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 authors and do not reflect the views of either the Boston Public Schools or the NYC Department of Education. 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 many helpful discussions along the way. Weiwei Hu and Miikka Rokkanen provided superb research assistance. Pathak 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, Revised September 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 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. 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. 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. Finally, we explore the external validity of RD estimates, arguing that as best we can tell, there is little effect of an exam school education on achievement even for the highest-ability marginal applicants and for applicants to the right of admissions cutoffs. 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 RD setup.

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 are typical of the schools in their host district. It's also common for exam schools to have an older, perhaps more experienced teaching staff, though exam school teaching assignments are usually considered a perquisite associated with seniority alone and not necessarily teaching quality (Stern, 2003). 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.

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. For the most part, this finding carries over to subgroups of minorities and women, though we find some evidence of an exam school achievement boost on high school English tests for minority applicants in Boston. Results for minorities are of special interest given the history of litigation around minority admissions. Our analysis of College Board outcomes also uncovers a small subset of admitted applicants who appear to earn higher SAT scores than they otherwise would have.

One explanation for the absence of broad exam school achievement gains is the local nature of RD: marginal applicants may be ill-positioned to benefit from an exam school education. At the same time, estimation for these applicants 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. For example, 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 Boston Public Schools (BPS) average (894 on the Reasoning test).

Although the local nature of RD estimates does not diminish our interest in them, the external validity of these estimates is still of considerable interest. A final contribution of our work is a systematic exploration of external validity. First, we exploit the fact that many of the applicants who are marginal on the exam school admissions test have exceptionally high scores on other tests. 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. We also report estimates from two econometric modeling strategies intended to extrapolate causal effects for applicants who are away from admissions cutoffs; both approaches fail to uncover any evidence that the results for marginal applicants are unusual.

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 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

[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.

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

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

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 one of two RD analyses of achievement effects at highly selective U.S. exam schools. In independent work, Dobbie and Fryer (2011b) examine the impact of New York City's exam schools on longer term academic outcomes; their analysis shows no impact on college enrollment and quality. Selective high schools have also 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. 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

[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.

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, 2011).

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 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.

[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.

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 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

[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).

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 a 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. 7th grade applicants are typically transferring out of their current middle school, while 9th grade applicants are picking a new high school and typically participate in the regular high school matching process, as well as the exam school process. Exam school 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).

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. Exam school assignment is then based on the student-proposing deferred acceptance algorithm (see, e.g., Pathak (2011) for a formal description) which generates a cutoff, C_k, the largest rank to obtain an offer for each school. 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, data in the Boston window were used to 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 a 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.

[10] As with the estimates discussed below, the sample used for to estimate smoothed lines in the plots includes only applicants within 10 units of each school-specific cutoff, though the plot window is wider. IK implementation is discussed in detail in the next section.

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 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. It's also worth nothing that the lower deciles of the baseline peer score distribution jump by a roughly similar amount. This implies that even if marginal admits are tracked within schools (so that they study primarily with the lower tail of the exam school enrolled population), the peer quality to which they are exposed nevertheless jumps at the cutoff.

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 they capture the effect of an exam school offer, without adjustment for the relationship between offers and enrollment or other mediating variables. The Boston empirical work is limited to sets of applicants with school-specific running variables in the interval [-10,+10]. Applicants outside this "Boston window" are well below or well above the relevant cutoffs. At the same time, the [-10,+10] window is wide enough to allow for reasonably precise inference using Boston applicant data.

The parametric estimating equation for applicants to school k is

$$y_{itk} = \alpha_{tk} + \sum_j \delta_{jk} d_{ij} + (1 - D_{ik}) f_{0k}(r_{ik}) + D_{ik} f_{1k}(r_{ik}) + \rho_k D_{ik} + \eta_{itk}, \tag{2}$$

where the variable D_{ik} is an indicator for $r_{ik} \geq 0$ and the coefficient of interest is ρ_k. 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 a pair of third-order polynomials that differ on either side of the cutoff, specifically

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

The reduced form estimates generated by this set-up reflect the fact that applicants to school k typically apply to more than one school and are likely to have (or to lose) other exam school options as distance from a given cutoff grows. For example, highly qualified Latin Academy applicants also qualify for admission to Latin School. Looking in the other direction, while many 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. In principle, the IK procedure adjusts for the additional nonlinearity by shortening the bandwidth. The possibility that the resulting bandwidth straddles multiple cutoffs, however, raises the question of how we should interpret the resulting estimates.

Figure 7 presents a stylized representation of the Boston admissions process that sheds some light on the nature of the causal effects captured by equation (2). The bottom of the figure sketches own-school and other-school offers for Latin Academy applicants. At the cutoff, Latin Academy offers naturally jump, but as we move to the right, Latin School offers are made, while moving to the left, O'Bryant offers fall away. The changing pattern of offers is reflected in the reduced form relation between achievement and the Latin Academy running variable. Near the Latin Academy cutoff, reduced form estimates capture the effects of a Latin Academy offer (only) on students who would also qualify for O'Bryant. In practice, however, even the non-parametric estimation strategy borrows information that that is close to and sometimes beyond the next cutoff up or down. Therefore, the reduced form effect of Latin Academy offers captures an average causal effect generated by the difference in outcomes between applicants who qualify for admission to Latin Academy, some of whom also qualify at Latin School, and applicants who don't qualify for Latin Academy, most, but not all of whom qualified for O'Bryant. Although not sketched in the figure, it should be clear that the same sort of mixing occurs to the left of the Latin School cutoff and to the right of the O'Bryant cutoff.

Following the discussion of reduced form estimates, we use 2SLS to simplify the interpretation of reduced form estimates by assigning reduced form effects to specific channels, such as peer means, that are manipulated at all cutoffs in some measure.

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 10.[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 equation as

$$\begin{aligned} y_{itk} &= \alpha_{tk} + \sum_j \delta_{jk} d_{ij} + \gamma_{0k}(1 - D_{ik}) r_{ik} + \gamma_{1k} D_{ik} r_{ik} + \rho_k D_{ik} + \eta_{itk}, \\ &= \alpha_{tk} + \sum_j \delta_{jk} d_{ij} + \gamma_{0k} r_{ik} + \gamma_k^* D_{ik} r_{ik} + \rho_k D_{ik} + \eta_{itk} \end{aligned} \tag{4}$$

for each of the three schools indexed by k. Non-parametric RD estimates come from a kernel-weighted least square fit of equation (4).

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 as before. Jumps in smoothed scores at admissions cutoffs are the IK sharp regression discontinuity estimates of

[13] The IK bandwidths for Table 4 range from about 8 to 24.

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 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 (4), 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-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 0.028 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 co-

[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.

variates. 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 0.51; for 9th graders in the Boston window (mostly applying to O'Bryant), the re-enrollment rate falls to 0.44. Exam school offers *increase* enrollment persistence by 0.11 for 7th grade applicants and by 0.17 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.

5.2 Estimates for Minorities and by Sex

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 blacks and Hispanics, with an estimated effect of 0.17σ. In fact, the full-sample ELA results appear to be driven primarily by

[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.

the minority impact, since the corresponding IK estimate for non-minorities is 0.07σ (se=0.042).

Our investigation of differences in student achievement by gender is motivated by long-standing academic interest in gender gaps in achievement. For instance, Ellison and Swanson (2010) document a substantial male advantage in elite high school math contests. The estimates in Table 5, however, show that exam schools have little effect on math scores for boys as well as girls, while effects on English are similar.

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 (4), 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

$$y_{itk} = \Gamma_k' X_{itk} + \theta M_{it} + \epsilon_{itk}, \tag{5}$$

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 (4). 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 1.5 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 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 (-0.024 with standard error of 0.033), 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 $0.53 - 0.63\sigma$ 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

[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.

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

$$y_{itk} = \Gamma_k' X_{itk} + \theta_0 E_{it} + \theta_1 E_{it} g_{ik} + \epsilon_{itk}, \tag{6}$$

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 (6) 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 -0.078 (se $= 0.078$) for math and -0.119 (se $= 0.080$) 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 External Validity

RD non-parametrically identifies causal effects for those near treatment cutoffs. How limiting is this? In the Boston exam school context, it seems worth emphasizing that the three cutoffs in our sample cover a wide range of ability. Among 7th grade applicants, the O'Bryant cutoff falls near the median of the ISEE distribution while the Latin School cutoff reaches the 75th quantile. 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 reaches the 92nd quantile. It's impressive that the effects of clearing these widely spaced thresholds are similar. This robustness notwithstanding, we'd also like to say something about how an exam school education affects achievement for unusually high achievers, even within the exam school population. We briefly explore two approaches to this question, the first based on covariates and the second an extrapolation using functional form.

High Achievers

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 (non-parametric) RD estimates for, say, O'Bryant students with ISEE scores in the upper 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: no more than one third top out in the sense of testing

[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.

at the Advanced (highest) MCAS proficiency level. Likewise, MCAS remains informative even for applicants in the upper baseline 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 high-baseline 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 columns 1-4 of Table 7, are mostly negative with few significantly different from zero. The exception again is a significant positive effect for 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 (baseline) achievers, there is little evidence of a consistent exam school boost.

Extrapolation Away from Cutoffs

The previous analysis looks at students with high baseline values, but says nothing about effects for applicants with ISEE scores above admissions cutoffs. Perhaps the ISEE has score special significance. For example, once admitted to an exam school, students may be tracked (perhaps implicitly) on the basis of their ISEE scores. Moreover, in ongoing theoretical work, Ellison (2011) argues that in school systems that sort by ability across campuses, the optimal design has no discontinuity in score levels at admissions cutoffs, though we should see the score gradient get steeper for those admitted to a more challenging program. We'd therefore like to say something about treatment effects to the right of exam school admissions cutoffs.

