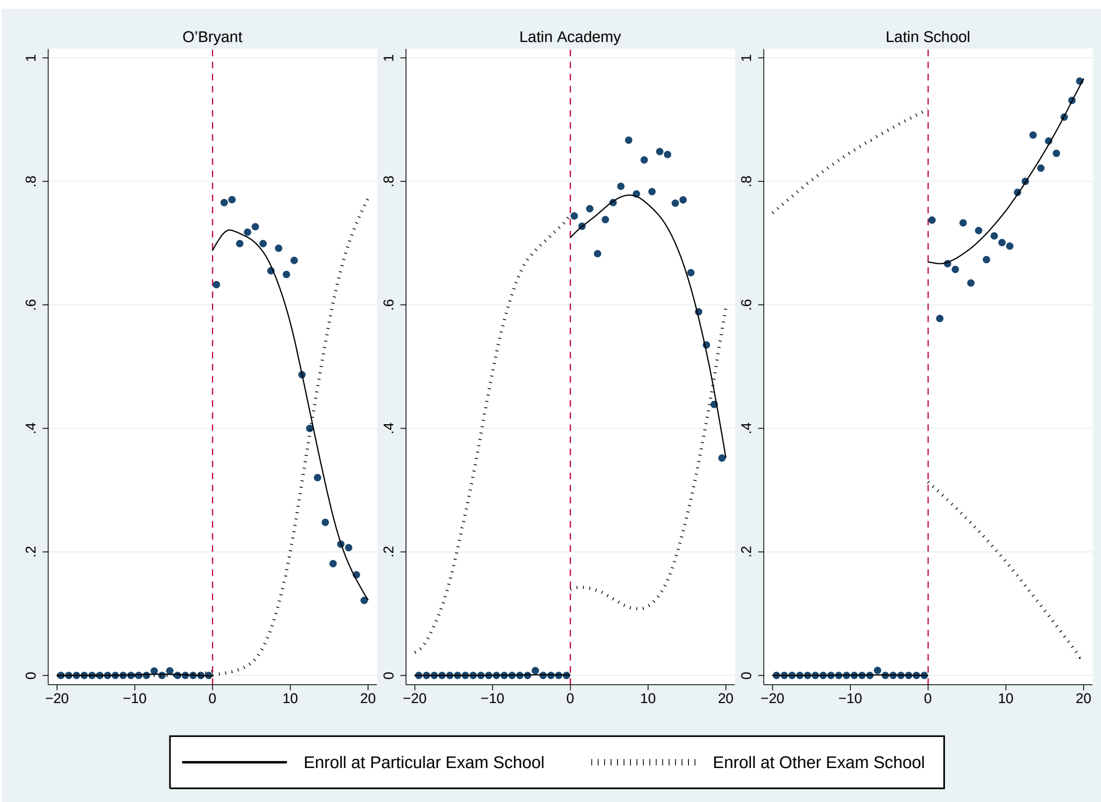

Figure 2. Enrollment at Each Boston Exam School for 7th Grade Applicants (1997-2008)

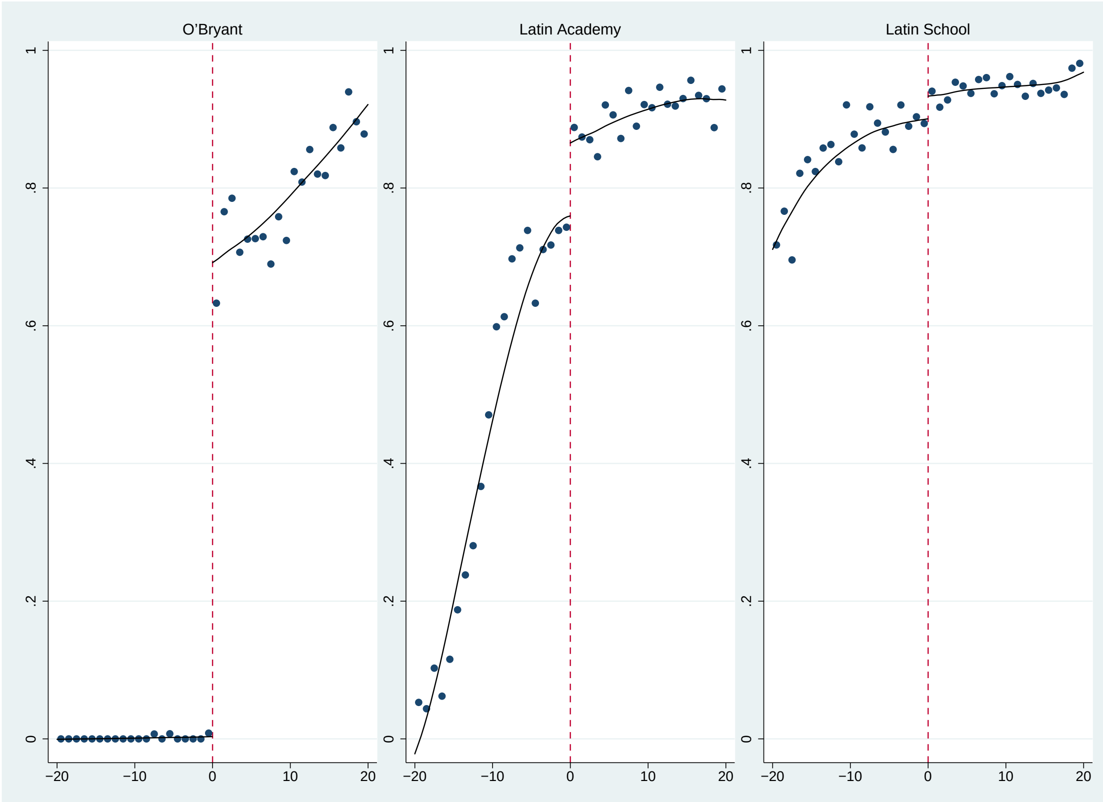

Figure 3. Enrollment at Any Boston Exam School for 7th Grade Applicants (1997-2008)

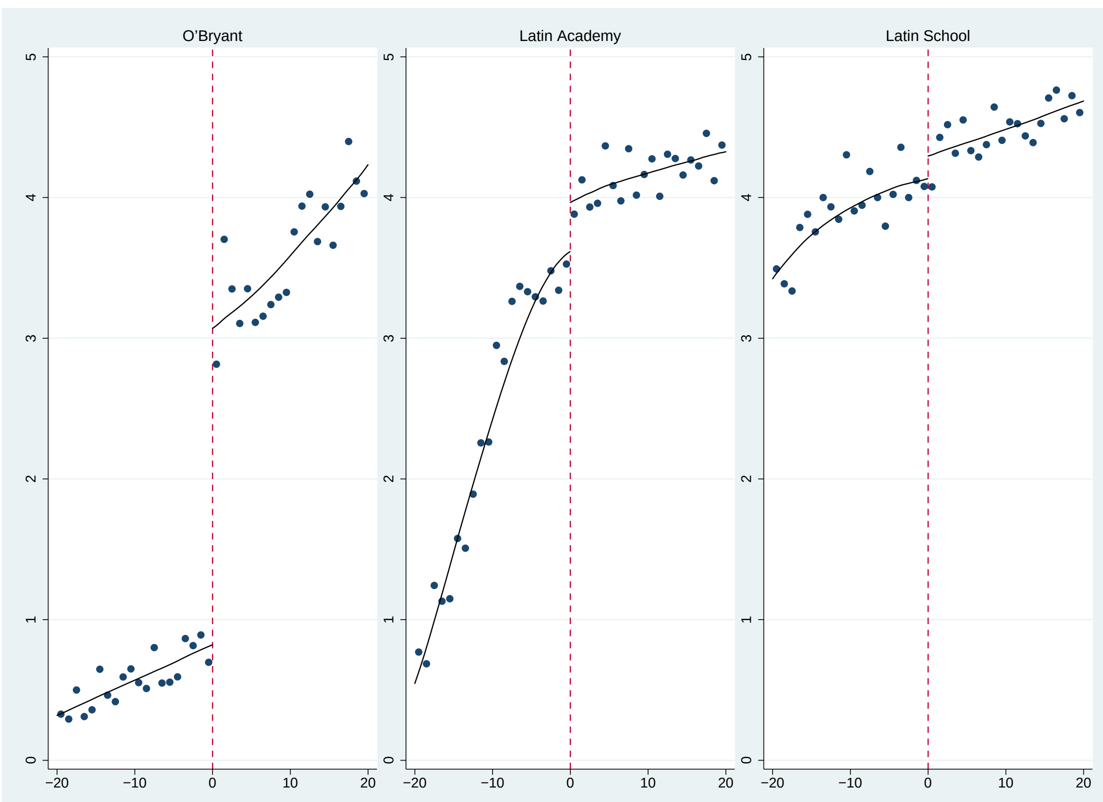

Figure 4. Years at Any Boston Exam School for 7th Grade Applicants (1997-2008)

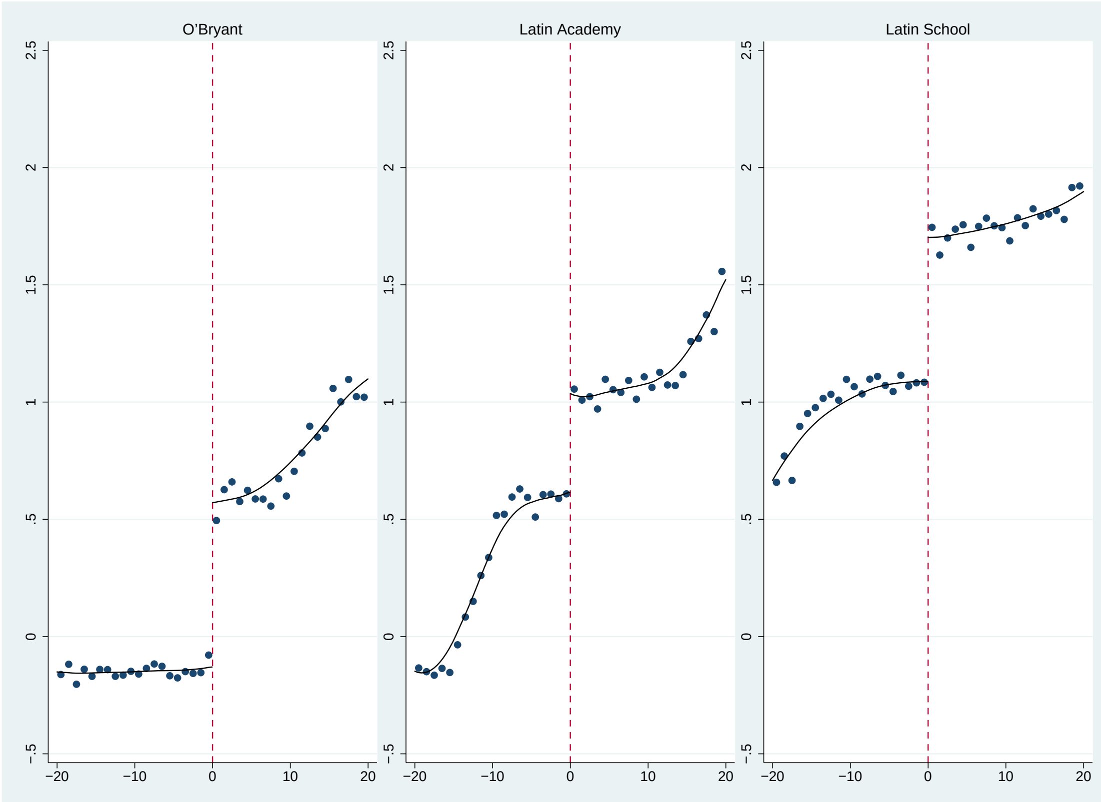

Figure 5. Average Baseline Math Scores of Peers for 7th Grade Applicants (1997-2008) in Boston

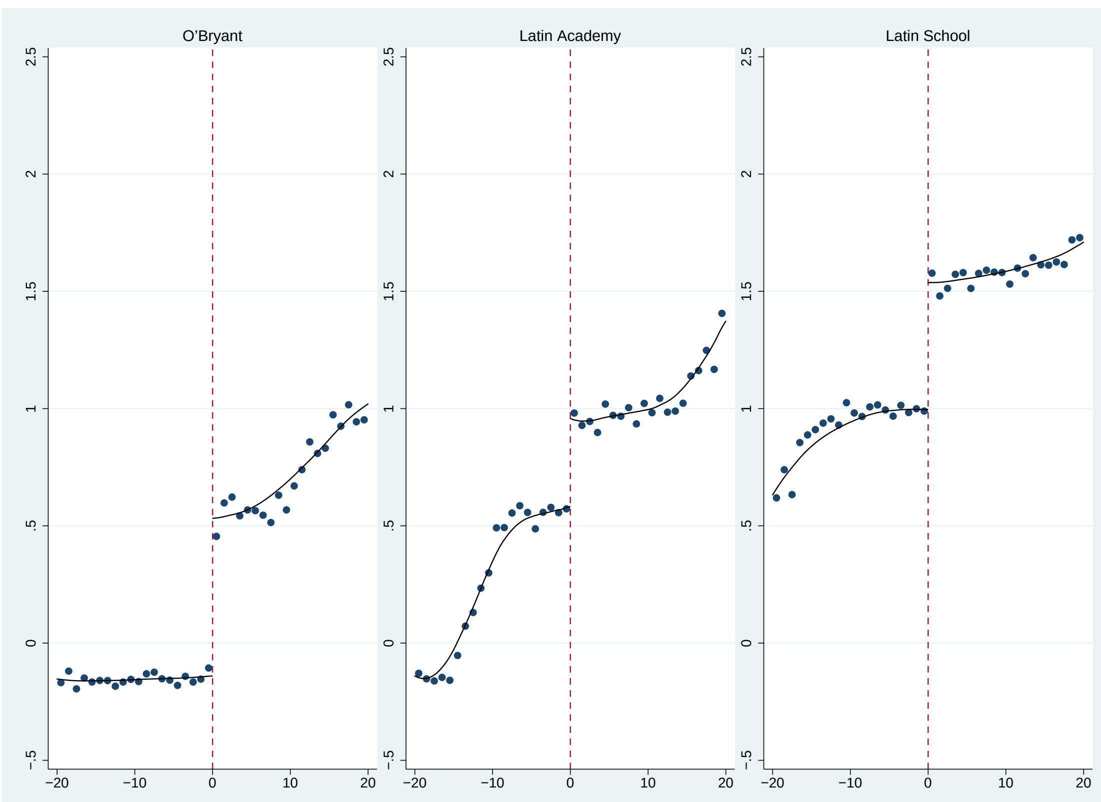

Figure 6. Average Baseline English Scores of Peers for 7th Grade Applicants (1997-2008) in Boston

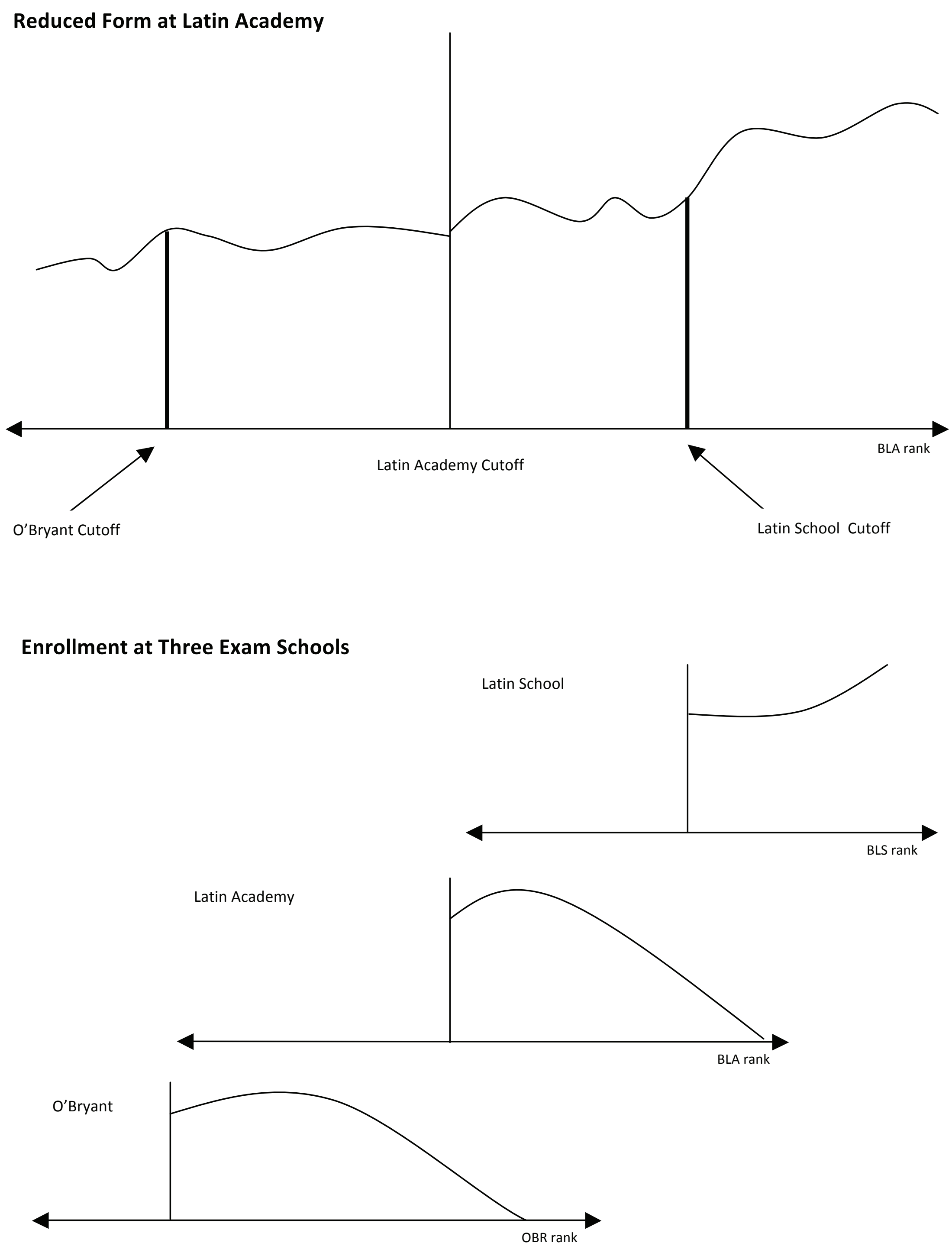

Figure 7. Reduced Form at Boston Latin Academy and Enrollment with Three Running Variables

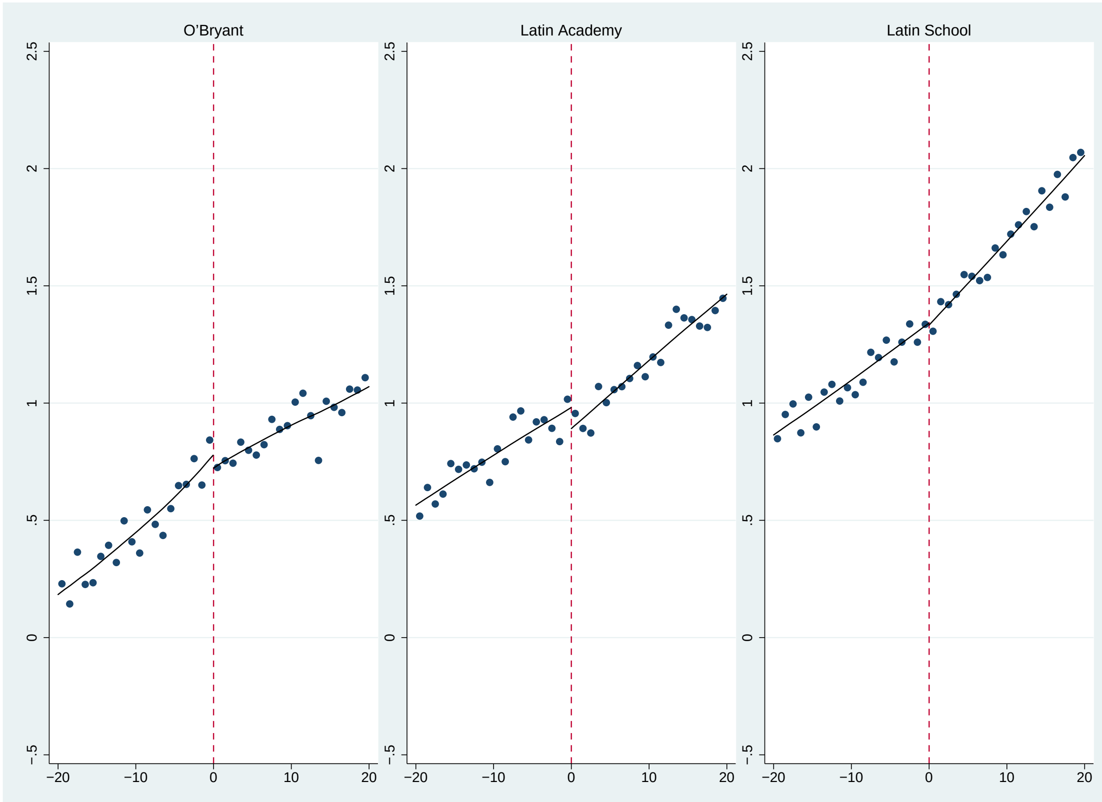

Figure 8. 7th (2006-2009) and 8th (1999-2009) Grade Math Scores for 7th Grade Applicants (1997-2007 / 2005-2008) in Boston

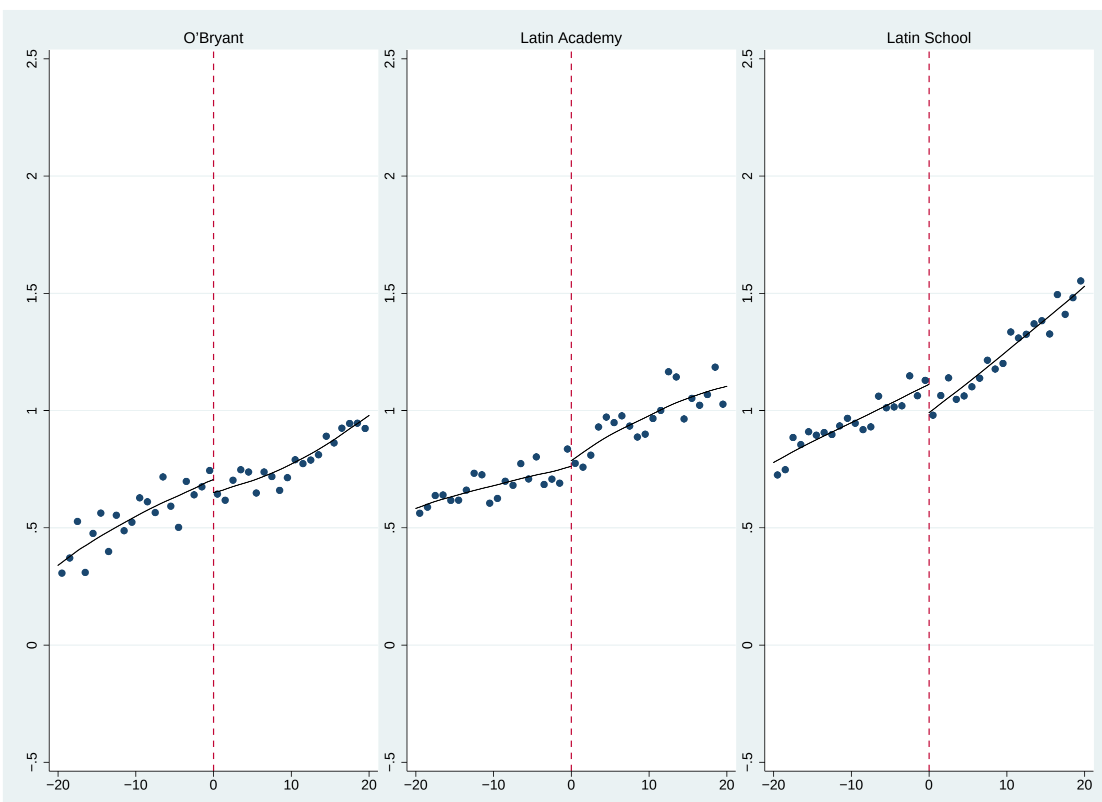

Figure 9. 7th (2001-2009) and 8th (2006-2009) Grade English Scores for 7th Grade Applicants (2000-2008 / 2004-2007) in Boston

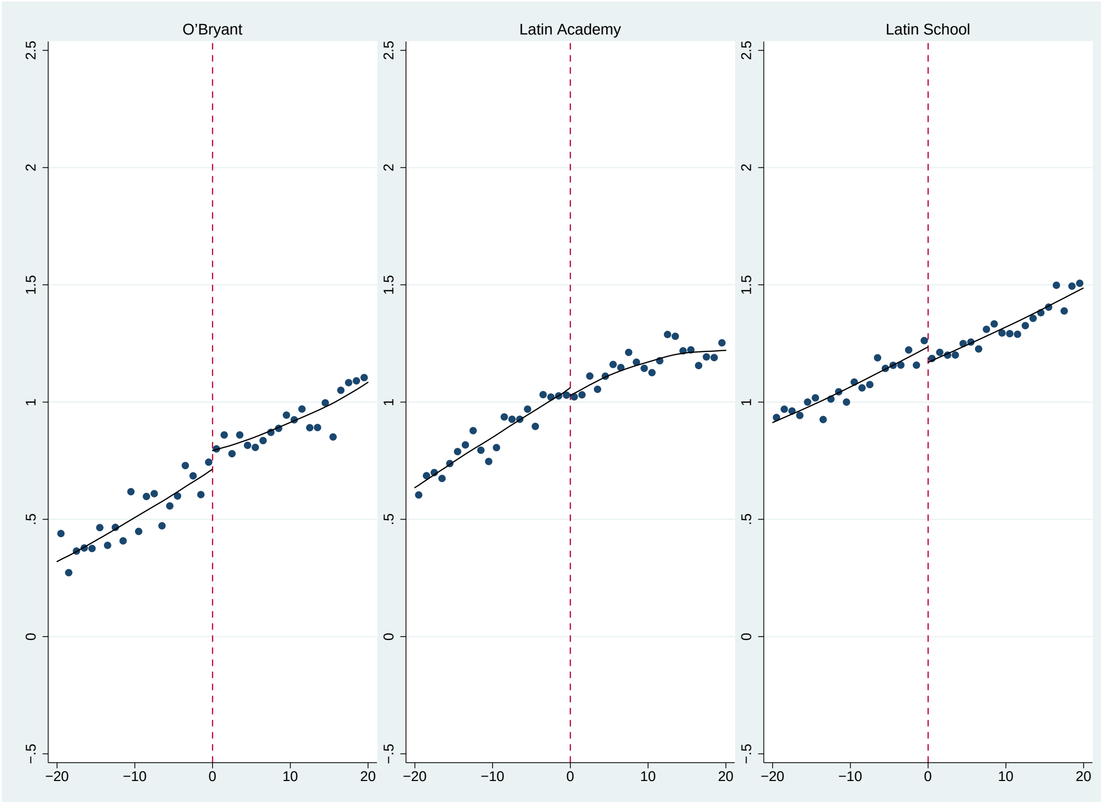

Figure 10. 10th Grade Math (2003-2009) Scores for 7th (1999-2005) and 9th (2001-2007) Grade Applicants in Boston

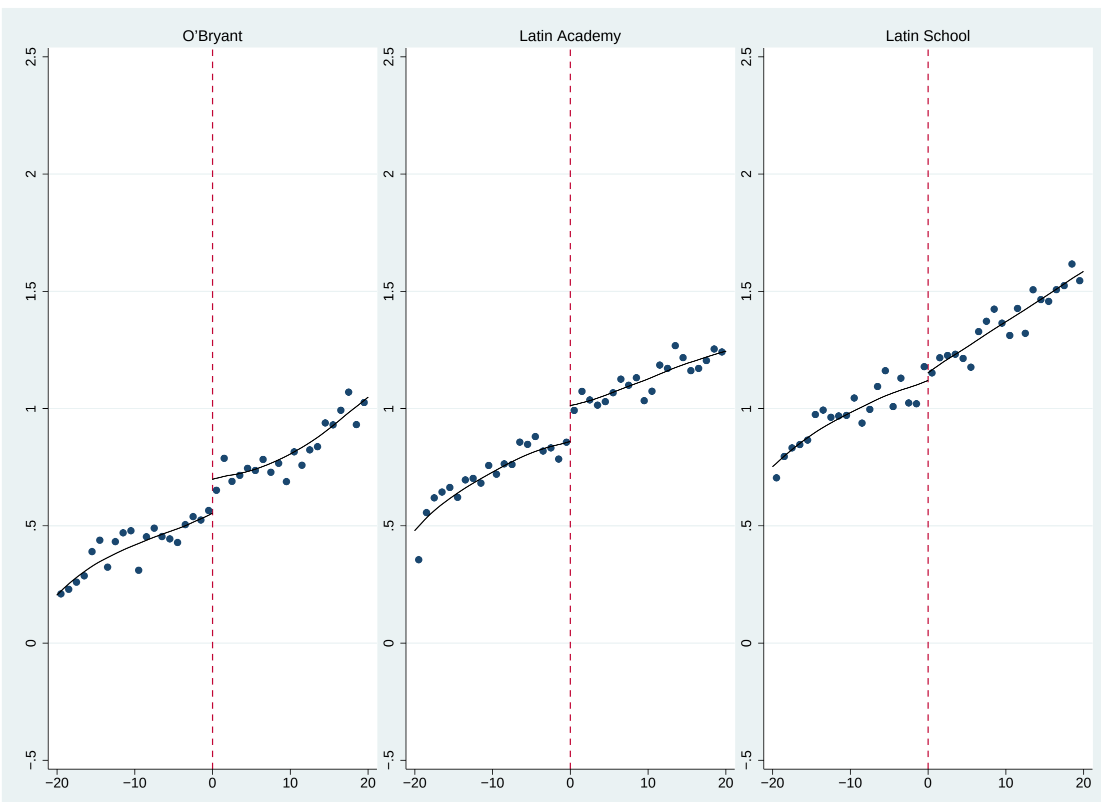

Figure 11. 10th Grade English (2003-2009) Scores for 7th (1999-2005) and 9th (2001-2007) Grade Applicants in Boston