Parametric RD identifies causal effects along the entire support of the running variable, an important feature of the parametric approach noted by Angrist and Pischke (2009). To see this, begin by letting $E[\mathrm{Y}_{0i}|r_i]$ and $E[\mathrm{Y}_{1i}|r_i]$ denote the conditional expectation functions (CEFs) for potential outcomes indexed against treatment assignment $\{\mathrm{Y}_{ji}; j = 0, 1\}$ and conditional on the running variable (here we suppress the application school subscript, k.) Modeling both of these CEFs with p^{th}-order polynomials, we have

$$\begin{aligned} E[\mathrm{Y}_{0i}|r_i] &= \alpha + \beta_{01} r_i + \beta_{02} r_i^2 + ... + \beta_{0p} r_i^p, \\ E[\mathrm{Y}_{1i}|r_i] &= \alpha + \rho + \beta_{11} r_i + \beta_{12} r_i^2 + ... + \beta_{1p} r_i^p. \end{aligned}$$

Since D_i is a deterministic function of r_i, we can write

$$E[\mathrm{Y}_i|r_i] = E[\mathrm{Y}_{0i}|r_i] + E[\mathrm{Y}_{1i} - \mathrm{Y}_{0i}|r_i] D_i.$$

Finally, substituting polynomials for conditional expectations, we have

$$\begin{aligned} \text{Y}_i &= \alpha + \beta_{01} r_i + \beta_{02} r_i^2 + ... + \beta_{0p} r_i^p \\ &\quad + \rho \text{D}_i + \beta_1^* \text{D}_i r_i + \beta_2^* \text{D}_i {r_i}^2 + ... + \beta_p^* \text{D}_i {r_i}^p + \eta_i, \end{aligned} \tag{8}$$

where $\beta_1^* = \beta_{11} - \beta_{01}$, $\beta_2^* = \beta_{12} - \beta_{02}$, and $\beta_p^* = \beta_{1p} - \beta_{0p}$ and the error term, η_i, is the CEF residual, $\text{Y}_i - E[\text{Y}_i|r_i]$. The treatment effect at $r_i = c > 0$ is

$$\tau(c) \equiv E[\text{Y}_{1i} - \text{Y}_{0i}|r_i = c] = \rho + \beta_1^* c + \beta_2^* c^2 + ... + \beta_p^* c^p,$$

while the treatment effect at the cutoff is just ρ.

Figure 14 illustrates the extrapolation implicit in parametric estimates of causal effects away from the cutoff. We observe $E[\text{Y}_{0i}|r_i]$ to the left of the cutoff and $E[\text{Y}_{1i}|r_i]$ to the right. But identification of treatment effects at $r_i = c > 0$ requires knowledge of $E[\text{Y}_{0i}|r_i = c]$. A parametric model for $E[\text{Y}_{0i}|r_i]$ constructs the missing counterfactual mean of Y_{0i} to the right of the cutoff by extrapolating $E[\text{Y}_{0i}|r_i]$ from the left, generating the parametric estimand labeled $\tau_A(c)$ in the figure.

The problem with parametric extrapolation is that the parametric model may be incorrect. Suppose, for example, that the mean of Y_{0i} to the right of the cutoff is described by the upper dotted line to the right of the cutoff in Figure 14. Then the causal effect at c is $\tau_B(c)$. An interesting question in this context is whether effects away from the cutoff are non-parametrically identified. In other words, given an infinitely large sample, could we tell whether $\tau_A(c)$ or $\tau_B(c)$ is the right answer? Without constraining the behavior of the underlying CEF in some way, the answer is no, even if we impose smoothness on the underlying functions. As suggested by the figure, whatever shape a continuously differentiable $E[\text{Y}_{0i}|r_i]$ takes to the left of the cutoff, there are many continuously differentiable pieces (continuous at the cutoff) consistent with this to the right. This suggests that continuous differentiability alone is not strong enough to allow us to extrapolate $E[\text{Y}_{0i}|r_i]$ from left to right.

Stronger regularity assumptions restricting the behavior of $E[\text{Y}_{0i}|r_i]$ buy more in the way of identification. Suppose we're prepared to assume that $E[\text{Y}_{0i}|r_i]$ is an analytic function over the interval we're concerned with, that is, from the cutoff through the intended extrapolation point. This means that on this interval, the function can be written as a power series or, equivalently, that a Taylor approximation converges to it. We can then imagine improving the approximation to the left of the cutoff to the point where we eventually learn what $E[\text{Y}_{0i}|r_i = c]$ is up to any desired level of accuracy. It seems reasonable to presume, however, that in practice the quality of a profligate approximation is likely to be limited. As an empirical matter, it makes sense to settle for some low-order polynomial such as the cubic used here. (And, in any case, the question of whether the underlying CEF is analytic seems hard to assess.)

Recently, Dong and Lewbel (2011) propose a derivative-based approach to RD extrapolation in the neighborhood of the cutoff, beginning with the observation that given continuous differentiability of the underlying CEFs, the derivative of the treatment effect at the cutoff, $\tau'(0)$, is non-parametrically identified.[19] Dong and Lewbel use this fact to motivate an approximate treatment effect

$$\tau(c) \approx \tau(0) + \tau'(0) \cdot c,$$

interpreting this as the effect were the cutoff to be moved by a small amount. Let $g_1(r_i)$ and $g_0(r_i)$ denote the continuously differentiable CEFs for potential outcomes. With cutoff at zero, the treatment effect moving the cutoff to the right is

$$E[\mathrm{Y}_{1i} - \mathrm{Y}_{0i} | r_i = c] = g_1(c) - g_0(c).$$

Dong and Lewbel's "move-the-cutoff" extrapolation uses derivatives to approximate both $g_1(c)$ and $g_0(c)$. Here, we use a similar local derivative argument to approximate the treatment effect for someone with $r_i = c$ for small c when the cutoff is fixed at zero. Move-the-cutoff and away-from-the-cutoff effects are the same when the location of the cutoff has no impact on potential outcomes.

Our extrapolation exploits the fact that we only need to impute $g_0(c)$; $g_1(c)$ is non-parametrically identified. We therefore impute effects away from the cutoff using the fact that $g_0(c) \approx g_0(0) + g_0'(0) \cdot c$. The idea behind this is rendered in Figure 14 as the vertical distance falling from $E[\mathrm{Y}_{1i}|r_i = c]$ to the line tangent to $E[\mathrm{Y}_{0i}|r_i = 0]$. This approximation distinguishes between possible counterfactuals according to where $g_0(r_i)$ is headed in a small neighborhood of the cutoff. Of course, two functions might have the same first derivative and yet be headed in different directions. In that case, we could look to higher-order derivatives to tell them apart, though the payoff to additional approximation terms falls off quickly, since the rate at which non-parametric estimates of derivatives converge falls with the order of differentiation (Stone, 1982). Derivate-based approximation is not *guaranteed* to answer the substantive question of interest in our context, which is whether treatment effects increase as the running variable increases. On the other hand, we might indeed be interested in what happens just to the right of the cutoff, which in this case means not being among the very last applicants admitted to an exam school.

We constructed non-parametric estimates of $g_1(c)$ by local linear regression with bandwidth centered at c. The imputed $g_0(c)$ was constructed from local linear estimates of the slope and

[19] The identification argument here is similar to that invoked for non-parametric identification of the treatment effect itself; see, e.g., Li and Racine (2007).

intercept in a bandwidth-determined neighborhood just to the left of the cutoff, as implemented earlier for RD estimates at the cutoff. The imputed treatment effect at c can be obtained from these pieces in one go by kernel-weighted estimation of equation (4). In this case, however, although we want the estimates to the left to be constructed at the cutoff as before, estimates to the right use a weighting scheme centered at the extrapolation point, c. This is accomplished by using a kernel that recenters to the right:

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

where $\tilde{u}_{ik} = \frac{r_{ik} - D_i c}{h}$ and h is the bandwidth.

Parametric extrapolation results for $c = 1$ and $c = 5$ for a sample combining applicants for all three schools are reported in columns (5) and (6) of Table 7.[20] These estimates show no evidence of a positive treatment effect for higher scoring students, though the extrapolation 5 units out is very imprecise. Dong-Lewbel extrapolation produces precisely estimated zeros for extrapolated math effects one unit to the right of the cutoff, as can be seen in column (7). Non-parametric extrapolation of ELA effects generates results similar to those at the cutoff, showing a significant negative effects on middle school math and a signficant positive effect on 10th grade reading. The similarity of the non-parametrically extrapolated estimates to those at the cutoff is a consequence of the small interaction terms generated by IK estimation of the reduced form. The estimated interactions appear in column (8), which reports IK estimates of the slope change at the cutoff.

5.5 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.[21] 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).

[20] As in the earlier reduced form All-Schools estimates, the specification allows school-specific running variable effects, but the cutoff main effect and the interactions between the cutoff and the running variable terms are restricted to be the same across schools.

[21] 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.[22]

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 15 and 16 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 8 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 (0.31 with se=0.13 at Latin Academy and 0.17 with se=0.08 at Latin School), but the overall SAT effect for minority students is 0.04, 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).

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

AP tests are scored on a scale of 1-5 and some colleges grant credit for some AP subjects 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 17 shows little evidence of a jump in the number of tests taken at any exam school cutoff. Figure 18 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 17 and 18 appear in Table 9. 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.[23] 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

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

scores. Seats are then allocated down this ranking, with the top scorer getting his first choice, the second highest scorer get his most preferred choice among schools with remaining seats, and so on.[24]

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 10 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.

[24] The NYC exam school assignment mechanism is a *serial dictatorship* with students ordered by the admissions test score. Though exam school assignment is a distinct process, students apply for exam schools at the same time they rank regular New York high schools, and may receive offers from both. Abdulkadiroğlu, Pathak, and Roth (2009) describe how exam school admissions interacts with admissions at regular high schools.