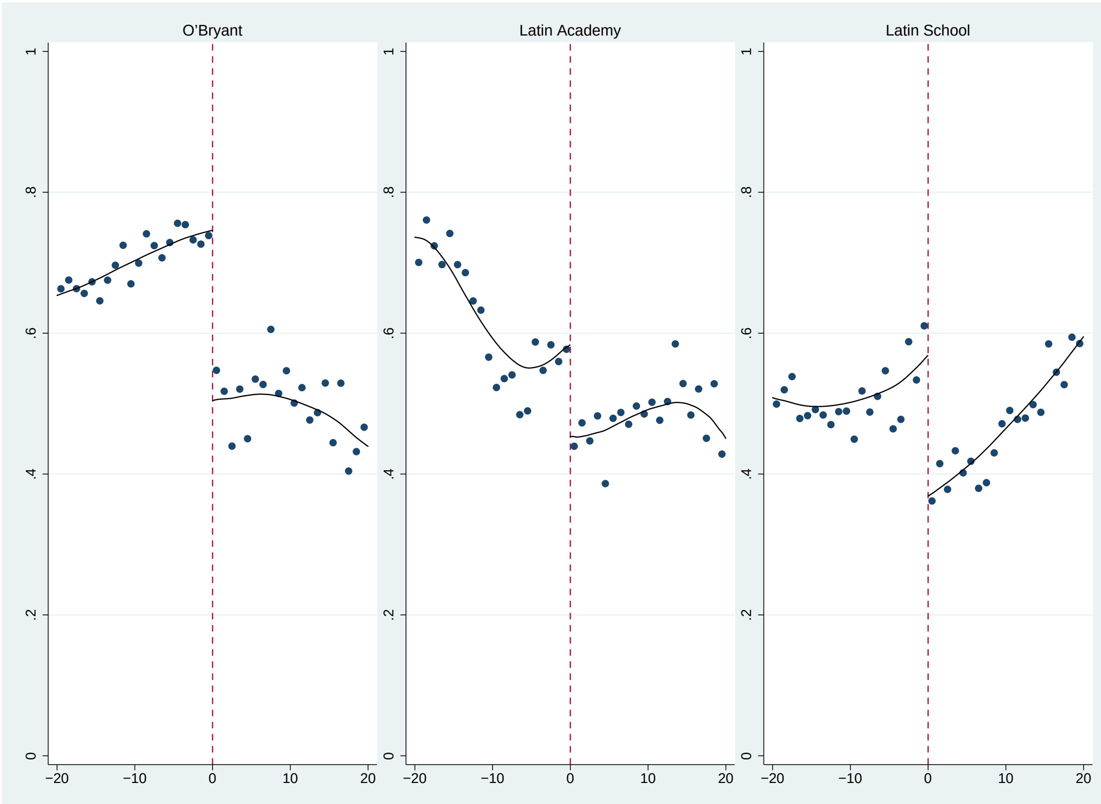

Figure 12. Rank in Baseline Math for 7th Grade Applicants (1997-2008) in Boston

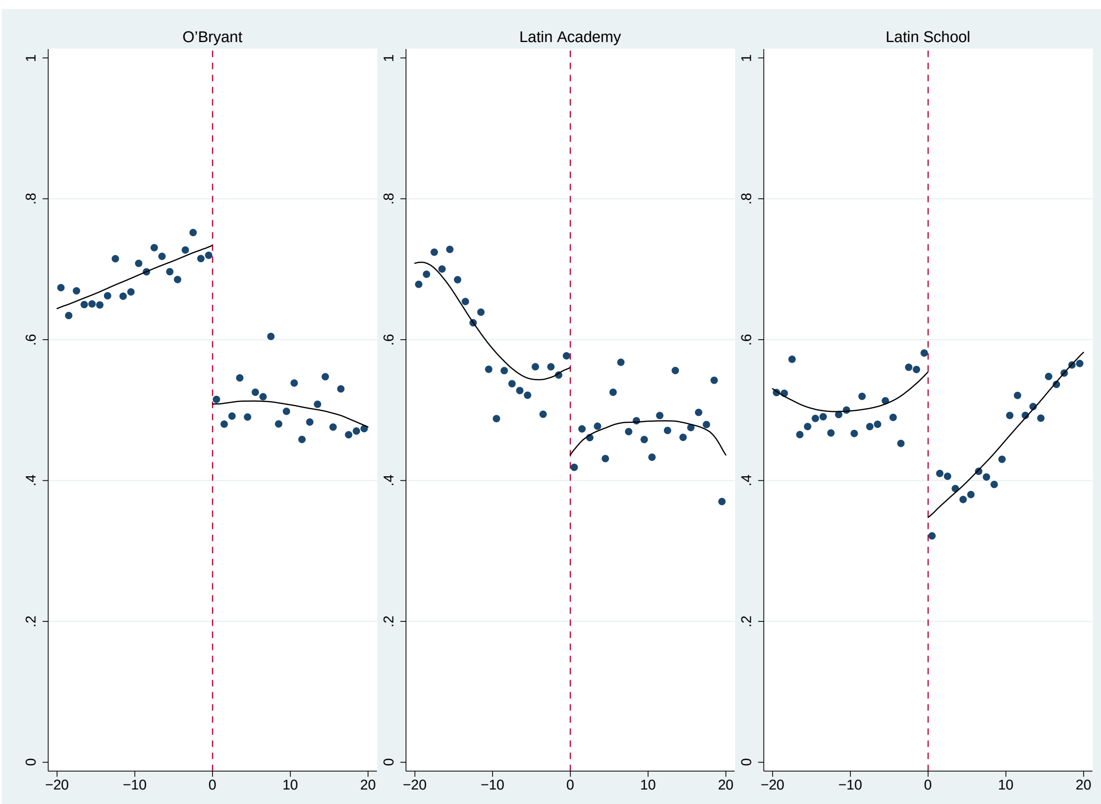

Figure 13. Rank in Baseline English for 7th Grade Applicants (1997-2008) in Boston

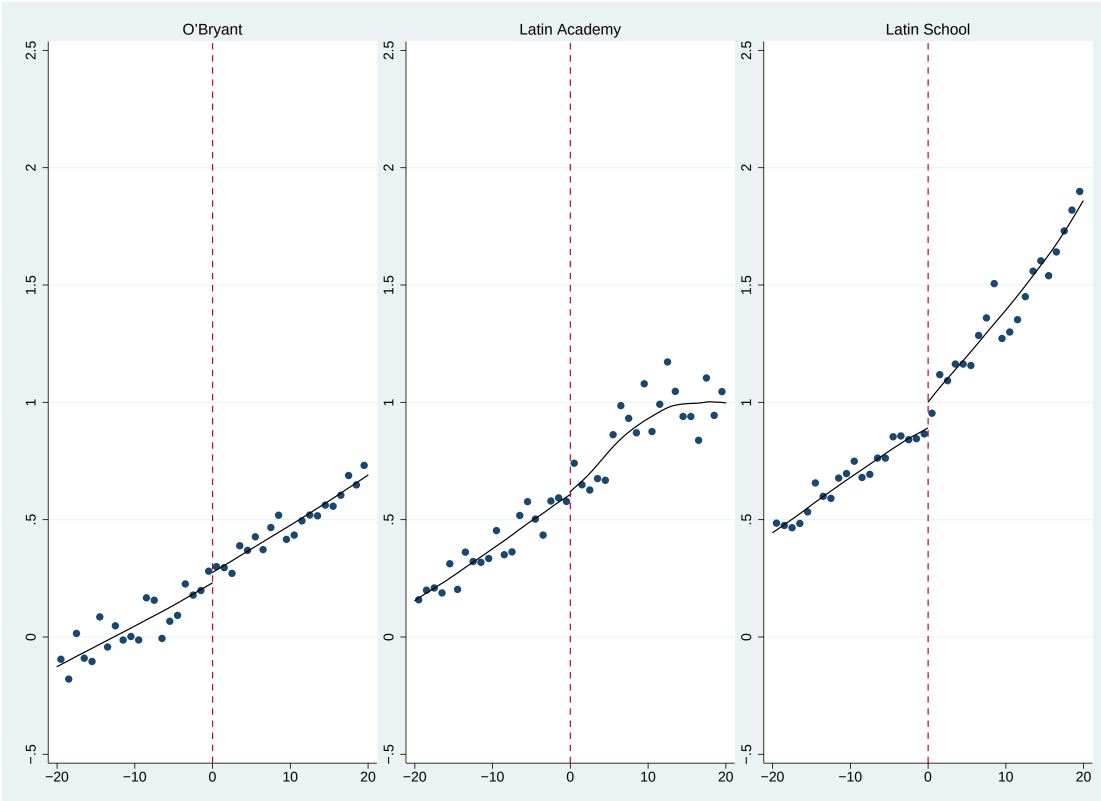

Figure 14. PSAT Scores for 7th (2000-2005) and 9th (2002-2007) Grade Applicants in Boston

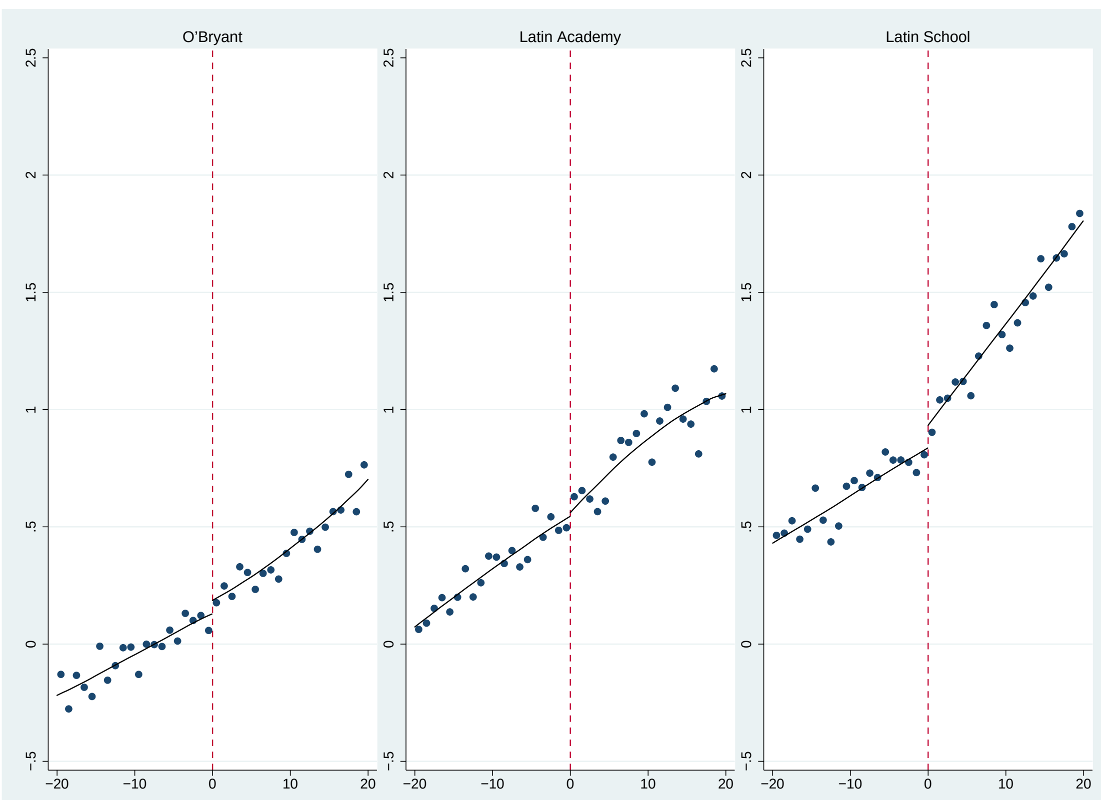

Figure 15. SAT Scores for 7th (2000-2005) and 9th (2001-2006) Grade Applicants in Boston

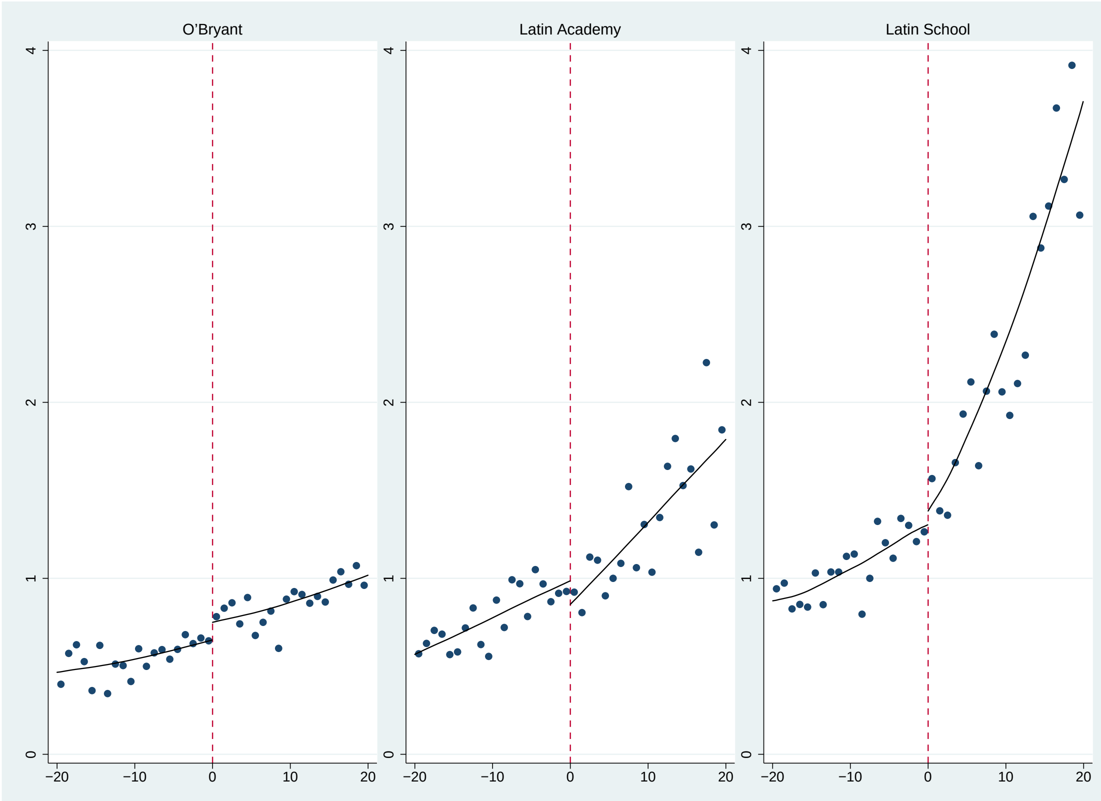

Figure 16. Number of AP Classes for 7th (1999-2004) and 9th (2001-2006) Grade Applicants in Boston

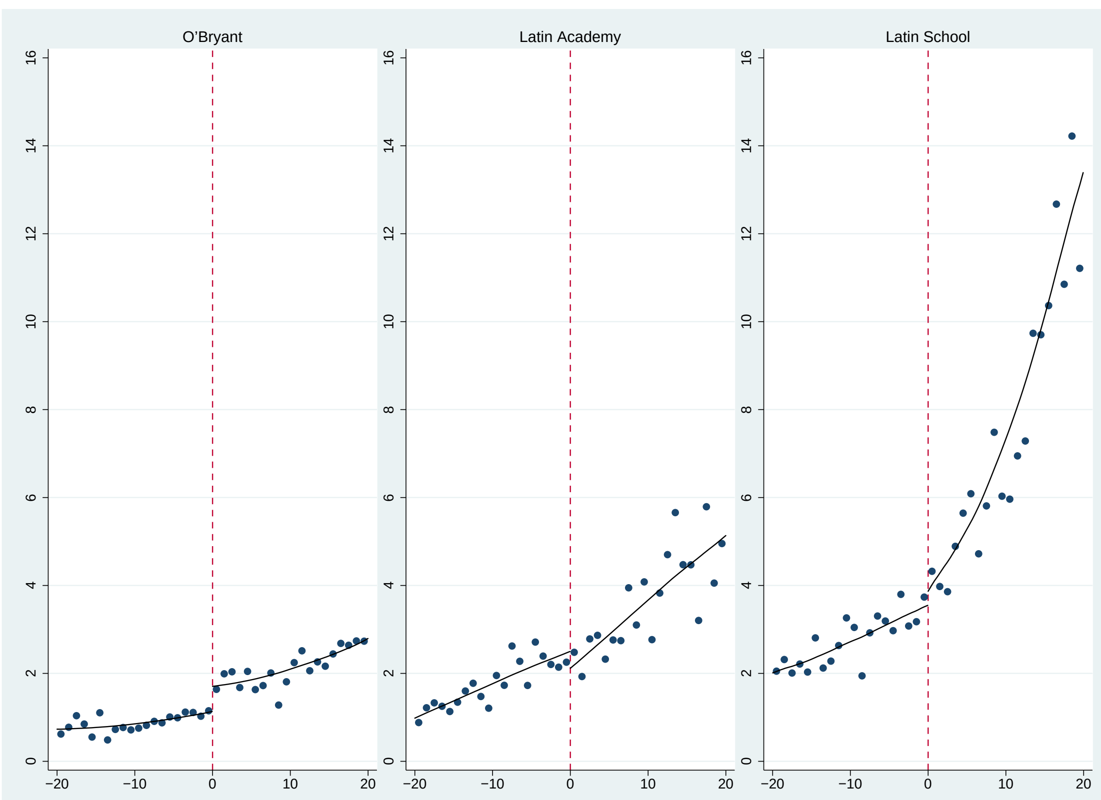

Figure 17. Sum of AP Scores for 7th (1999-2004) and 9th (2001-2006) Grade Applicants in Boston

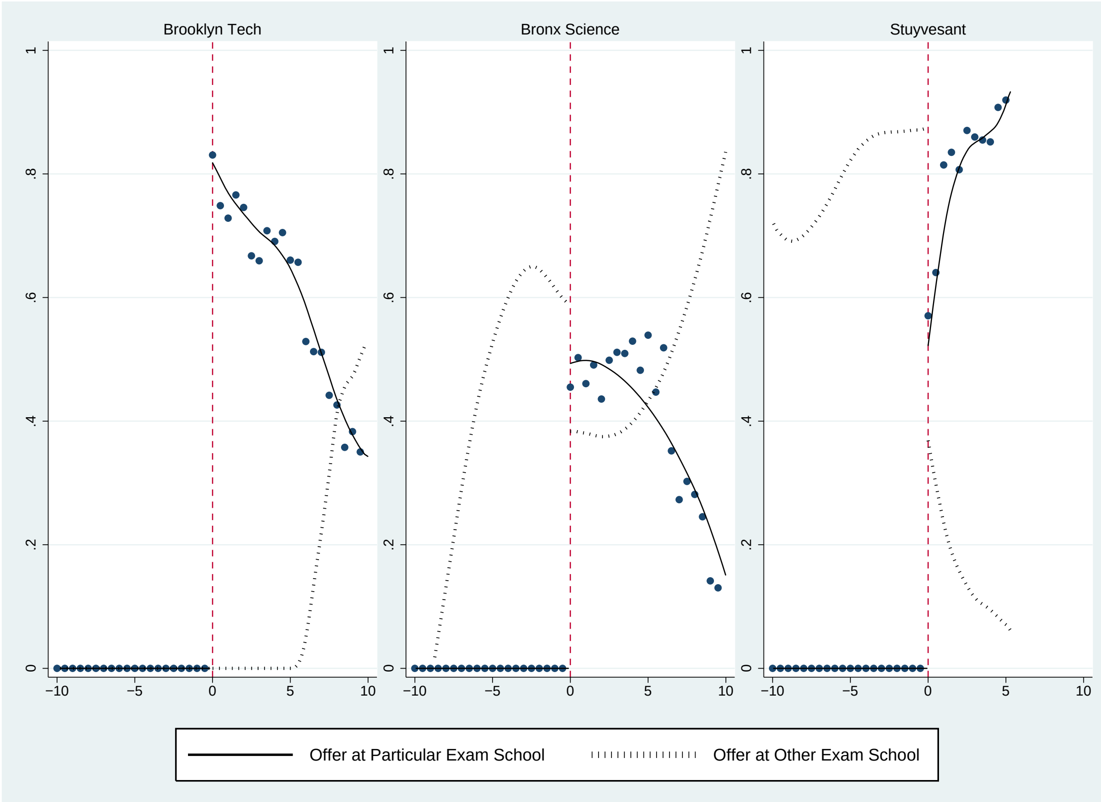

Figure 18. Offers at Each NYC Exam School for 9th Grade Applicants (2004-2007)

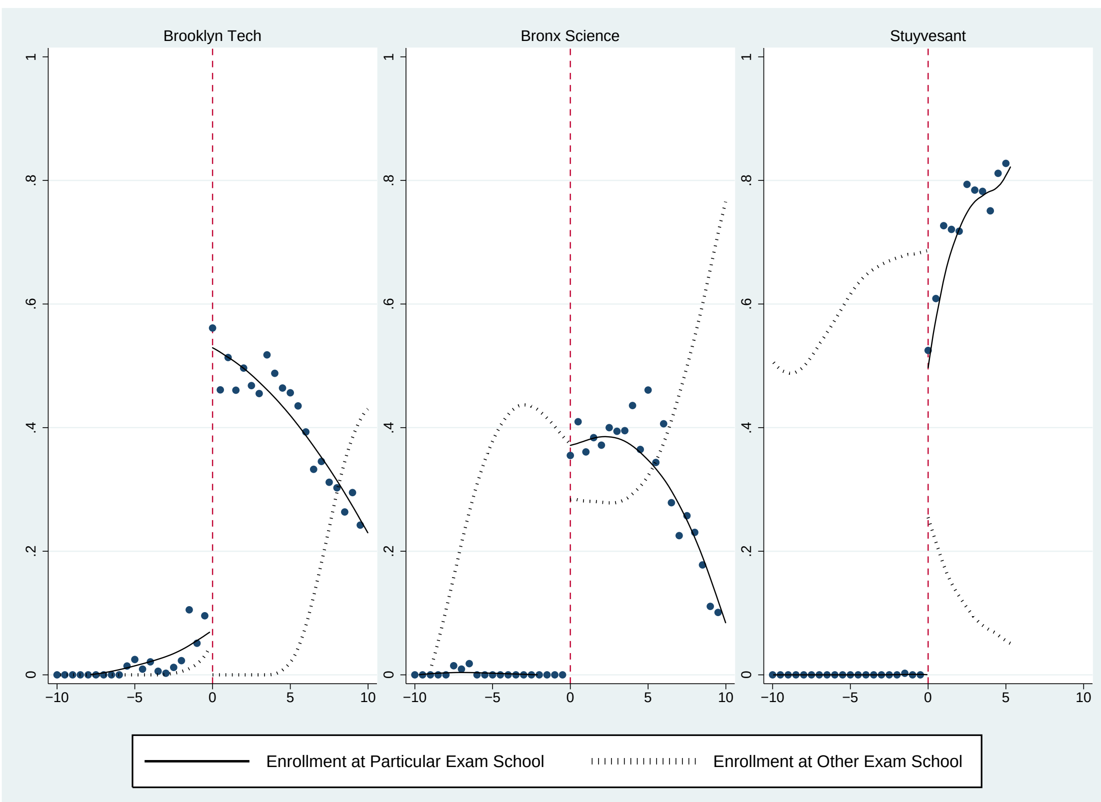

Figure 19. Enrollment at Each NYC Exam School for 9th Grade Applicants (2004-2007)

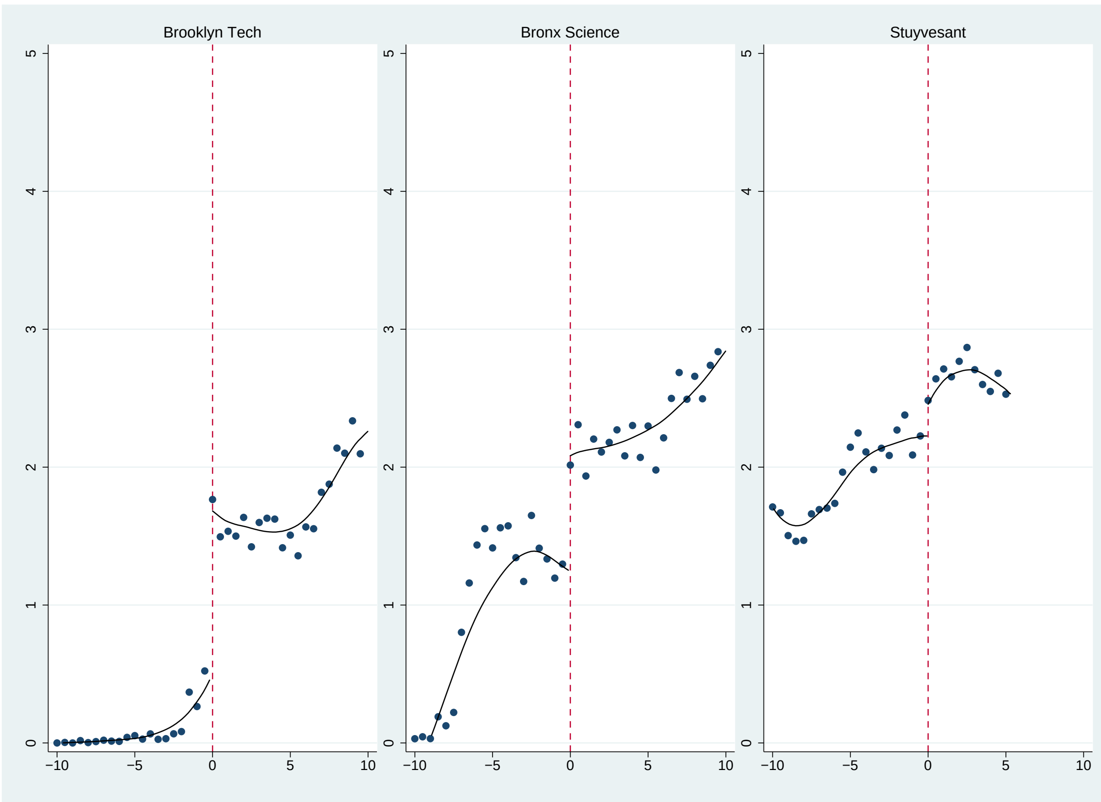

Figure 20. Years at Any Exam School for 9th Grade Applicants (2004-2007)

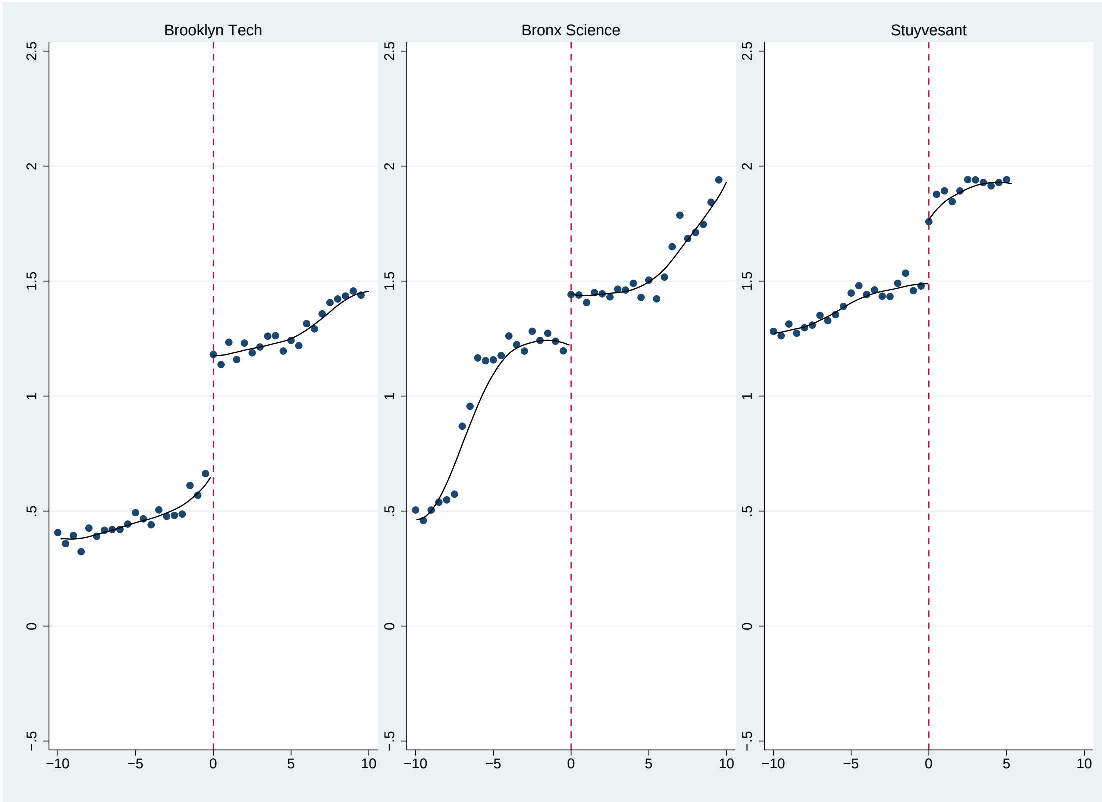

Figure 21. Average Baseline Math Score of Peers for 9th Grade Applicants (2004-2007) in NYC