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 +/-10 because there are many more New York applicants, so the sample is much larger than Boston's even in a window half the size, and because trimming at five greatly mitigates problems of interpretation due to 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 19 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 19 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 20. 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 21.[25]

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 22-23. 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 24 and 25, 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 con-

[25]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.

firmed in Table 11, 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 (4) for applicants in the $[-5, +5]$ interval.[26]

The New York estimates are precise enough to rule out even modest score gains. For 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 11 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 12. 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 minorities and by sex in New York appear in Appendix Table B3. Unlike Boston, New York's exam schools do not appear to boost Regents achievement for blacks and Hispanics. Moreover, neither boys nor girls appear to score higher because of exam school attendance. The models with baseline interactions and extrapolations away from cutoffs are in Appendix Table B4. Few of the estimates for high achievers are significantly different from zero and most are negative, in line with the Boston findings. The results from the extrapolation are also mostly zeroes, though they are less precise further away from the cutoff.

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 13, 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 0.02, while the corresponding peer effects have estimated standard errors on the order of 0.03.

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

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 (2011a) 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 though Dobbie and Fryer's (2011b) recent analysis of college matriculation and enrollment for New York city exam school students weighs against this possibility. In any case, if there are other later life gains, 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), Hsieh and Urquiola (2006), Rothstein (2006), MacLeod and Urquiola (2009), 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 (1)	Exam Applicants (2)	Offered Students (3)	Enrolled Students (4)	All Boston (5)	Exam Applicants (6)	Offered Students (7)	Enrolled Students (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. LEP means Limited English Proficient. SPED means Special Education. N is the number of observations with at least one non-missing value for the variable listed.

‡ LEP only available beginning in year 1998.

¥ SPED 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 tables 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, are shown in parentheses.
* significant at 10%; ** significant at 5%; *** significant at 1%

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

		Parametric Estimates				Non-parametric (IK) Estimates			
			Latin	Latin	All		Latin	Latin	All
Application	Test	O'Bryant	Academy	School	Schools	O'Bryant	Academy	School	Schools
Grade	Grade	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
					A. Math				
7th	7th and 8th	-0.121	-0.029	-0.038	-0.062	-0.091	-0.037	-0.026	-0.049
		(0.115)	(0.114)	(0.112)	(0.061)	(0.069)	(0.065)	(0.072)	(0.036)
		2779	2856	2610	8245	2590	2856	2537	7983
7th and 9th	10th	0.015	-0.031	-0.059	-0.024	0.051	-0.035	-0.063*	-0.011
		(0.099)	(0.078)	(0.065)	(0.043)	(0.055)	(0.044)	(0.035)	(0.027)
		1876	1822	1606	5304	1876	1803	1474	5153
7th and 9th	7th, 8th, and 10th	-0.065	-0.030	-0.046	-0.047	-0.030	-0.036	-0.038	-0.035
		(0.086)	(0.085)	(0.076)	(0.046)	(0.051)	(0.045)	(0.052)	(0.027)
		4655	4678	4216	13549	4466	4659	4011	13136
					B. English				
7th	7th and 8th	-0.186*	-0.079	-0.125	-0.129***	-0.091	-0.023	-0.115**	-0.079**
		(0.107)	(0.079)	(0.085)	(0.050)	(0.061)	(0.052)	(0.046)	(0.031)
		2505	2549	2257	7311	2087	1992	2173	6252
7th and 9th	10th	0.037	0.216*	0.058	0.103*	0.133**	0.206***	0.045	0.130***
		(0.104)	(0.121)	(0.086)	(0.061)	(0.060)	(0.058)	(0.064)	(0.036)
		1879	1825	1609	5313	1879	1633	1543	5055
7th and 9th	7th, 8th, and 10th	-0.088	0.046	-0.049	-0.030	0.025	0.091*	-0.045	0.022
		(0.074)	(0.079)	(0.072)	(0.044)	(0.043)	(0.047)	(0.046)	(0.028)
		4384	4374	3866	12624	3966	3625	3716	11307

Notes: This table reports estimates of the effects of exam school offers on MCAS scores. The sample covers students within 10 standardized units of offer cutoffs. Parametric models include a cubic function of the running variable, allowed to differ on either side of offer cutoffs. Non-parametric estimates use the edge kernel, with bandwidth computed following Imbens and Kalyanaraman (2010). 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

		By Race				By Sex	
				Black or	Not Black or		
Application	Test	Black	Hispanic	Hispanic	Hispanic	Men	Women
Grade	Grade	(1)	(2)	(3)	(4)	(5)	(6)
				A. Math			
7th	7th and 8th	-0.014	-0.137*	-0.078	-0.032	-0.016	-0.060
		(0.065)	(0.077)	(0.052)	(0.047)	(0.059)	(0.049)
		2363	1569	3759	3410	3386	4547
7th and 9th	10th	-0.031	-0.040	-0.031	-0.018	0.052	-0.051
		(0.051)	(0.070)	(0.049)	(0.034)	(0.035)	(0.036)
		1588	1001	2495	2281	2165	2990
7th and 9th	7th, 8th, and 10th	-0.021	-0.098*	-0.059	-0.026	0.014	-0.057
		(0.047)	(0.057)	(0.039)	(0.033)	(0.042)	(0.036)
		3951	2570	6254	5691	5551	7537
				B. English			
7th	7th and 8th	-0.026	-0.143**	-0.059	-0.083**	-0.058	-0.079*
		(0.047)	(0.070)	(0.043)	(0.042)	(0.052)	(0.041)
		2187	1431	3548	3144	2780	3829
7th and 9th	10th	0.192***	0.117	0.173***	0.070*	0.144***	0.109***
		(0.053)	(0.080)	(0.046)	(0.042)	(0.051)	(0.042)
		1560	962	2673	2355	2027	3028
7th and 9th	7th, 8th, and 10th	0.065	-0.048	0.045	-0.017	0.028	0.005
		(0.042)	(0.062)	(0.038)	(0.036)	(0.042)	(0.037)
		3747	2393	6221	5499	4807	6857

Notes: This table reports reduced form estimates for minorities and by sex. The table shows non-parametric 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	-0.050			0.053			-0.010			0.032		
	(0.069)			(0.058)			(0.066)			(0.057)		
Exam Years		-0.024			0.026			0.003			0.018	
		(0.033)			(0.031)			(0.032)			(0.030)	
Peer Mean			-0.089*			-0.001			-0.040			0.001
			(0.049)			(0.052)			(0.043)			(0.046)
	First Stage Estimates											
O'Bryant	0.730***	1.530***	0.753***	0.737***	1.403***	0.687***						
	(0.064)	(0.125)	(0.073)	(0.068)	(0.127)	(0.070)						
Latin Academy	0.118**	0.219*	0.388***	0.142**	0.227	0.378***						
	(0.057)	(0.128)	(0.085)	(0.062)	(0.143)	(0.079)						
Latin School	0.032	0.072	0.628***	0.022	0.045	0.534***						
	(0.022)	(0.053)	(0.093)	(0.022)	(0.058)	(0.080)						
N	13130	13136	11116	11305	11307	10604	13130	13136	11116	11305	11307	10604

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 non-parametric estimates using the reduced form bandwidths computed for Table 4. Excluded 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 Estimates for High Achievers and Away from Admission Cutoffs

		Conditional on Baseline				Extrapolation			
		Baseline in Upper Half		Baseline in Upper Quartile		Parametric		Non-parametric (IK)	
Application Grade	Test Grade	Baseline Mean (1)	Estimates (2)	Baseline Mean (3)	Estimates (4)	1 unit away from cutoff (5)	5 units away from cutoff (6)	1 unit away from cutoff (7)	Derivative (8)
					A. Math				
7th	7th and 8th	1.445	-0.104**	2.045	-0.040	-0.143	-0.740	-0.049	0.001
			(0.050)		(0.087)	(0.107)	(0.646)	(0.039)	(0.007)
		3768	3831	1723	1742	8245	8245	7983	7983
7th and 9th	10th	1.334	-0.007	1.779	-0.020	-0.058	-0.352	-0.016	-0.006
			(0.028)		(0.031)	(0.072)	(0.430)	(0.028)	(0.006)
		3469	3163	1935	1948	5304	5304	5153	5153
7th and 9th	7th, 8th, and 10th	1.392	-0.060*	1.904	-0.029	-0.109	-0.586	-0.036	-0.001
			(0.032)		(0.047)	(0.079)	(0.465)	(0.029)	(0.006)
		7237	6994	3658	3690	13549	13549	13136	13136
					B. English				
7th	7th and 8th	1.336	-0.095***	1.758	-0.114**	-0.166**	-0.677	-0.075**	0.004
			(0.036)		(0.058)	(0.081)	(0.487)	(0.033)	(0.007)
		4159	3767	1922	1752	7311	7311	6252	6252
7th and 9th	10th	1.200	0.074*	1.501	0.087**	0.023	-0.619	0.130***	0.000
			(0.039)		(0.043)	(0.089)	(0.475)	(0.038)	(0.005)
		3206	3065	1770	1568	5313	5313	5055	5055
7th and 9th	7th, 8th, and 10th	1.277	-0.013	1.635	-0.022	-0.085	-0.657*	0.022	0.001
			(0.033)		(0.046)	(0.068)	(0.380)	(0.030)	(0.004)
		7365	6832	3692	3320	12624	12624	11307	11307

Notes: This table reports reduced form estimates for students with high baseline scores and for applicants away from admission cutoffs. Baseline means and the proportion of applicants at an advanced level are computed for those who belong to at least one discontinuity sample. Conditional on baseline are non-parametric estimates with bandwidth computed as in the all schools model in Table 4. Parametric extrapolation estimates use the parametric model to form counterfactuals 1 and 5 units from the cutoff. Non-parametric estimates are based on approximations of the derivative of the treatment effect at the cutoff.

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

Table 8. Boston Reduced Form Estimates - PSAT and SAT Scores

	Probability Tested				Test Score for Takers			
		Latin	Latin	All		Latin	Latin	All
Application	O'Bryant	Academy	School	Schools	O'Bryant	Academy	School	Schools
Grade	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				A. PSAT				
7th	0.101**	0.103**	-0.072	0.050*	0.048	-0.075	0.021	0.001
	(0.051)	(0.048)	(0.049)	(0.030)	(0.074)	(0.060)	(0.067)	(0.036)
	1366	1348	1164	3878	917	965	890	2772
9th	0.116**	-0.008	0.113	0.073**	-0.100	0.205***	0.036	0.036
	(0.056)	(0.053)	(0.071)	(0.034)	(0.104)	(0.068)	(0.120)	(0.055)
	889	701	442	2032	478	376	328	1182
7th and 9th	0.107***	0.069*	-0.034	0.058**	0.004	0.007	0.025	0.011
	(0.038)	(0.037)	(0.041)	(0.023)	(0.059)	(0.050)	(0.063)	(0.032)
	2255	2049	1606	5910	1395	1341	1218	3954
				B. SAT				
7th	0.060	0.111**	0.026	0.067**	0.091	-0.107*	0.111	0.020
	(0.054)	(0.052)	(0.052)	(0.031)	(0.085)	(0.057)	(0.083)	(0.042)
	1349	1354	1207	3910	623	855	860	2338
9th	0.040	-0.065	0.019	0.003	0.017	0.360***	0.189**	0.139**
	(0.064)	(0.069)	(0.078)	(0.039)	(0.083)	(0.111)	(0.081)	(0.056)
	859	716	533	2108	550	318	299	1167
7th and 9th	0.052	0.057	0.024	0.046*	0.052	0.001	0.129*	0.059*
	(0.041)	(0.042)	(0.043)	(0.025)	(0.063)	(0.043)	(0.069)	(0.035)
	2208	2070	1740	6018	1173	1173	1159	3505

Notes: This table reports estimates of the effects of exam school offers on PSAT and SAT test taking and scores. Outcome-specific non-parametric estimates, bandwidths, and standard errors were computed as for Table 4.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 9. 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	0.180	-0.099	0.205	0.110	1.049***	-0.205	0.573	0.562**
	(0.149)	(0.233)	(0.176)	(0.091)	(0.359)	(0.671)	(0.491)	(0.266)
	864	787	898	2549	864	724	868	2456
9th	-0.127	0.100	-0.103	-0.057	-0.004	0.849	-0.181	0.188
	(0.262)	(0.318)	(0.367)	(0.142)	(0.605)	(1.086)	(1.232)	(0.392)
	626	391	370	1387	509	342	342	1193
7th and 9th	0.065	-0.033	0.120	0.055	0.710*	0.126	0.378	0.450**
	(0.162)	(0.207)	(0.178)	(0.084)	(0.365)	(0.560)	(0.487)	(0.227)
	1490	1178	1268	3936	1373	1066	1210	3649
				B. Exams with 500+ Takers				
7th	0.108	-0.105	0.020	0.022	0.690**	-0.458	0.073	0.209
	(0.131)	(0.212)	(0.157)	(0.074)	(0.287)	(0.585)	(0.452)	(0.196)
	864	699	867	2430	864	652	859	2375
9th	-0.285	0.141	-0.353	-0.171	-0.435	0.901	-0.602	-0.105
	(0.248)	(0.288)	(0.320)	(0.138)	(0.518)	(1.009)	(0.961)	(0.363)
	580	444	401	1425	482	363	402	1247
7th and 9th	-0.035	-0.009	-0.093	-0.046	0.351	0.026	-0.134	0.110
	(0.146)	(0.183)	(0.161)	(0.076)	(0.291)	(0.527)	(0.434)	(0.176)
	1444	1143	1268	3855	1346	1015	1261	3622

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/BC, Statistics, Biology, Chemistry, Physics B/C, English Language and Composition, English Literature and Composition, European History, US Government and Politics, US History, Microeconomics, and Macroeconomics. Outcome-specific non-parametric estimates, bandwidths, and standard errors were computed as for Table 4.

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

Table 10. Descriptive Statistics for NYC Exam School Applicants

	All NYC	Any Exam: Exam Applicants	Any Exam: Offered Students	Any Exam: Enrolled Students	Enrolled in: Brooklyn Tech	Enrolled in: Bronx Science	Enrolled in: Stuyvesant
	(1)	(2)	(3)	(4)	(5)	(6)	(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. LEP means Limited English Proficient. SPED means Special Education.
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 11. NYC Reduced Form Estimates - Regents Exams

	Parametric Estimates				Non-parametric (IK) Estimates			
	Brooklyn Tech	Bronx Science	Stuyvesant	All Schools	Brooklyn Tech	Bronx Science	Stuyvesant	All Schools
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Math	0.083	-0.096*	-0.056	-0.021	0.012	-0.129***	-0.037	-0.056***
	(0.064)	(0.056)	(0.039)	(0.029)	(0.040)	(0.033)	(0.038)	(0.018)
	4264	3746	3800	11810	3743	3746	3417	10906
Advanced Math	-0.029	-0.024	-0.030	-0.028	0.000	-0.059	-0.022	-0.025
	(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)	(0.043)	(0.030)	(0.025)	(0.039)	(0.023)	(0.022)	(0.014)
	4950	4581	5150	14681	4950	4217	5150	14317
Global History	-0.085	-0.025	-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	-0.070*	-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	-0.06	0.092**	-0.072**	-0.015	-0.080***	0.057**	-0.031	-0.025**
	(0.041)	(0.039)	(0.033)	(0.021)	(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 sample includes applicants 5 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 12. New York 2SLS Estimates for Enrollment, Years in Exam School, and Peer Means

	Advanced Math			English		
	(1)	(2)	(3)	(4)	(5)	(6)
				2SLS		
Enrollment	-0.073			0.071		
	(0.082)			(0.070)		
Exam Years		-0.027			0.028	
		(0.038)			(0.029)	
Peer Mean			-0.038			0.066
			(0.061)			(0.052)
				First Stage		
Brooklyn Tech	0.401***	0.967***	0.559***	0.406***	0.901***	0.523***
	(0.075)	(0.156)	(0.064)	(0.090)	(0.221)	(0.057)
Bronx Science	0.325***	0.629***	0.177***	0.304***	0.785***	0.155**
	(0.114)	(0.217)	(0.068)	(0.114)	(0.296)	(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 non-parametric estimates using the same bandwidths as for the reduced form estimates in Table 11. 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 13. 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	-0.059			0.059			-0.022			0.038		
	(0.053)			(0.044)			(0.051)			(0.042)		
Exam Years		-0.025			0.027			-0.008			0.016	
		(0.025)			(0.021)			(0.025)			(0.021)	
Peer Mean			-0.070*			0.025			-0.030			0.027
			(0.038)			(0.037)			(0.036)			(0.034)
	First Stage Estimates											
O'Bryant	0.730***	1.530***	0.753***	0.737***	1.403***	0.687***						
	(0.064)	(0.125)	(0.073)	(0.068)	(0.127)	(0.070)						
Latin Academy	0.118**	0.219*	0.388***	0.142**	0.227	0.378***						
	(0.057)	(0.128)	(0.085)	(0.062)	(0.143)	(0.079)						
Latin School	0.032	0.072	0.628***	0.022	0.045	0.534***						
	(0.022)	(0.053)	(0.093)	(0.022)	(0.058)	(0.080)						
Brooklyn Tech	0.401***	0.967***	0.559***	0.406***	0.901***	0.523***						
	(0.075)	(0.156)	(0.064)	(0.090)	(0.221)	(0.057)						
Bronx Science	0.325***	0.629***	0.177***	0.304***	0.785***	0.155**						
	(0.114)	(0.217)	(0.068)	(0.114)	(0.296)	(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	30857	30863	28843	25622	25624	24921	30857	30863	28843	25622	25624	24921

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 non-parametric estimates using bandwidths computed one city at a time. Excluded 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%

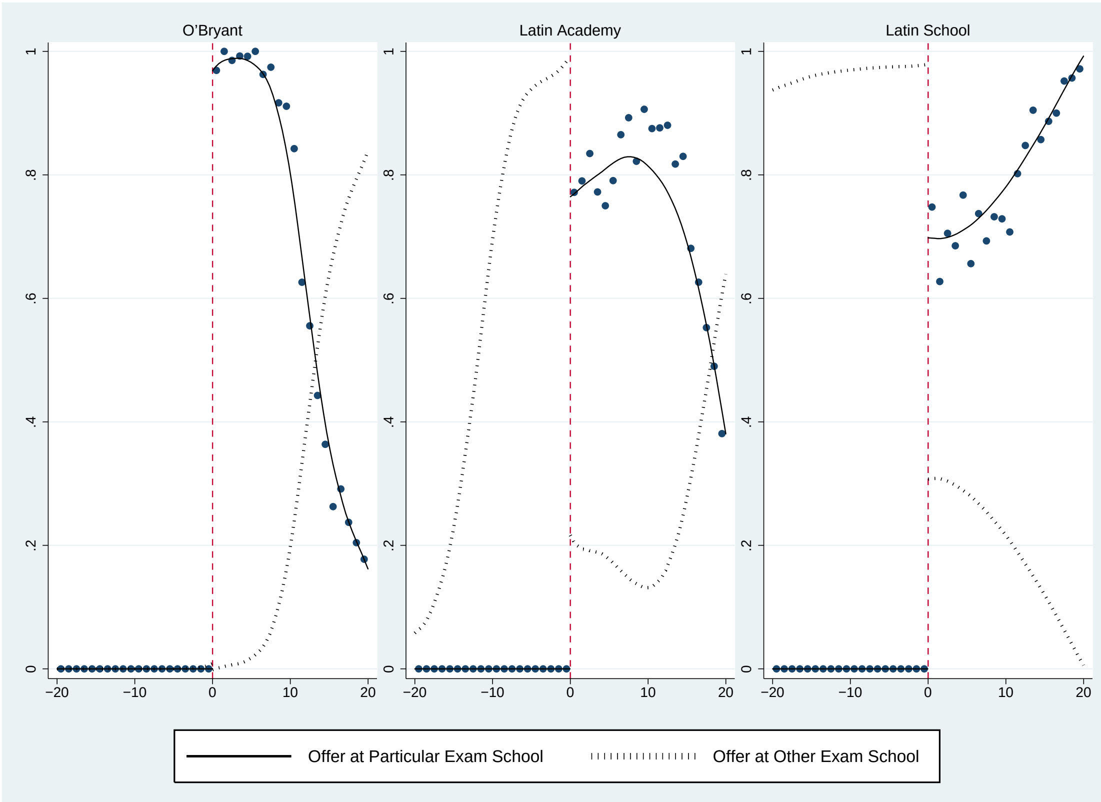

Figure 1. Offers at Each Boston Exam School for 7th Grade Applicants (1997-2008)