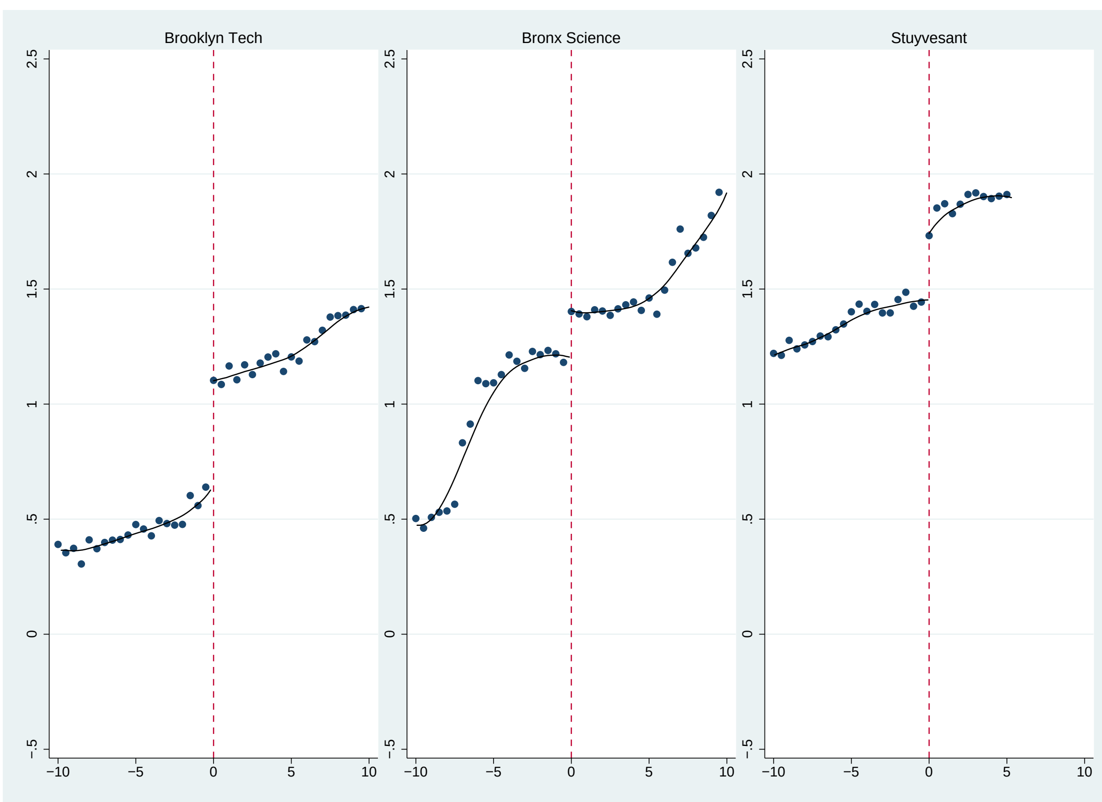

Figure 22. Average Baseline English Score of Peers for 9th Grade Applicants (2004-2007) in NYC

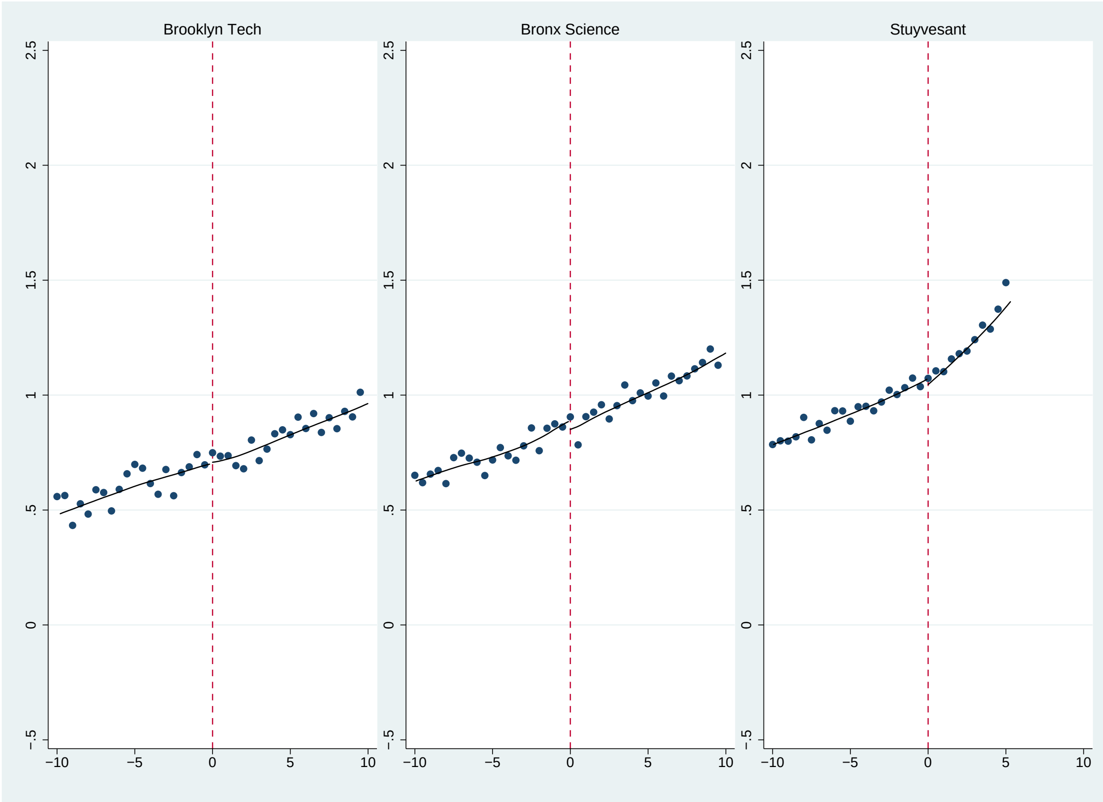

Figure 23. Advanced Math Regents Scores for 9th Grade Applicants (2004-2007) in NYC

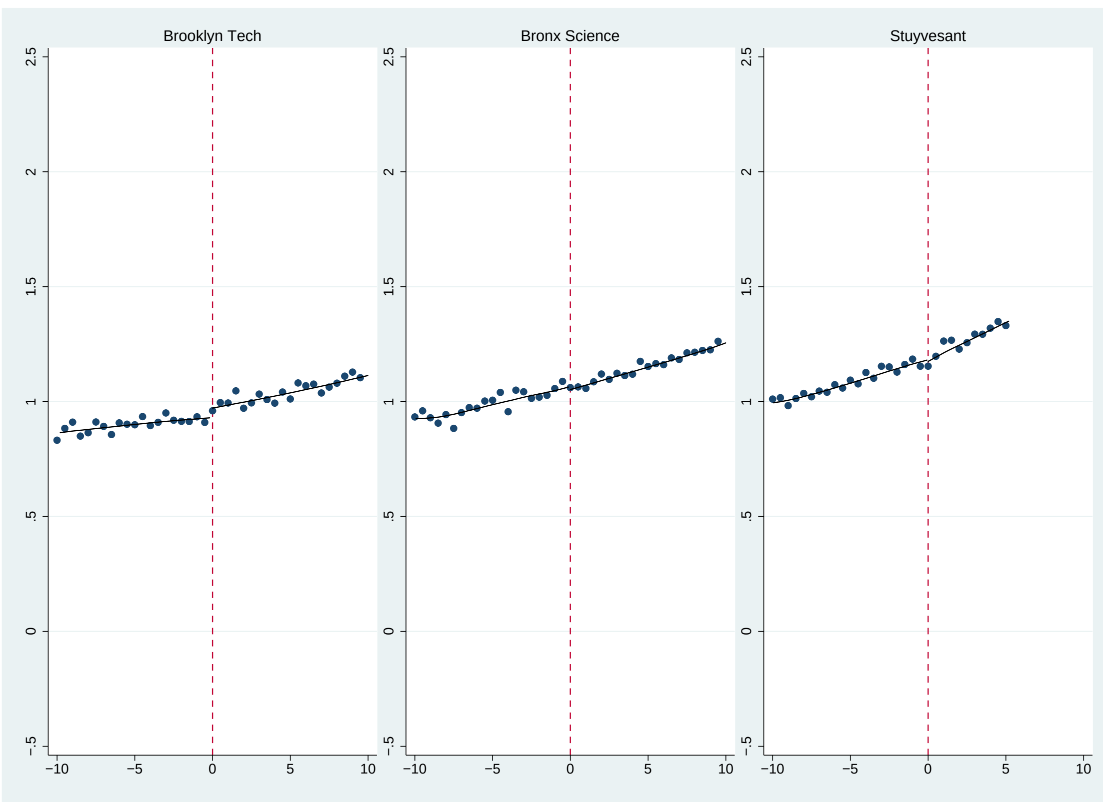

Figure 24. English Regents Scores for 9th Grade Applicants (2004-2007) in NYC

A Boston Appendix

Differential Attrition

Figures 8-11 and Tables 4-6 were constructed from samples of students who apply to exam schools and for whom we have post-admissions test scores. Students near admissions cutoffs should be similar at the time of application if the regression discontinuity is to approximate an experimental design. Subsequent attrition may lead to differences in the follow-up sample, unless the attrition process is also random. In other words, a threat to our research design is differential and selective attrition by exam offer status. For instance, students just below the cutoff may be less likely to be found than students above the cutoff if students below the cutoff leave the public school system when they do not obtain an exam offer. Differential attrition generates selection bias which in turn may compromise the estimates. One simple test for selection bias is to look at the effect offers have on the likelihood that an applicant contributes MCAS scores to our sample. If differences in follow-up rates are small, then selection bias from differential attrition is also likely to be modest.

Table A1 reports the fraction of exam school applicants with follow-up scores in the discontinuity sample. Between 76-89% of applicants have a follow-up score. This relatively high follow-up rate is likely due to the requirement that our applicant sample is limited to students who were enrolled in BPS at baseline. Follow-up differentials are estimated using both the parametric and IK approach that parallels the estimates presented in Table 4. Most of the estimated differentials for Math and English are small and not significantly different from zero using either the parametric or IK method. While the follow up differential is about 3% in the All Schools column, this difference seems unlikely to explain our findings as the most likely scenario is that relatively high achievers who miss the cutoff exit the public school system.

Discontinuities in Covariates

Another potential concern with our research design is that exam school offers are not the only variable that changes in a discontinuous manner at admissions cutoffs. If covariates other than the ranking of the applicant are used in the assignment mechanism, then these covariates may confound the interpretation of test score differences at cutoffs as being based solely on exam school offers. The fact that exam school admissions take place in the BPS central office suggest that it is unlikely that schools have much discretion in selecting which applicants obtain offers at particular schools. Nonetheless, discontinuities in the characteristics of applicants may arise in situations where the admissions process is compromised.

We briefly examine this possibility in Table A2. The table reports estimates from models which parallel the reduced-form, but each dependent variable is a covariate. Most of the entries in the table are not statistically significant, with few exceptions. 9th grade offers at Boston Latin are less likely to be black and to receive subsidized lunches according to estimates from the parametric model. These differences are no longer statistically significant with the IK method, which suggests that these are chance findings due to the small number of black and low-income students in the Latin school's grade 9 discontinuity sample. Moreover, the 9th grade applicant sample begins in 2001, after the end of racial preferences. For 7th grade, applicants above the admissions cutoff are *more* likely to be black, casting doubt on a situation where the admissions process is compromised. The p-values from the joint tests of significance of the coefficients lead us to conclude that the few significant differences in covariates seem like chance findings, and discontinuities in covariates do not explain the pattern of our results.

B New York Appendix

Differential Attrition

Table B1 reports estimates of NYC follow-rates as following Table A1. The follow-up rate is lowest for Math since many applicants take these Regents exams before 9th grade. For the other subjects, the follow-up rates range from 80-87%. For instance, 80% of students in at least one discontinuity sample have follow up Advanced Math scores, while 87% have follow-up English scores. While some of the attrition differentials are significantly different from zero at school cutoffs, in the All Schools column, the differences are relatively small. For instance, the follow-up differential for English is about 3%, a result which is only significant with the IK method. Advanced Math and Global History have attrition differences of about 4-5% in the All Schools model either with the parametric or IK method, and most of this difference is driven by students from the Brooklyn Technical cutoff.

Discontinuities in Covariates

As in Boston, the NYC admissions process is run in the central office, suggesting limited scope for school discretion in making assignments. In Table B2, we report differences in covariates on either side of cutoffs which parallel those reported in Table A2. While students are more likely to be Hispanic and less likely to obtain a free lunch near the Stuyvesant cutoff, these differences are muted with the IK method. Even though there are a few small differences at particular school cutoffs, the estimates from the All Schools model do not suggest discontinuous changes in covariates at offer cutoffs. The p-values from the joint test that all covariate discontinuities are significant supports this conclusion in the All Schools model.

Subgroups

Following Table 5, Table B3 presents reduced-form estimates for New York's high achievers and minorities. Almost no student in the discontinuity sample scores below a 65 (a failing grade) on Regents exams. However, there is still variation in score performance for the high achieving subgroups. Regents scores of 85 or higher are needed for Regents Diplomas with Advanced Designation. For Advanced Math, only 57% of students in the upper half of the baseline score distribution score 85 or more. For other outcomes, this fraction ranges from 71% for Living Environment to 94% for U.S. History. The estimates for these outcomes indicate little evidence for Regents score gains for high achievers, just as in Boston.

New York's exam schools enroll fewer blacks and Hispanic students than in Boston's exam schools. About 7.6% of enrolled students are black and 6.7% are Hispanic in New York, compared to 24% and 15%, respectively for 7th graders in Boston. The last three columns of Table B3 report reduced form estimates for black and Hispanic students in New York. Unlike Boston, the results for minorities do not support Regents achievement gains at exam schools. For instance, the impact on Advanced Math is 0.005, with standard error of 0.050.

Table A1. Boston Attrition Differentials

			Parametric (Discontinuity Sample)				Optimal Bandwidth (IK)			
				Latin	Latin	All		Latin	Latin	All
Application Grade	Test Grade	Fraction with Follow Up	O'Bryant (1)	Academy (2)	School (3)	Schools (4)	O'Bryant (5)	Academy (6)	School (7)	Schools (8)
					A. Math					
7th	7th and 8th	0.874	0.054	0.056*	-0.021	0.030	0.033	0.034*	0.020	0.029**
			(0.040)	(0.032)	(0.031)	(0.022)	(0.022)	(0.018)	(0.017)	(0.014)
		10803	6518	6384	5898	18800	6166	6302	5676	18144
7th and 9th	10th	0.759	0.066	0.065	-0.002	0.047	0.021	0.017	0.054*	0.030
			(0.050)	(0.055)	(0.052)	(0.030)	(0.029)	(0.032)	(0.029)	(0.019)
		7769	5133	4473	3886	13492	4633	3592	3426	11651
7th and 9th	7th, 8th, and 10th	0.826	0.060*	0.060**	-0.014	0.037*	0.028	0.027*	0.033**	0.029**
			(0.032)	(0.029)	(0.028)	(0.021)	(0.018)	(0.017)	(0.015)	(0.013)
		18572	11651	10857	9784	32292	10799	9894	9102	29795
					B. English					
7th	7th and 8th	0.891	0.046	0.072**	-0.022	0.032	0.041*	0.038**	0.008	0.028*
			(0.041)	(0.032)	(0.031)	(0.023)	(0.024)	(0.019)	(0.017)	(0.015)
		9472	5734	5615	5099	16448	4411	5289	4867	14567
7th and 9th	10th	0.761	0.074	0.045	0.011	0.048	0.044	0.013	0.052	0.037*
			(0.050)	(0.055)	(0.052)	(0.030)	(0.032)	(0.036)	(0.034)	(0.020)
		7769	5133	4473	3886	13492	3646	2680	2451	8777
7th and 9th	7th, 8th, and 10th	0.832	0.060*	0.061**	-0.009	0.039*	0.043**	0.030*	0.022	0.032**
			(0.033)	(0.030)	(0.028)	(0.021)	(0.020)	(0.018)	(0.016)	(0.014)
		17241	10867	10088	8985	29940	8057	7969	7318	23344

Notes: This table reports estimates of the effects of exam school offers on an indicator for non-missing outcome scores. The specificatio and estimation procedures are the same as used to construct the estimates in Table 4. The fraction with follow-up is the follow-up rate for appliants who appear in any school-speciifc discontinuity sample.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table A2. Boston Covariate Discontinuities