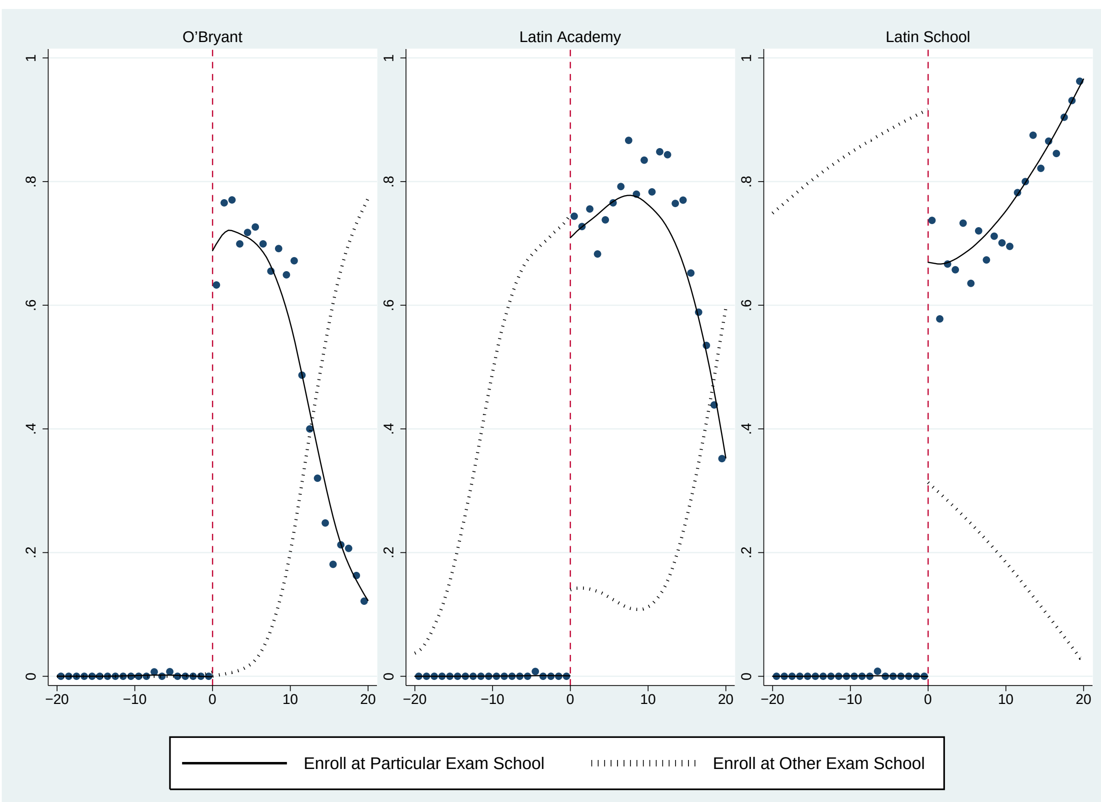

Figure 2. Enrollment at Each Boston Exam School for 7th Grade Applicants (1997-2008)

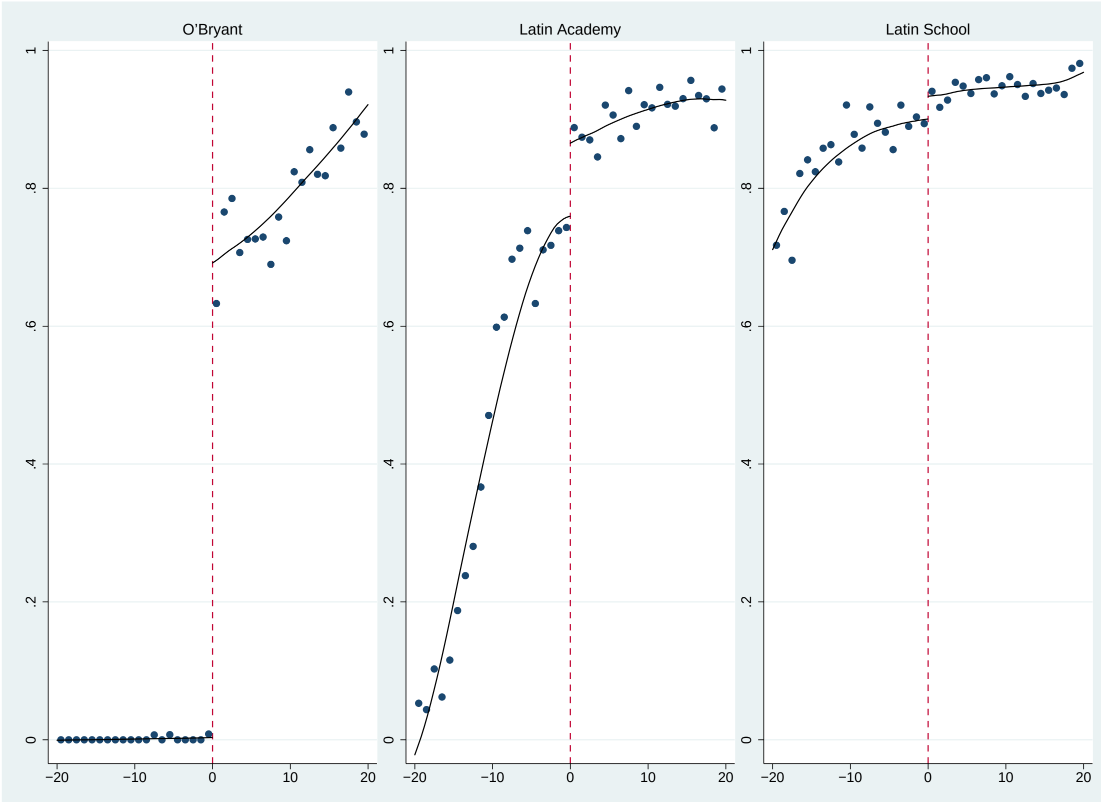

Figure 3. Enrollment at Any Boston Exam School for 7th Grade Applicants (1997-2008)

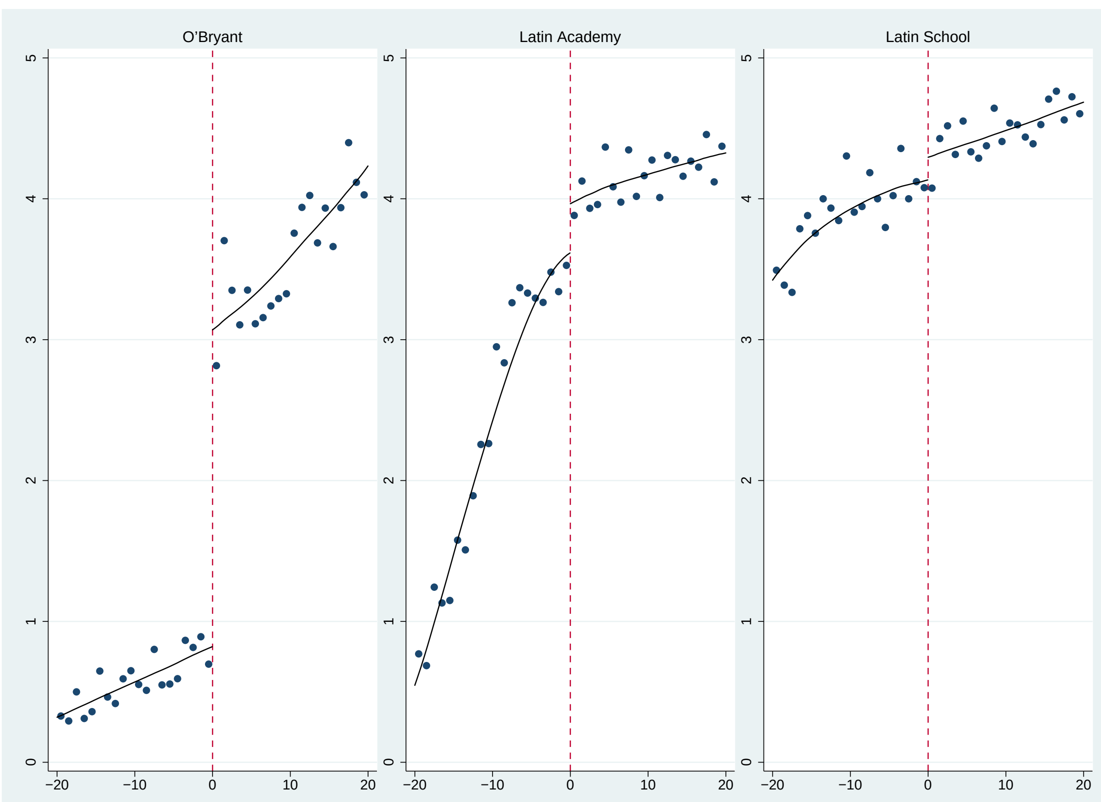

Figure 4. Years at Any Boston Exam School for 7th Grade Applicants (1997-2008)

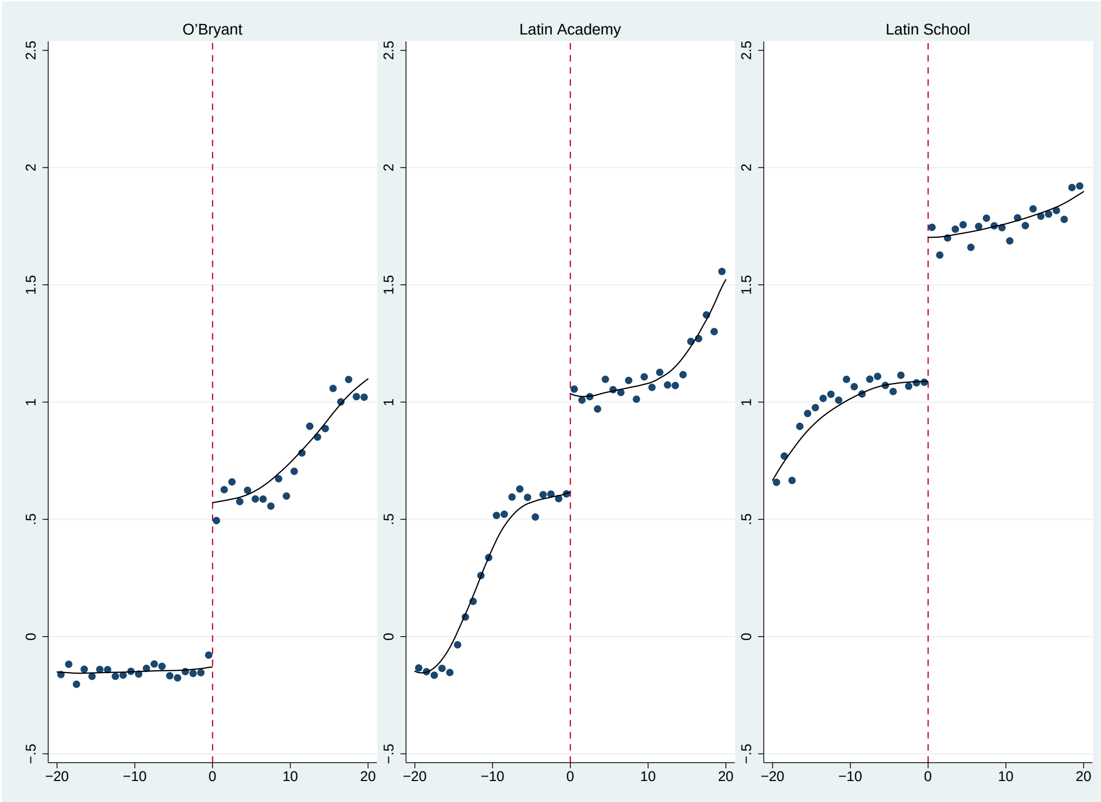

Figure 5. Average Baseline Math Scores of Peers for 7th Grade Applicants (1997-2008) in Boston

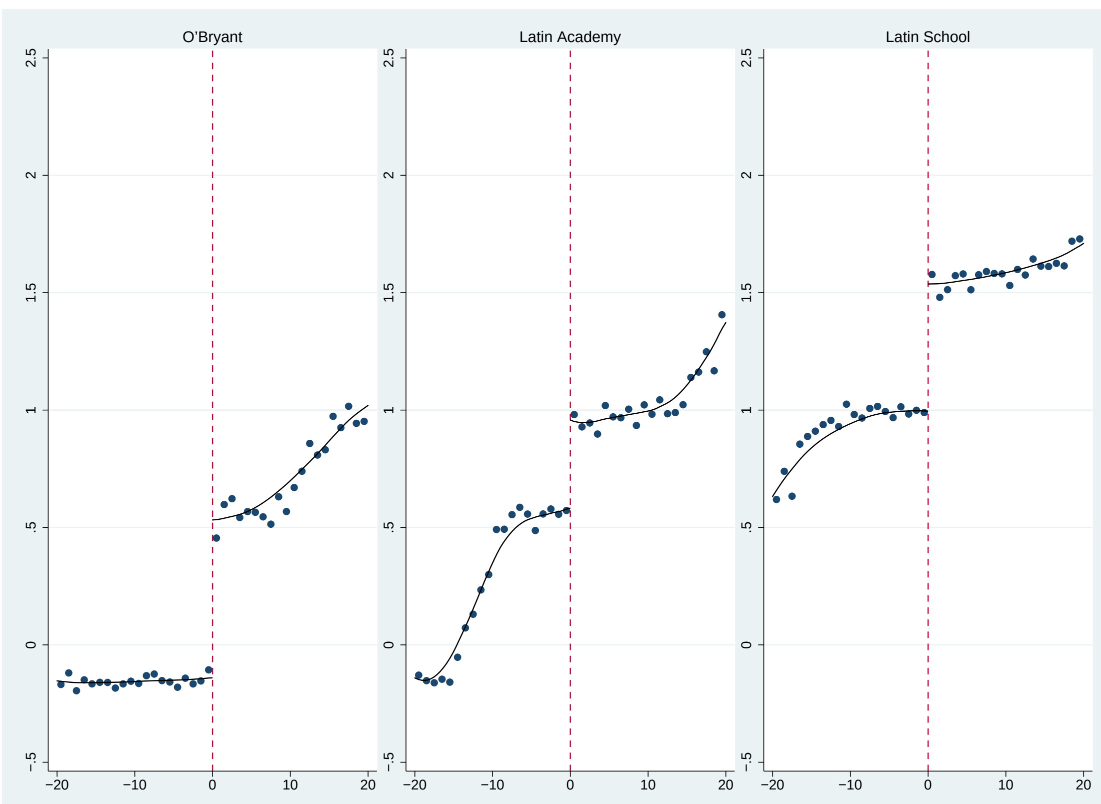

Figure 6. Average Baseline English Scores of Peers for 7th Grade Applicants (1997-2008) in Boston

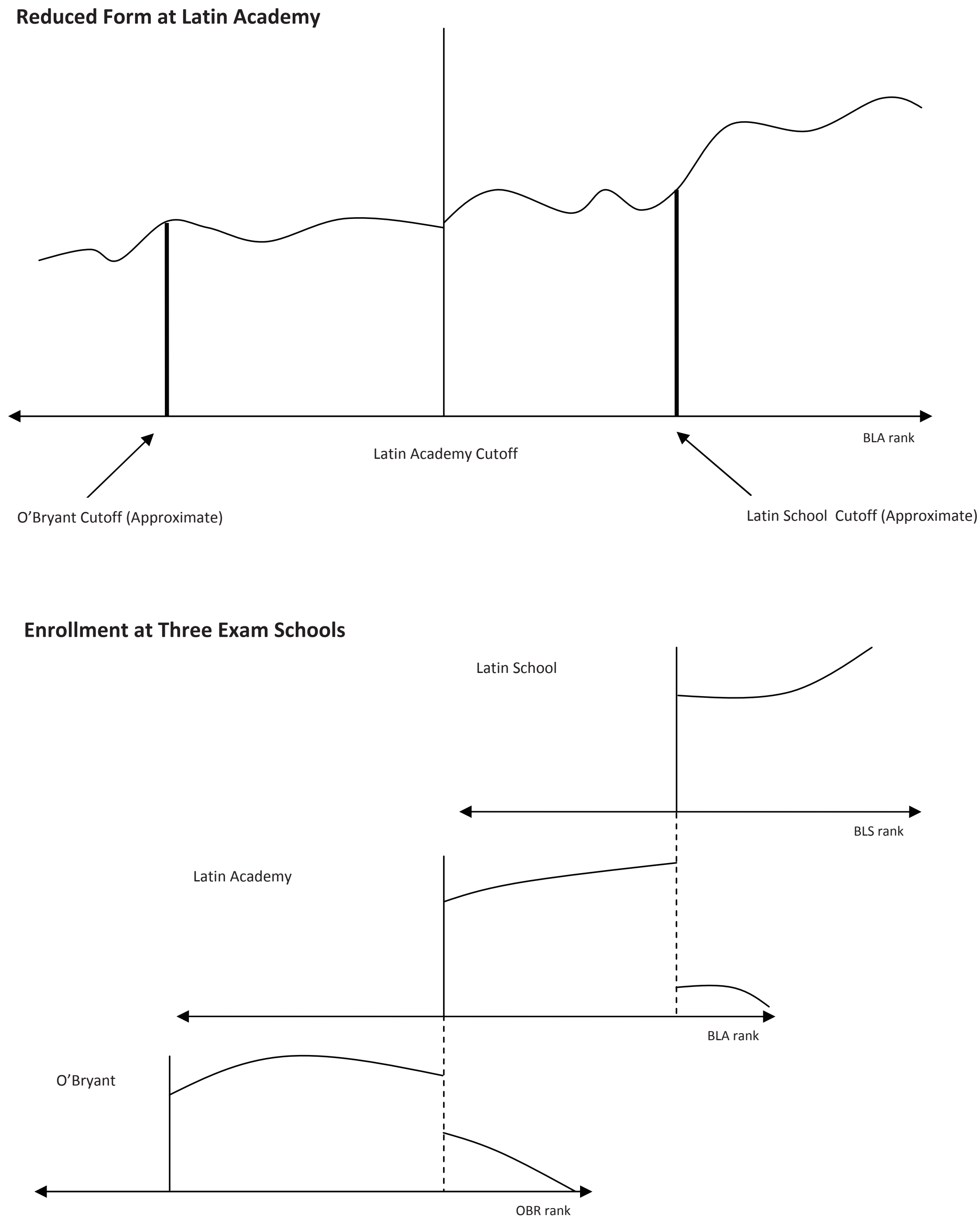

Figure 7. Reduced Form at Boston Latin Academy and Enrollment with Three Running Variables