		Parametric (Discontinuity)				Optimal Bandwidth (IK)			
			Latin	Latin	All		Latin	Latin	All
		O'Bryant	Academy	School	Schools	O'Bryant	Academy	School	Schools
Covariate	Mean	(1)	(2)	(3)	(5)	(5)	(6)	(7)	(8)
					A. 7th Grade Applicants				
Female	0.563	-0.013	-0.034	0.010	-0.012	-0.008	0.027	0.066**	0.027
		(0.057)	(0.056)	(0.058)	(0.032)	(0.030)	(0.037)	(0.031)	(0.020)
	8628	5211	5126	4736	15073	5211	3524	4736	13471
Black	0.322	0.059	-0.017	0.057	0.033	-0.001	-0.008	0.032	0.008
		(0.055)	(0.054)	(0.049)	(0.030)	(0.034)	(0.032)	(0.029)	(0.020)
	8620	5210	5121	4729	15060	4223	4330	4015	12568
Hispanic	0.189	-0.065	-0.025	-0.078*	-0.056**	-0.024	-0.023	-0.001	-0.016
		(0.047)	(0.046)	(0.046)	(0.026)	(0.027)	(0.025)	(0.024)	(0.016)
	8620	5210	5121	4729	15060	4732	5121	4729	14582
Free Lunch	0.694	0.010	-0.110**	-0.083	-0.060**	0.001	-0.078***	-0.022	-0.038**
		(0.045)	(0.047)	(0.051)	(0.027)	(0.027)	(0.026)	(0.031)	(0.018)
	8628	5211	5126	4736	15073	4308	5126	3967	13401
LEP[‡]	0.124	-0.026	-0.018	-0.081**	-0.041*	0.002	-0.011	-0.055***	-0.019
		(0.041)	(0.038)	(0.034)	(0.022)	(0.023)	(0.022)	(0.021)	(0.014)
	7948	4782	4718	4340	13840	4614	4589	3268	12471
SPED[¥]	0.017	-0.010	0.009	-0.003	-0.002	-0.016	-0.007	0.011	-0.003
		(0.026)	(0.016)	(0.018)	(0.012)	(0.012)	(0.008)	(0.009)	(0.006)
	5125	3081	3060	2818	8959	3081	3060	2818	8959
Joint p-value		0.717	0.342	0.052	0.075	0.858	0.066	0.021	0.170
					B. 9th Grade Applicants				
Female	0.602	-0.052	-0.118	-0.090	-0.075	0.014	-0.022	-0.051	-0.009
		(0.084)	(0.121)	(0.167)	(0.064)	(0.054)	(0.066)	(0.095)	(0.040)
	2524	1995	1367	1065	4427	1512	1367	577	3456
Black	0.415	0.040	0.006	-0.306**	-0.022	-0.003	0.002	-0.097	-0.019
		(0.084)	(0.120)	(0.128)	(0.062)	(0.048)	(0.073)	(0.075)	(0.037)
	2521	1992	1364	1062	4418	1935	900	954	3789
Hispanic	0.232	-0.073	0.118	0.115	0.006	0.001	0.005	0.053	0.009
		(0.071)	(0.109)	(0.135)	(0.056)	(0.039)	(0.057)	(0.076)	(0.032)
	2521	1992	1364	1062	4418	1992	1364	636	3992
Free Lunch	0.789	-0.051	0.002	-0.283**	-0.072	0.005	0.101*	0.044	0.042
		(0.069)	(0.096)	(0.128)	(0.053)	(0.043)	(0.053)	(0.076)	(0.032)
	2524	1995	1367	1065	4427	1714	1367	610	3691
LEP[‡]	0.118	-0.001	-0.035	-0.009	-0.011	0.020	-0.008	0.034	0.016
		(0.057)	(0.091)	(0.090)	(0.042)	(0.030)	(0.051)	(0.052)	(0.024)
	2524	1995	1367	1065	4427	1995	1063	708	3766
SPED[¥]	0.026	0.028	0.036*	0.026	0.030	-0.003	-0.021	-0.036	-0.012
		(0.038)	(0.020)	(0.027)	(0.024)	(0.020)	(0.014)	(0.029)	(0.014)
	1363	1062	772	614	2448	1062	538	395	1995
Joint p-value		0.841	0.474	0.049	0.624	0.997	0.414	0.579	0.844

Notes: This table reports estimated discontinuities in covariates using models like those used to construct the reduced form estimates in Table 4. The joint p-value is from a F-test looking at all covariate discontinuities at once.

[‡] LEP only available beginning in year 1998.

[¥] SPED only available for years 1998-2004.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table B1. New York Attrition Differentials

	Fraction with Follow Up	Parametric (Discontinuity Sample)				Optimal Bandwidth (IK)			
		Brooklyn Tech	Bronx Science	Stuyvesant	All Schools	Brooklyn Tech	Bronx Science	Stuyvesant	All Schools
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Math	0.535	-0.087*	0.058	0.059	0.012	0.009	0.017	0.050*	0.025*
		(0.049)	(0.049)	(0.046)	(0.028)	(0.026)	(0.025)	(0.026)	(0.015)
	17713	7622	6829	7553	22004	6707	6829	6786	20322
Advanced Math	0.804	0.127***	0.039	-0.010	0.049**	0.069***	0.031	0.025	0.043***
		(0.043)	(0.039)	(0.033)	(0.022)	(0.021)	(0.021)	(0.016)	(0.011)
	17713	7622	6829	7553	22004	7622	6829	7553	22004
English	0.874	0.02	0.032	0.028	0.027	0.036*	0.031	0.025	0.031***
		(0.038)	(0.036)	(0.031)	(0.020)	(0.019)	(0.021)	(0.016)	(0.011)
	13147	5867	5268	5695	16830	5867	4834	5695	16396
Global History	0.866	0.082**	0.033	0.035	0.049***	0.062***	0.036*	0.020	0.039***
		(0.037)	(0.033)	(0.028)	(0.019)	(0.022)	(0.020)	(0.017)	(0.011)
	17713	7622	6829	7553	22004	5782	5396	5739	16917
US History	0.814	0.039	0.032	0.008	0.026	0.043	0.039	0.009	0.030**
		(0.046)	(0.044)	(0.035)	(0.024)	(0.030)	(0.024)	(0.022)	(0.014)
	13147	5867	5268	5695	16830	3706	5086	4342	13134
Living Environment	0.797	0.016	0.038	0.024	0.026	0.033*	0.011	0.003	0.017
		(0.041)	(0.038)	(0.035)	(0.022)	(0.020)	(0.021)	(0.017)	(0.011)
	17713	7622	6829	7553	22004	7622	6829	7553	22004

Notes: This table reports estimates of the effect of exam school offers on indicators for non-missing outcome scores. Models and estimation procedures are the same as for Table 10. The fraction with follow-up is the follow-up rate for appliants who appear in any school-speciifc discontinuity sample.

*significant at 10%; **significant at 5%; ***significant at 1%.

Table B2. New York Covariate Discontinuities

	Mean of Variable	Parametric (Discontinuity)				Optimal Bandwidth (IK)			
		Brooklyn Tech	Bronx Science	Stuyvesant	All Schools	Brooklyn Tech	Bronx Science	Stuyvesant	All Schools
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	0.468	-0.039	-0.015	-0.004	-0.018	0.001	-0.011	-0.018	-0.009
		(0.049)	(0.049)	(0.046)	(0.028)	(0.025)	(0.029)	(0.025)	(0.015)
	17713	7622	6829	7553	22004	7622	5848	7469	20939
Black	0.106	-0.056	-0.001	0.033	-0.006	-0.033*	-0.001	0.015	-0.005
		(0.037)	(0.031)	(0.022)	(0.017)	(0.018)	(0.016)	(0.011)	(0.009)
	17713	7622	6829	7553	22004	7622	6829	7553	22004
Hispanic	0.104	0.010	-0.011	0.048**	0.017	0.030*	-0.015	0.025*	0.011
		(0.035)	(0.030)	(0.021)	(0.017)	(0.017)	(0.015)	(0.015)	(0.010)
	17713	7622	6829	7553	22004	7622	6829	4626	19077
Free Lunch[#]	0.668	-0.064	0.061	-0.080*	-0.030	-0.008	0.039*	-0.042*	-0.001
		(0.046)	(0.046)	(0.043)	(0.026)	(0.024)	(0.023)	(0.024)	(0.014)
	17713	7622	6829	7553	22004	7339	6829	6972	21140
LEP	0.005	0.009	-0.002	0.002	0.003	0.004	-0.003	-0.002	-0.001
		(0.006)	(0.004)	(0.006)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
	17713	7622	6829	7553	22004	5330	6829	7553	19712
Joint test:									
p-value		0.230	0.793	0.056	0.590	0.233	0.404	0.136	0.819

Notes: This table reports estimated discontinuities in covariates using models like those used to construct the reduced form estimates in Table 11. The joint p-value is from an F-test looking at all covariate discontinuities at once.

* significant at 10%; ** significant at 5%; *** significant at 1%.

Table B3. New York Reduced Form Estimates for Subgroups

	High Baseline Scores						Black and Hispanic		
	Upper Half			Upper Quartile					
	Baseline Mean	Proportion above 85 on Regents	IK Estimates	Baseline Mean	Proportion above 85 on Regents	IK Estimates	Baseline Mean	Proportion above 85 on Regents	IK Estimates
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Math	1.425	0.872	-0.053***	1.495	0.927	-0.038*	1.253	0.723	0.027
			(0.018)			(0.021)			(0.047)
	9049		10471	6744		7570	2006		2381
Advanced Math	0.937	0.571	-0.019	1.027	0.637	-0.006	0.600	0.346	0.005
			(0.021)			(0.019)			(0.050)
	13854		17261	10990		13446	2596		3126
English	1.097	0.816	0.013	1.161	0.869	0.008	0.975	0.716	0.021
			(0.015)			(0.018)			(0.043)
	10772		12624	7703		9018	2298		2947
Global History	1.236	0.841	-0.026	1.284	0.879	-0.025	1.085	0.723	-0.064
			(0.016)			(0.015)			(0.043)
	14251		12296	10365		10858	2994		3060
US History	1.151	0.937	0.003	1.186	0.955	-0.003	1.037	0.871	-0.013
			(0.016)			(0.015)			(0.033)
	10034		10096	7216		8114	2092		2522
Living Environment	1.335	0.711	-0.020*	1.380	0.754	-0.020	1.178	0.563	-0.074**
			(0.012)			(0.014)			(0.036)
	13610		16975	10585		13242	2612		3092

Notes: This table reports reduced form estimates for students with high baseline scores and for minorities. Baseline means and the proportion of applicants above 85 are computed for those who belong to at least one discontinuity sample. Math scores are from either Regents Math A (Elementary Algebra and Planar Geometry) or Integrated Algebra I. Advanced Math scores are from either Regents Math B (Intermediate Algebra and Trigonometry) or Geometry. The Table reports IK estimates using bandwidths computed as for the all schools model in Table 10.

* significant at 10%; ** significant at 5%; *** significant at 1%

C Data Appendix: Boston

Boston Public Schools is the source for four datasets: the exam school application file, the enrollment file which contains student demographic and school attendance information, the Massachusetts Comprehensive Assessment System (MCAS) test score file, and the College Board test file which contains PSAT, SAT and AP scores. This appendix describes these data sets and the procedures used to construct the analysis sample.

C.1 Data Sets

Exam school application file

Data description and sample restrictions

The exam school application file contains a record for each student consisting of a registration number, application year, grade, date of birth, preferences over three exam schools, and scores on the ISEE verbal, quantitative, reading and math sections. Each record also includes the rank of each student by the exam schools on their preference list and the school where the student receives an offer (if any). This dataset covers students in grades 7, 9, and 10 and application years 1997-2008. Since there are a small number of grade 10 applicants, we kept students applying for grades 7 and 9 only.

Table C1 indicates the steps involved in processing the exam application file. We excluded duplicate observations, applicants from private schools and those who did not rank or were not ranked by any exam school. We also dropped students who obtained an offer at an exam school that is not on their preference list.

Coding the offer variable

For each applicant, the exam school application file indicates whether the student receives an offer at one school on their preference list. For a given application year, grade, and school, we computed the lowest-ranked student to obtain an offer from that school. Each student is then coded as obtaining an offer at an exam school if her score is above this minimum cutoff for any school that is on her preference list.

Enrollment file

Data description

The BPS enrollment file spans school years 1995-1996 through 2008-2009. Each record contains an end-of-year (June) snapshot for each student enrolled in Boston Public Schools, with unique student identifier (the BPS ID), the student's grade and school, and demographic information.

Coding of demographics and attendance

The variables of interest in the enrollment file are grade, year, date of birth, sex, race, special education (SPED) and limited English proficiency (LEP) status, subsidized lunch eligibility, and school. Students are coded as attending an exam school if their year-end enrollment is at an exam school. Years at an exam school is the total number of years where a student is at an exam school at the end of the year. We transformed the enrollment file into a wide-format layout for each student where we compute the grade and exam school years attended for a given year. Finally, we kept only students that attend Boston Public Schools in 6th or 8th grade and use their demographic information from that year.

MCAS test file

Data description and sample restrictions

Each record in the MCAS test file contains a student identifier (BPS ID) and scores on MCAS tests in a given year. We used data from school years 1999-2000 through 2008-2009. The scores we look at are Math and English Language Arts (ELA) for grades 4-10. The MCAS test file contains raw scores for all BPS test takers for Math, English Language Arts, Writing, and Science. As shown in Table C2, the number of grades tested has increased over time. MCAS Math for grade 8 was the first examination offered in 1999. By the end of our data, there are tests for Math and English tests at grade 7, 8, and 10. Baseline scores for grade 7 applicants are from 4th grade MCAS exams. For 9th grade applicants, baseline Math is from 8th grade Math and baseline English is from 7th grade English, since the 8th grade English exam is first offered in 2006.

We standardized scores to have mean zero and standard deviation one within a subject-grade-year among all test-takers in Boston Public Schools. When there is more than one test score for a student for a particular subject, we used the earliest available one. We converted the file to a wide-format structure where each row contains all available MCAS scores for a student.

College Board test file

Data description and sample restrictions

The College Board provides BPS with reports on the test performance of all BPS test-takers from 2004-2005 through 2009-2010. These files come with the name, date of birth, address, gender, school of test, and test year for each exam. BPS matched the PSAT file for October 2004 and October 2005, the SAT file which is available from 2005-2009, and the Advanced Placement test file, available from 2005-2009. The timing of these tests for our applicant cohorts is shown in Table C2.

The PSAT file is not matched to BPS student IDs for years 2006-2009, so we had to link College Board files to BPS files for these years. The address information in the College Board file is entered by the test-taker and does not immediately concord with the BPS address system. There also appear to be small errors in the date of birth in the College Board file for similar reasons. Our procedure to match these files to the BPS registration files is as follows. First, we take all unique year, date of birth, gender, school of test, and zip code matches between the BPS registration file and the College Board PSAT file. Among the remaining unmatched PSAT records, we take all unique year, date of birth, gender, and school of test matches between the two files. Finally, for the remaining unmatched PSAT test records, we hand-matched the records for these four years to the closest record in the registration file, attempting to correct mismatches due to address misspellings or typos in the date of birth.

BPS students take AP exams across a range of subjects. The tests with 500+ or more takers are Calculus AB, Statistics, Biology, Chemistry, Physics B, English Language and Composition, English Literature and Composition, European History, US Government and Politics, US History, Microeconomics, Macroeconomics, and Spanish Language. The other tests are Art History, Art: Drawing, Art: 2D Design, Art: 3D design, Chinese Language and Culture, Computer Science A, Computer Science AB, Environmental Science, French Language, French Literature, German Language, Comparative Government and Politics, Latin: Vergil, Latin: Literature, Calculus BC, Music Theory, Physics C: Mechanics, Physics C: Electricity and Magnetism, Spanish Literature, and World History.

We standardized the PSAT and SAT scores to have mean zero and standard deviation one within a year among all test-takers in Boston Public Schools. When there is more than one test score for a student, we used the earliest available one. We only use applicant cohorts for whom we might expect to observe PSAT, SAT or AP exams; these are summarized in Table C2.

C.2 Matching Data Sets

Match from the MCAS test file to the enrollment file

Match criteria

The MCAS test file and enrollment files are merged by grade, year, and BPS ID. Any test record that is not be matched to the enrollment file is dropped.

The exam applicant file is matched to the enrollment/MCAS file using an auxiliary table that links exam registration number to BPS ID. This table provides a BPS ID for each exam registration number. For a small number of cases, an exam registration number is matched to more than one BPS ID. In these cases, we matched to the registration number to the BPS ID where the date of birth is the same between the exam applicant and enrollment file.

Match rates

Table C3 reports match rates from exam applicant file to the enrollment/MCAS file. The overall match rate is 96.6 percent (13,730 out of 14,212) for grade 7 applicants and 99.6 (6,155 out of 6,181) for grade 9 applicants. The match rate for offered students in grade 7 is 96.9%, while the match rate for students who were not offered is 96.3%. The lower match rates come from earlier application years 1997-2000. The match rate for not offered is larger than for offered for three of these years, and the differences in match rates are small. For grade 9, where the application cohorts start in 2001, the match rate for offered students is 99.9%, while for non-offered it is 99.5%. Applicants who are not matched to the enrollment file at baseline are dropped as are applicants who enrolled in an exam school before application. This latter restriction only impacts grade 9 applicants, as can be seen comparing columns (7) and (8) of Table C1.

C.3 Construction of the Analysis Sample

The size of the final analysis sample is presented in column (8) of Table C1.

Stacking grades

Some of the analysis stacks grades and includes multiple test scores for individual students. For each student in an application year, Table C4 reports the number of students with at least one follow up test score (column (2)). It also presents the number of test scores expected for each cohort and the number of test scores observed for both Math and English. For example, a 7th grade applicant for the 2005-2006 school year contributes Math scores in grade 7 (Spring 2006), 8 (Spring 2007), and 10 (Spring 2009). Hence, we expect 3,285 Math scores from the 1,095 applicants for this cohort, and we observe at least one score for 1,001 students, which corresponds to a total of 2,650 student-score observations. On the other hand, a 7th grade applicant for the previous school year contributes one fewer test score (no grade 7 Math). Table C5 shows a related analysis of expected follow up for PSAT, SAT and AP scores.

D Data Appendix: NYC

The New York City Department of Education is the source for three datasets: the exam school application and Student Enrollment Office (formerly, OSEPO) files which contains demographic information, the registration file which contains school attendance information, and the NYSED and Regents test score file. This appendix describes these data sets and details the procedures used to construct the analysis sample.

D.1 Data

Exam school application and Enrollment Office files

Data description and sample restrictions

The exam school application file is maintained by the Enrollment Office, which runs high school admissions. All applicants must take the Specialized High School Admissions Test (SHSAT) to apply to an exam school. On test day, students also submit a ranking of exam schools. At a later date, students are also required complete a New York City Public High School Admissions Application and submit it to their guidance counselor.

Several Enrollment Office files are used in the analysis. The first contains a record for each student indexed by their ID number (OSISID) and their score on the SHSAT. For each student, the exam school offer file contains a list of the schools ranked and an indicator for the school at which the student obtains an offer (if any). The Enrollment Office student file has demographic information such as grade, sex, race, home language code, and borough of residence for each student. There are also separate files indicating special education and limited English proficiency for each student. Each file for a given application year contains an OSISID number for each student, which allows us to merge the files together.

Registration and enrollment files

Data description and sample restrictions

The NYC registration file is from the Office of School Performance and Accountability and is available as part of data underlying school progress reports. The registration and enrollment cover all public school students in grades 9 to 12 for school years 2002-2003 through 2008-2009. This data set includes each student's NYC ID, grade, and current school as of October in the school year. The registration data are used to determine whether and for how many years a student enrolls in an exam school, where a student who is enrolled in October is counted as enrolling for the entire year. Starting in 2004-2005, there is a separate file which contains a list of all students who obtain a subsidized school lunch in that year. This variable is used to code subsidized lunch status for applicants using the application year. For applicants in 2003-2004, 2004-2005, and 2005-2006, we used the lunch status record from 2004-2005. For application cohort in 2006-2007, we used the lunch status record from 2005-2006.

Table D1 indicates the steps involved in processing the exam application file and merging it with the Student Enrollment files. From the file of exam applicants, we eliminated private school applicants (based on whether their OSISID starts with the letter "A") and those who do not submit a New York City Public School Admissions Application (based on the Round 1 HS ranking file). The 4,000-5,000 private school applicants are excluded because these students do not have a NYC ID at the time of application, they do not have baseline information, and the relevant counterfactual for this population is unlikely to be a regular NYC public high school. We also excluded students who did not rank at least one of the three original academic exam schools: Bronx Science, Brooklyn Technical, and Stuyvesant.

Baseline test files

Data description and sample restrictions

The NYC Department of Education also provided us with NYSED grade 8 standardized exams in Math and English Language Arts for all public school students for years 2002-2003 through 2007-2008. These tests are taken in the winter of grade 8 and are required of all public school students in the state. These tests serve as our baseline Math and English scores.

Regents test file

Data description and sample restrictions

The NYC Regents test file contains the date and raw score for each tested student. Regents exams are mandatory state examinations where performance determines whether a student is eligible for a high school diploma in New York. There are Regents examinations in English, Global History, US History, and multiple exams in Mathematics and Science. A Regents exam typically has a multiple choice section and a long answer or essay component, and each exam usually lasts for three hours. The English exam, however, consists of two three-hour pieces over two days. The exam has a locally graded component and Dee, Jacob, McCrary, and Rockoff (2011) illustrate how test scores bunch near performance thresholds.

The New York State Board of Regents governs and designs the Regents exams. Starting in 2005, they started to modify the Mathematics exams. At the beginning of our sample, the two Mathematics examinations were Elementary Algebra and Planar Geometry (Math A) and Intermediate Algebra and Trigonometry (Math B). Two new mathematics examinations, Integrated Algebra I (Math E) and Geometry (Math G), have since been phased in. Since students typically either take Math A or Math E, we focus on the score on the test taken first, taking the Math A score when both are contemporaneous. Likewise, students typically either take Math B or Math G, so we focus on the score which comes first, taking the Math B score when both are contemporaneous. We denote the first test outcome as 'Math' and the second outcome as 'Advanced Math'. There are Regents science exams in Earth Science, Living Environment, Chemistry, and Physics. The science outcome we focus on is Living Environment because it is the only Regents science exam required to obtain a state high school diploma.

In Table D3, for each test, we report the number of applicants and the number of test scores we observe. English and U.S. History Regents exams are typically taken in 11th grade. For the 2006-2007 applicants, we expect to observe these scores in the 2009-2010, a year after the Regents test score file's last date. Even though there are a small number of students who take these exams before the 11th grade, we do not examine Regents English and U.S. History outcomes for the 2006-2007 applicant cohort, since the vast majority do not.

Since students may take Regents exams multiple times, there can be multiple test scores per student in the Regents test file. Table D3 presents the number of students who have taken each exam more than once among the exam applicant sample. This fraction is about 10%, with slightly higher retake rates for Math and Global History. Some students may also take Regents exams before exam school enrollment. Table D3 shows the fraction of students who take exams before enrolling in an exam school. A large fraction of exam school applicants take Math before enrolling. Most Regents exams are offered in January, June, and August, with most students usually taking tests in June.

For some subjects, such as Global History, most applicants take the test at the end of 10th grade. For other tests, such as Math (Math A or E), many students take the exam before entering high school and some students take the exam multiple times. The exact number of students who take the exam before 9th grade, the number who take the exam more than once after 9th grade, and the number who take the exam on a date other than the most common date are presented in Table D3. For each test where there is a re-take, we only use the first test outcome.

For each test, students who have scores before the 9th grade are omitted because they tested prior to potential exam enrollment. If a student takes the test more than once after 9th grade, we used the test score from the earliest date. There are a small number of cases where there is more than one score on the same date, and this date is the first date after entering 9th grade. In some of these cases, there are two different test codes, where one code ends with a "2." We used the score corresponding to the test that does not end with a "2." Otherwise, we treated the score as missing.

For each subject, we standardized scores to have mean zero and standard deviation one within year-semester-subject among the universe of students: 8th graders from public school who participated in Round 1 of the HS Admissions process, have valid demographic information, and did not take the SHSAT test in a previous year.

D.2 Matching Data Sets

Match between Exam Applicant file and Enrollment Office student file

We matched the exam application to the student file using the OSISID. Table D2 shows the match rates. Nearly every student who has applied to an exam school can be matched to the corresponding Enrollment Office student file. The student file allows us to identify whether an applicant is in grade 8 or 9. Since there are a limited number of 9th grade applicants for grade 10 spots, we kept only students applying for grade 9. Finally, our sample is limited to first-time SHSAT takers.

Coding the offer variable

For each exam school and applicant year, the exam school offer file indicates the school at which a student obtains an offer (if any). The offered school is the student's most preferred school where a student has a high enough SHSAT score. For each school, we computed the minimum score needed to obtain an offer at each exam school. We coded anyone with an SHSAT score above the lowest score offered as having received an offer.

Coding Attendance

Students are coded as attending an exam school if they are enrolled at an exam school in the registration file.

D.3 Construction of the Analysis Sample

After processing the exam application file, we next matched it to the registration file for grade 9. An exam applicant may not match to the registration file if she leaves New York City's Public Schools following application. Such an applicant would not contribute any follow up scores.

To generate the final analysis dataset, we merged the student registration and test file with the exam application file. The exam application file contains the NYC ID, a list of exam schools that students have ranked, and the student's raw SHSAT test score. This data spans four cycles of admissions years: 2003-04 through 2006-07.

Next, we merged baseline scores for students for whom they are available. Finally, we merged the dataset of cleaned Regents outcome scores. For each test, we compute the implied years of exam school attendance based on the test date and enrollment status. If a student took a Regents test in the fall semester, we computed years assuming the exam date is January 31st. Otherwise, we compute years assuming the exam date is June 1st. The resulting file is our analysis sample. An applicant who is matched to the registration file for grade 9 may not contribute follow up scores if the applicant leaves New York City's Public Schools before taking a Regents exam. The last column of Table D1 indicates the sample of students who contribute at least one follow up score.

Table C1. Processing of Boston Exam School Application Data

Application Year	Total number of records (1)	Excluding duplicate observations (2)	Excluding applicants from private schools (3)	Excluding students who did not rank an exam school (4)	Excluding students who are not ranked by an exam school (5)	Excluding students who obtain an offer at a school they do not rank (6)	Excluding students not matched to Boston Public Schools at baseline (7)	Excluding students previously enrolled in exam school (8)	Excluding students with no observed outcome MCAS test scores (9)
					A. 7th Grade Applicants				
1997	2376	2375	1319	1299	1299	1299	1185	1185	1009
1998	2264	2264	1237	1215	1215	1215	1081	1081	917
1999	2353	2353	1353	1307	1307	1307	1180	1180	1000
2000	2283	2283	1252	1165	1165	1165	1125	1125	1032
2001	2317	2317	1299	1196	1196	1196	1193	1193	1100
2002	2365	2365	1304	1237	1236	1236	1235	1235	1118
2003	2494	2494	1386	1251	1251	1251	1240	1240	1127
2004	2217	2217	1206	1174	1174	1174	1172	1172	1083
2005	2062	2062	1116	1105	1105	1099	1095	1095	1001
2006	2079	2079	1184	1166	1166	1161	1158	1158	1052
2007	1992	1992	1086	1081	1080	1073	1068	1068	974
2008	1874	1874	1050	1049	1040	1036	998	998	898
All Years	26676	26675	14792	14245	14234	14212	13730	13730	12311
					B. 9th Grade Applicants				
2001	1520	1520	863	787	787	787	783	680	496
2002	1607	1607	876	829	828	828	826	755	553
2003	1750	1750	951	812	812	812	809	727	546
2004	1723	1723	936	918	918	918	912	815	631
2005	1630	1630	936	924	924	924	918	832	642
2006	1729	1729	992	981	981	981	977	889	677
2007	1684	1683	945	936	931	931	930	842	612
All Years	11643	11642	6499	6187	6181	6181	6155	5540	4157

Notes: This table summarizes the steps going from the raw application data to the analysis sample.

Table C2. Data Structure and Test Outcomes for Boston

Application Year	Math 7 (1)	Math 8 (2)	Math 10 (3)	English 7 (4)	English 8 (5)	English 10 (6)	PSAT (7)	SAT (8)	AP (9)
				A. 7th Grade Applicants					
1997		1999							
1998		2000							
1999		2001	2003			2003		2005	2005
2000		2002	2004	2001		2004	2004	2006	2006
2001		2003	2005	2002		2005	2005	2007	2007
2002		2004	2006	2003		2006	2006	2008	2008
2003		2005	2007	2004		2007	2007	2009	2009
2004		2006	2008	2005	2006	2008	2008	2010	2010
2005	2006	2007	2009	2006	2007	2009	2009		
2006	2007	2008		2007	2008				
2007	2008	2009		2008	2009				
2008	2009			2009					
				B. 9th Grade Applicants					
2001			2003			2003		2005	2005
2002			2004			2004	2004	2006	2006
2003			2005			2005	2005	2007	2007
2004			2006			2006	2006	2008	2008
2005			2007			2007	2007	2009	2009
2006			2008			2008	2008	2010	2010
2007			2009			2009	2009		

Notes: This table reports the applicant cohorts and test year outcomes. Application year refers to the fall of application year, while test outcome year refers to the spring of year. Test outcomes are available based on the schedule of the MCAS and availability of SAT, PSAT and AP score outcomes.

Table C3. Match from Boston Exam Application to Enrollment Data

Application Year	Number of Students (1)	Fraction with Match: Total (2)	Offered (3)	Not Offered (4)
		A. 7th Grade Applicants		
1997	1299	0.912	0.906	0.917
1998	1215	0.890	0.888	0.891
1999	1307	0.903	0.919	0.890
2000	1165	0.966	0.958	0.972
2001	1196	0.997	0.996	0.998
2002	1236	0.999	1.000	0.999
2003	1251	0.991	0.996	0.987
2004	1174	0.998	1.000	0.997
2005	1099	0.996	0.996	0.996
2006	1161	0.997	0.995	1.000
2007	1073	0.995	1.000	0.991
2008	1036	0.963	0.980	0.946
All Years	14212	0.966	0.969	0.963
		B. 9th Grade Applicants		
2001	787	0.995	1.000	0.993
2002	828	0.998	1.000	0.997
2003	812	0.996	1.000	0.995
2004	918	0.993	1.000	0.992
2005	924	0.994	0.992	0.994
2006	981	0.996	1.000	0.994
2007	931	0.999	1.000	0.999
All Years	6181	0.996	0.999	0.995

Notes: This table provides summary statistics on the match between the exam school application data and the Boston Public School enrollment file. The sample in column (1) is the sample in column (6) of Table C1.