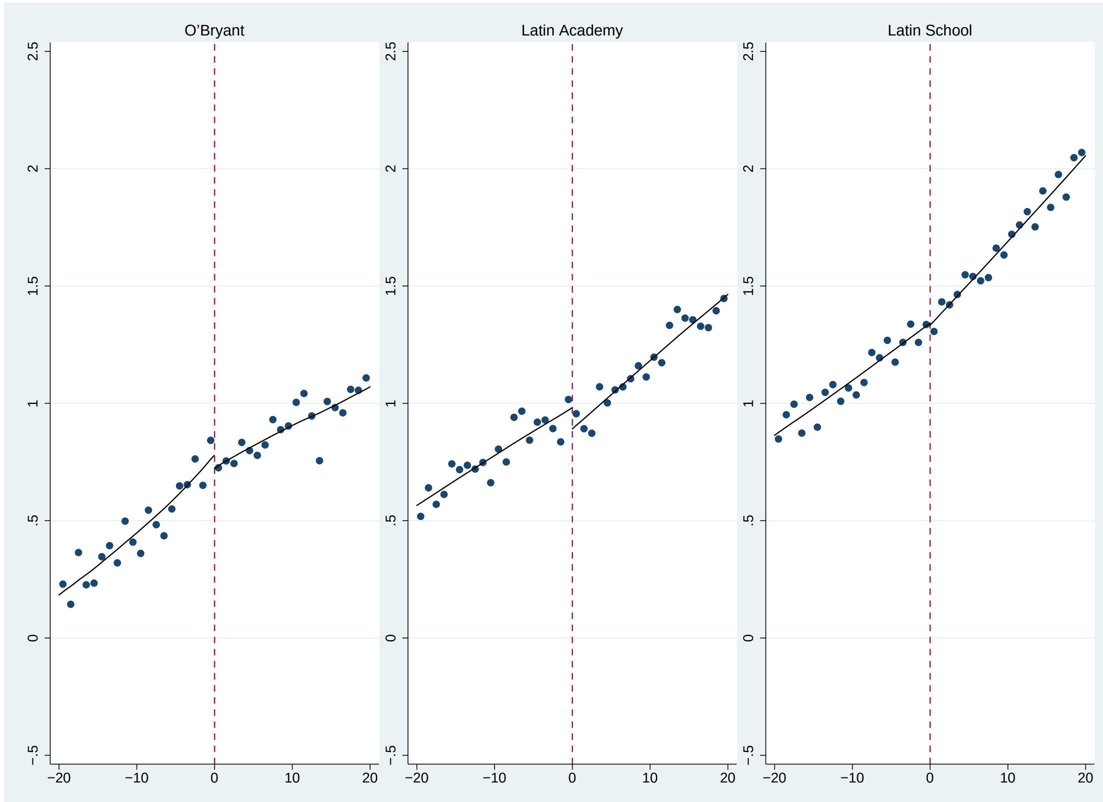

Figure 8. 7th (2006-2009) and 8th (1999-2009) Grade Math Scores for 7th Grade Applicants (1997-2007 / 2005-2008) in Boston

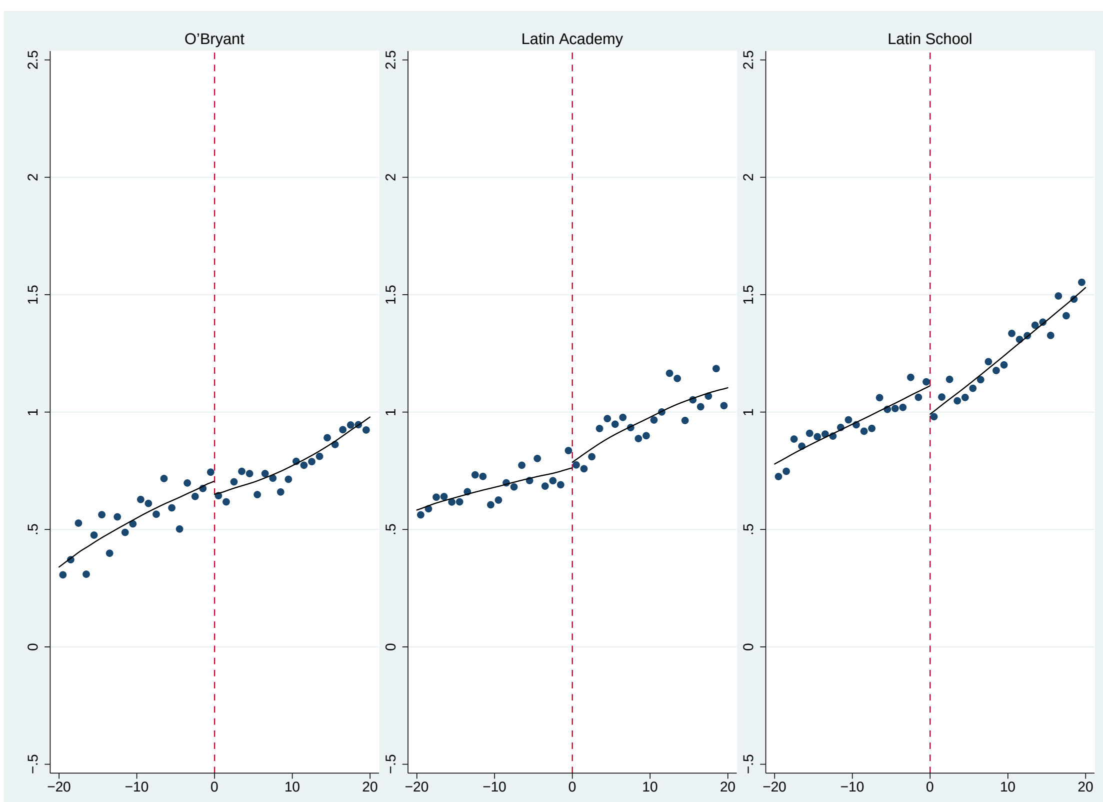

Figure 9. 7th (2001-2009) and 8th (2006-2009) Grade English Scores for 7th Grade Applicants (2000-2008 / 2004-2007) in Boston

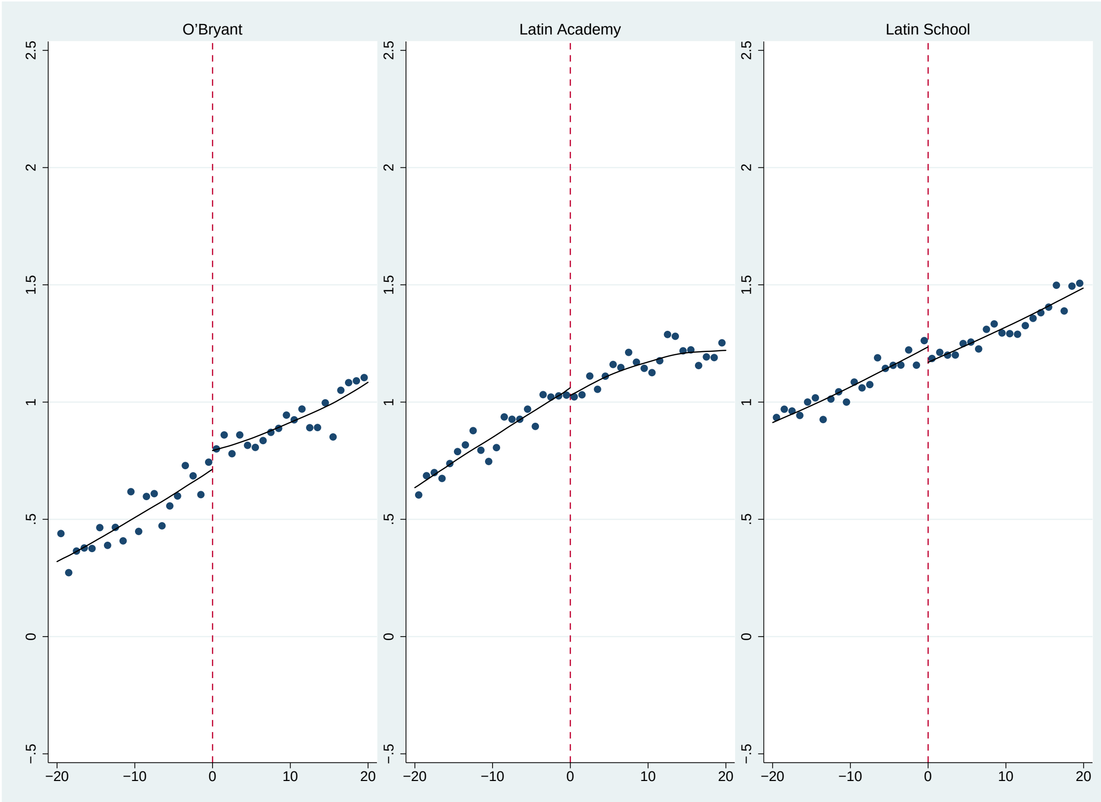

Figure 10. 10th Grade Math (2003-2009) Scores for 7th (1999-2005) and 9th (2001-2007) Grade Applicants in Boston

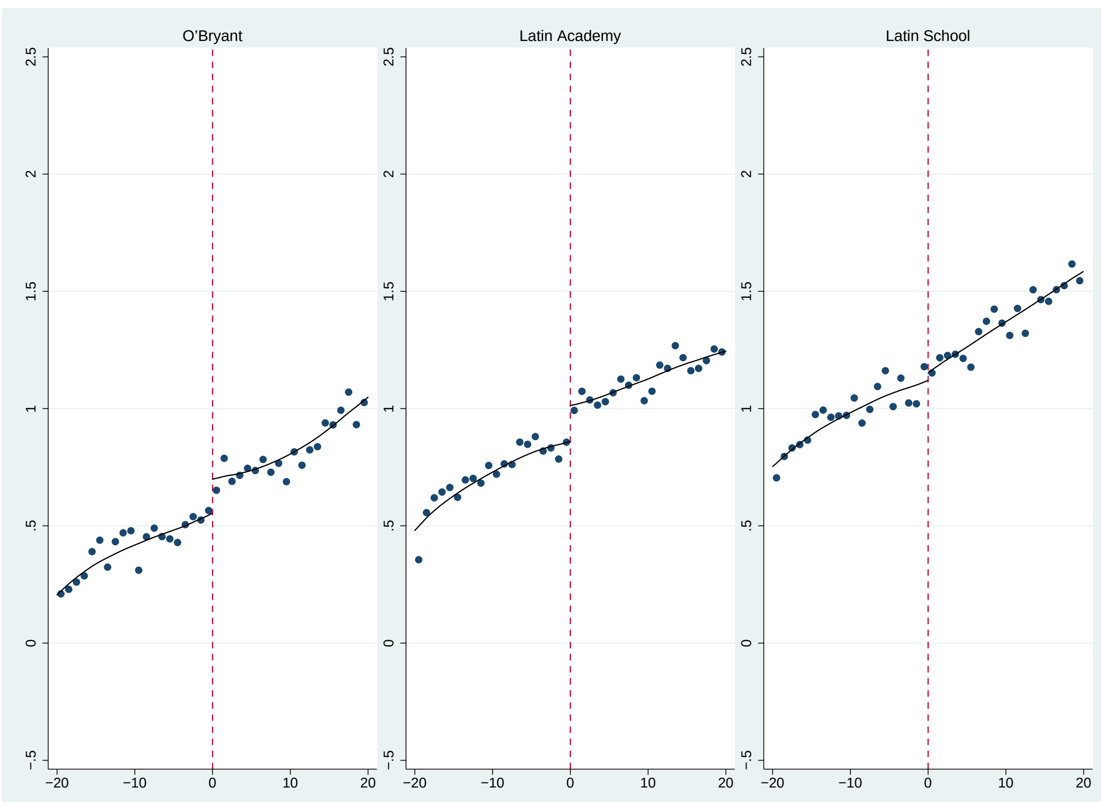

Figure 11. 10th Grade English (2003-2009) Scores for 7th (1999-2005) and 9th (2001-2007) Grade Applicants in Boston