Table C4. Test Outcome Data for Boston Exam School Applicants

Application Year	Number of students (1)	Number with an observed test score (2)	Number of Math test scores expected (3)	Math test scores observed (4)	Number of English test scores expected (5)	English test scores observed (6)
			7th Grade			
1997	1185	1009	1185	1017	0	9
1998	1081	917	1081	978	0	67
1999	1180	1000	2360	1765	1180	800
2000	1125	1032	2250	1776	2250	1792
2001	1193	1100	2386	1843	2386	1897
2002	1235	1118	2470	1894	2470	1945
2003	1240	1127	2480	1897	2480	2006
2004	1172	1083	2344	1842	3516	2890
2005	1095	1001	3285	2650	3285	2650
2006	1158	1052	2316	2039	2316	2038
2007	1068	974	2136	1884	2136	1879
2008	998	898	998	895	998	897
All Years	13730	12311	25291	20480	23017	18870
			9th Grade			
2001	680	496	680	496	680	495
2002	755	553	755	551	755	550
2003	727	546	727	545	727	543
2004	815	631	815	621	815	630
2005	832	642	832	630	832	636
2006	889	677	889	662	889	673
2007	842	612	842	603	842	610
All Years	5540	4157	5540	4108	5540	4137

Notes: This table summarizes the observed test score outcomes for exam school applicants. The sample is restricted to students in column (8) of Table C1.

Table C5. Matching of College Board Test Outcome Data for Boston Applicants

Application Year	Number of applicants (1)	Number with an observed PSAT test score (2)	Number with an expected PSAT test score (enrolled as of grade 11) (3)	Number with an observed SAT test score (4)	Number with an expected SAT test score (enrolled as of grade 11) (5)	Number with an observed AP test score (6)	Number with an expected AP test score (enrolled as of grade 12) (7)
				A. 7th Grade			
1997	1,185	1	0	5	0	0	0
1998	1,081	7	0	46	0	12	0
1999	1,180	50	0	640	0	291	1,180
2000	1,125	707	1,125	647	1,125	341	1,125
2001	1,193	826	1,193	710	1,193	432	1,193
2002	1,235	834	1,235	683	1,235	427	1,235
2003	1,240	844	1,240	687	1,240	481	1,240
2004	1,172	788	1,172	679	1,172	499	1,172
2005	1,095	664	1,095	3	0	345	0
2006	1,158	10	0	0	0	14	0
2007	1,068	0	0	0	0	0	0
2008	998	0	0	0	0	0	0
All Years	13,730	4,731	7,060	4,100	5,965	2,842	7,145
				B. 9th Grade			
2001	680	22	0	374	680	113	680
2002	755	462	755	413	755	159	755
2003	727	520	727	426	727	177	727
2004	815	635	815	478	815	235	815
2005	832	598	832	454	832	255	832
2006	889	612	889	481	889	290	889
2007	842	528	842	2	0	142	0
All Years	5,540	3,377	4,860	2,628	4,698	1,371	4,698

Notes: This table summarizes the observed College Board test score outcomes for exam school applicants. The sample is restricted to students in column (8) of Table C1.

Table D1. Processing of NYC Exam School Application Data

Application Year	Total number of records	Excluding applicants from private schools	Excluding applicants not in Round 1 of the application process	Excluding students who did not rank an exam school	Excluding students who did not rank Brooklyn Tech, Bronx Science or Stuyvesant
	(1)	(2)	(3)	(4)	(5)
2003-04	28,136	23,637	22,293	22,287	22,205
2004-05	28,279	24,123	22,894	22,859	22,776
2005-06	28,442	23,971	22,810	22,810	22,376
2006-07	26,616	22,377	21,278	21,278	20,824
All Years	111,473	94,108	89,275	89,234	88,181

Application Year	Excluding students not matched to student file	Excluding 9th graders	Excluding students who took SHSAT in previous years	Excluding students without post-assignment numeric outcome test scores at all
	(6)	(7)	(8)	(9)
2003-04	22,108	21,091	21,091	18,361
2004-05	22,776	21,883	21,880	19,106
2005-06	22,376	21,448	21,446	18,842
2006-07	20,824	20,124	20,122	17,431
All Years	88,084	84,546	84,539	73,740

Notes: This table summarizes the steps going from raw application data to the analysis sample.

Table D2. Match from NYC Exam Application to Student Data

Application Year	Number of Students	Fraction with Match		
		Total	Offered	Not Offered
	(1)	(2)	(3)	(4)
2003-04	22,205	0.996	0.997	0.995
2004-05	22,776	1	1	1
2005-06	22,376	1	1	1
2006-07	20,824	1	1	1
All Years	88,181	0.999	0.999	0.999

Notes: This table reports the fraction of applicants with a match between the exam application file and the student demographic file. The sample corresponds to column (5) of Table D1.

Table D3. Match from NYC Exam Applicants to Regents Test Score Outcomes

Record Availability	Application School Year				
	2003-04	2004-05	2005-06	2006-07	All Years
	(1)	(2)	(3)	(4)	(5)
			I. Math		
Number of applicants	21,091	21,880	21,446	20,122	84,539
Number with score observed before treatment	2,685	3,157	3,673	3,975	13,490
Number with score observed after treatment	15,055	15,307	14,206	12,492	57,060
Number with different multiple scores observed after treatment	1,795	2,360	2,000	1,821	7,976
Number with different multiple scores observed after treatment, on first date	3	20	10	2	35
Number with score observed on most common date	5,822	5,873	6,078	8,033	25,806
Number with score observed before most common date	3,522	3,875	4,348	2,022	13,767
Number with score observed after most common date	5,711	5,559	3,779	2,437	17,486
			II. Advanced Math		
Number of applicants	21,091	21,880	21,446	20,122	84,539
Number with score observed before treatment	7	9	13	29	58
Number with score observed after treatment	10,375	10,691	10,939	12,130	44,135
Number with different multiple scores observed after treatment	1,469	1,750	898	235	4,352
Number with different multiple scores observed after treatment, on first date	13	4	0	0	17
Number with score observed on most common date	3,913	3,938	5,496	11,177	24,524
Number with score observed before most common date	4,310	4,671	5,443	953	15,377
Number with score observed after most common date	2,152	2,082	0	0	4,234
			III. English		
Number of applicants	21,091	21,880	21,446	20,122	84,539
Number with score observed before treatment	2	1	0	n.a	14
Number with score observed after treatment	16,847	17,322	17,202	n.a	54,410
Number with different multiple scores observed after treatment	1,979	2,024	1,501	n.a	5,641
Number with different multiple scores observed after treatment, on first date	11	3	0	n.a	14
Number with score observed on most common date	9,333	8,614	8,985	n.a	29,389
Number with score observed before most common date	1,829	2,587	2,705	n.a	7,703
Number with score observed after most common date	5,685	6,120	5,512	n.a	17,317
			IV. Global History		
Number of applicants	21,091	21,880	21,446	20,122	84,539
Number with score observed before treatment	3	19	18	8	48
Number with score observed after treatment	17,057	17,735	16,434	15,429	66,655
Number with different multiple scores observed after treatment	2,321	2,882	1,771	203	7,177
Number with different multiple scores observed after treatment, on first date	19	59	1	0	79
Number with score observed on most common date	13,746	13,471	13,100	14,328	54,645
Number with score observed before most common date	796	844	1,037	1,101	3,778
Number with score observed after most common date	2,514	3,420	2,296	0	8,230
			V. US History		
Number of applicants	21,091	21,880	21,446	20,122	84,539
Number with score observed before treatment	41	23	91	n.a	256
Number with score observed after treatment	15,766	16,015	14,270	n.a	47,906
Number with different multiple scores observed after treatment	1,152	1,102	496	n.a	2,962
Number with different multiple scores observed after treatment, on first date	20	0	3	n.a	23
Number with score observed on most common date	10,252	10,365	11,844	n.a	33,431
Number with score observed before most common date	1,464	2,068	2,426	n.a	6,013
Number with score observed after most common date	4,049	3,582	0	n.a	8,461
			VI. Living Environment		
Number of applicants	21,091	21,880	21,446	20,122	84,539
Number with score observed before treatment	440	878	894	922	3,134
Number with score observed after treatment	16,562	16,807	16,310	14,102	63,781
Number with different multiple scores observed after treatment	1,356	1,807	1,484	977	5,624
Number with different multiple scores observed after treatment, on first date	2	7	8	0	17
Number with score observed on most common date	11,601	11,455	11,286	11,071	45,413
Number with score observed before most common date	207	324	344	209	1,084
Number with score observed after most common date	4,754	5,027	4,679	2,822	17,282

Notes: This table summarizes the match between Regents test score outcomes and exam school applicants. The sample is restricted to students in column (8) of Table D1.

References

Abdulkadiroğlu, A., J. D. Angrist, S. M. Dynarski, T. J. Kane, and P. A. Pathak (2011): "Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots," *Quarterly Journal of Economics*, 126(2), 699–748.

Abdulkadiroğlu, A., P. A. Pathak, and A. E. Roth (2009): "Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the New York City High School Match," *American Economic Review*, 99(5), 1954–1978.

Angrist, J. D. (1998): "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants," *Econometrica*, 66(2), 249–288.

Angrist, J. D., S. M. Dynarski, T. J. Kane, P. A. Pathak, and C. R. Walters (2010): "Who Benefits from KIPP?," NBER Working paper 15740.

Angrist, J. D., and K. Lang (2004): "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program," *American Economic Review*, 94(5), 1613–1634.

Angrist, J. D., P. A. Pathak, and C. R. Walters (2011): "Why Are Some Charter Schools Better than Others? Evidence from Massachusetts," Working paper, MIT.

Bowen, W. G., and D. Bok (2000): *The Shape of the River*. Princeton University Press.

Bui, S., S. Craig, and S. Imberman (2011): "Is Gifted Education a Bright Idea? Assessing the Impacts of Gifted and Talented Program," NBER Working Paper, 17089.

Clarke, D. (2008): "Selective Schools and Academic Achievement," IZA Discussion Paper 3182.

Cullen, J. B., and B. Jacob (2008): "Is gaining access to a selective elementary school gaining ground? Evidence from randomized lotteries," in *An Economics Perspective on the Problems of Disadvantaged Youth*, ed. by J. Gruber. Chicago IL: University of Chicago Press.

Cullen, J. B., B. A. Jacob, and S. Levitt (2006): "The Effect of School Choice on Participants: Evidence from Randomized Lotteries," *Econometrica*, 74(5), 1191–1230.

Cunha, F., and J. Heckman (2007): "The Technology of Skill Formation," *American Economic Review*, 97(2), 31–47.

Dale, S., and A. B. Krueger (2002): "Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables," *Quarterly Journal of Economics*, 117(4), 1491–1527.

——— (2011): "Estimating the Return to College Selectivity over the Career Using Administrative Earnings Data," Princeton University, Industrial Relations Working paper, 563.

Dee, T. S., B. A. Jacob, J. McCrary, and J. Rockoff (2011): "Rules and Discretion in the Evaluation of Students and Schools: The Case of the New York Regents Examinations," Work in progress, Columbia University.

Dobbie, W., and R. Fryer (2011): "Are High-Quality Schools Enough to Increase Achievement Among the Poor? Evidence from the Harlem Children's Zone," *American Economic Journal: Applied Economics*, 3(3), 158–187.

Duflo, E., P. Dupas, and M. Kremer (2010): "Peer Effects and the Impacts of Tracking: Evidence from a Randomized Evaluation in Kenya," forthcoming, *American Economic Review*.

Goldin, C., and L. F. Katz (2008): *The Race between Education and Technology*. Harvard University Press.

Hahn, J., P. Todd, and W. van der Klaauw (2001): "Identification and Estimation of Treatment Effects with a Regression Discontinuity Design," *Econometrica*, 69(1), 201–209.

Hoekstra, M. (2009): "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach," *Review of Economics and Statistics*, 91(4), 717–724.

Hoxby, C. (2003): "School Choice and School Productivity (Or, Could School Choice be a Rising Tide that Lifts All Boats)," in *The Economics of School Choice*, ed. by C. Hoxby. Chicago: University of Chicago Press.

Hoxby, C., and G. Weingarth (2006): "Taking race out of the equation: School reassignment and the structure of peer effects," Working paper, Harvard University, Available at: http://www.hks.harvard.edu/inequality/Seminar/Papers/Hoxby06.pdf.

Imbens, G., and K. Kalyanaraman (2010): "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," CEMMAP Working paper CWP05/10.

Jackson, K. (2010): "Do Students Benefit from Attending Better Schools? Evidence from Rule-Based Student Assignments in Trinidad and Tobago," Economic Journal, forthcoming.

Jan, T. (2006): "Growing a Boston Latin in Brooklyn," *Boston Globe*, Local Desk, March 4.

Katnani, S. (2010): "Pics, Grutter, and Elite Public Secondary Education: Using Race as a Means in Selective Admissions," *Washington University Law Review*, 87, 625–668.

KUGEL, S. (2005): "Battle Over a Principal Spills Outside a Schools Walls," *New York Times, Neighborhood Report, June 12*, p. Section 14.

LAVY, V., O. SILVA, AND F. WEINHARDT (2009): "The Good, The Bad, and the Average: Evidence on the Scale and Nature of Ability Peer Effects," NBER Working paper, No. 15600.

LEE, D., E. MORETTI, AND M. J. BUTLER (2004): "Do Voters Affect or Elect Policies? Evidence from the U.S. House," *Quarterly Journal of Economics*, 119(3), 807–860.

MARSH, H., D. CHESSOR, R. CRAVEN, AND L. ROCHE (1995): "The Effects of Gifted and Talented Programs on Academic Self-Concept: The Big Fish Strikes Again," *American Educational Research Journal*, 32, 285–319.

POP-ELECHES, C., AND M. URQUIOLA (2010): "Going to a Better School: Effects and Behavioral Responses," Working paper, Columbia University, Available at: http://www.nber.org/c̃onfer/2010/EDf10/Pop-Eleches_Urquiola.pdf.

PORTER, J. (2003): "Estimation in the Regression Discontinuity Model," Working paper, University of Wisconsin.

ROCKOFF, J., AND M. HERRMANN (2010): "Does Menstruation Explain Gender Gaps in Work Absenteeism?," forthcoming, *Journal of Human Resources*.

ROTHSTEIN, J. (2006): "Good Principals or Good Peers: Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition Among Jurisdictions," *American Economic Review*, 96(4), 1333–1350.

SACERDOTE, B. (2001): "Peer Effects with Random Assignment: Results from Dartmouth Roommates," *Quarterly Journal of Economics*, 116, 681–704.

SAULNY, S. (2005): "New York Tops Advanced Placement Tests," *New York Times*, Jan 26.

STEINBERG, J. (1998a): "Alumni to Give Brooklyn Tech Huge Donation," *New York Times, Metropolitan Desk*, pp. March 20, Page 1, Column 1.

——— (1998b): "Bronx High School Gets $1 Million Pledge," *New York Times, Metropolitan Desk*, pp. June 9, Page 4, Column 1.

STERN, S. (2003): "Façade of Excellence," *Education Next, June 25*.

ZHANG, H. (2010): "Magnet Schools and Student Achievement: Evidence from a Randomized Natural Experiment in China," Unpublished mimeo, Chinese University of Hong Kong.