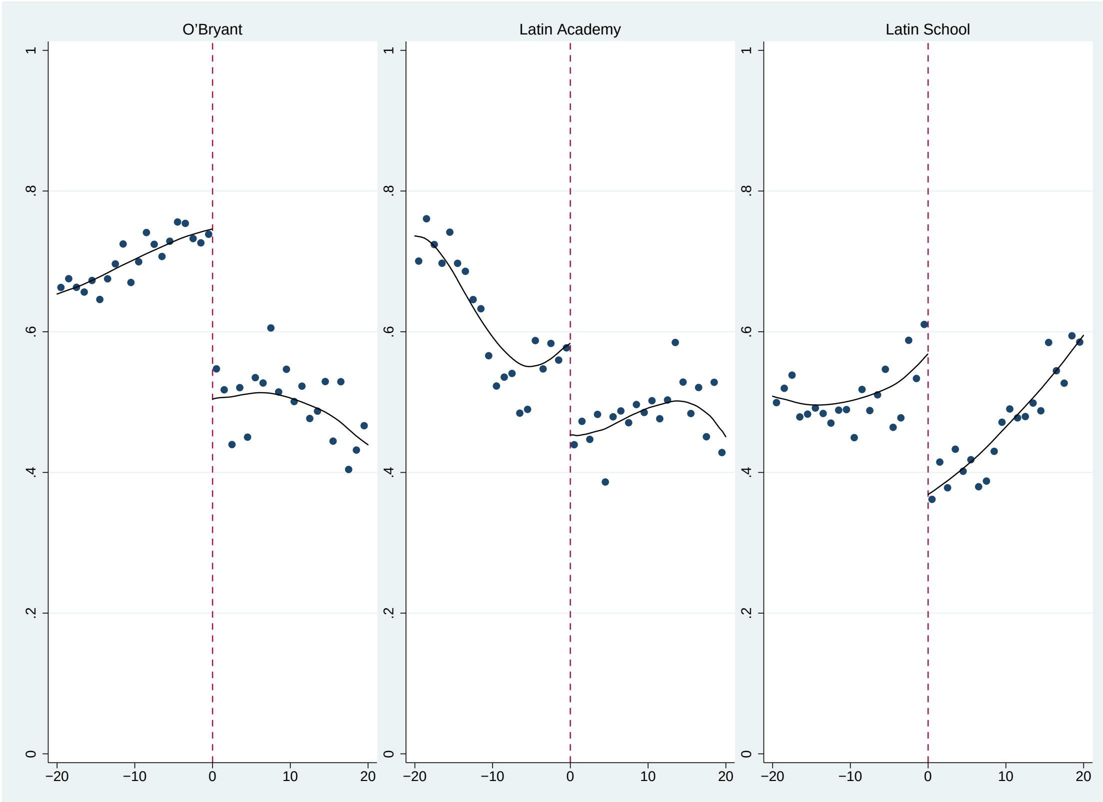

Figure 12. Rank in Baseline Math for 7th Grade Applicants (1997-2008) in Boston

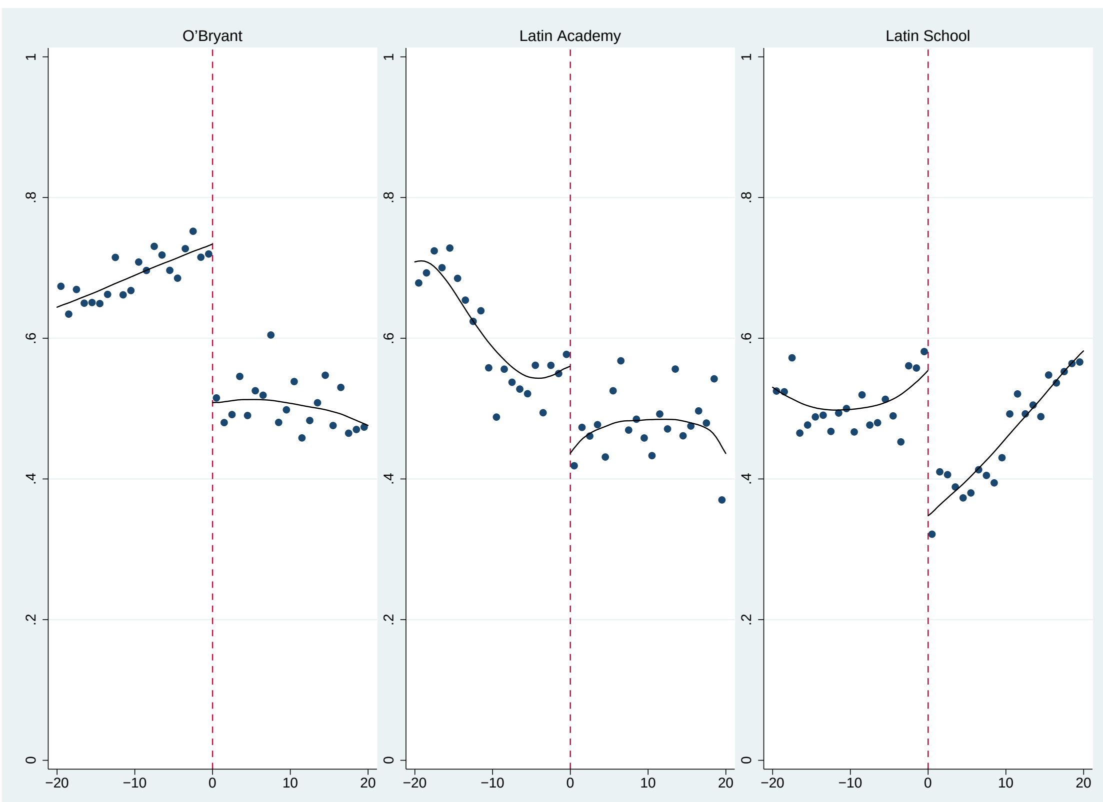

Figure 13. Rank in Baseline English for 7th Grade Applicants (1997-2008) in Boston

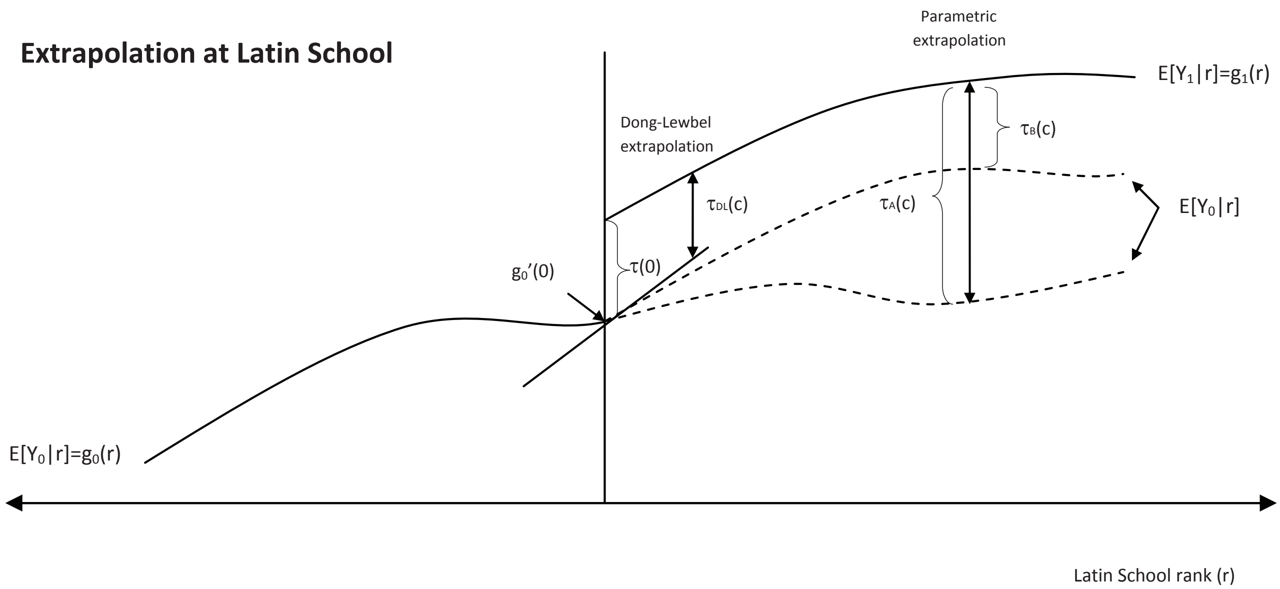

Figure 14. Parametric and Non-parametric Extrapolation Above Cutoff

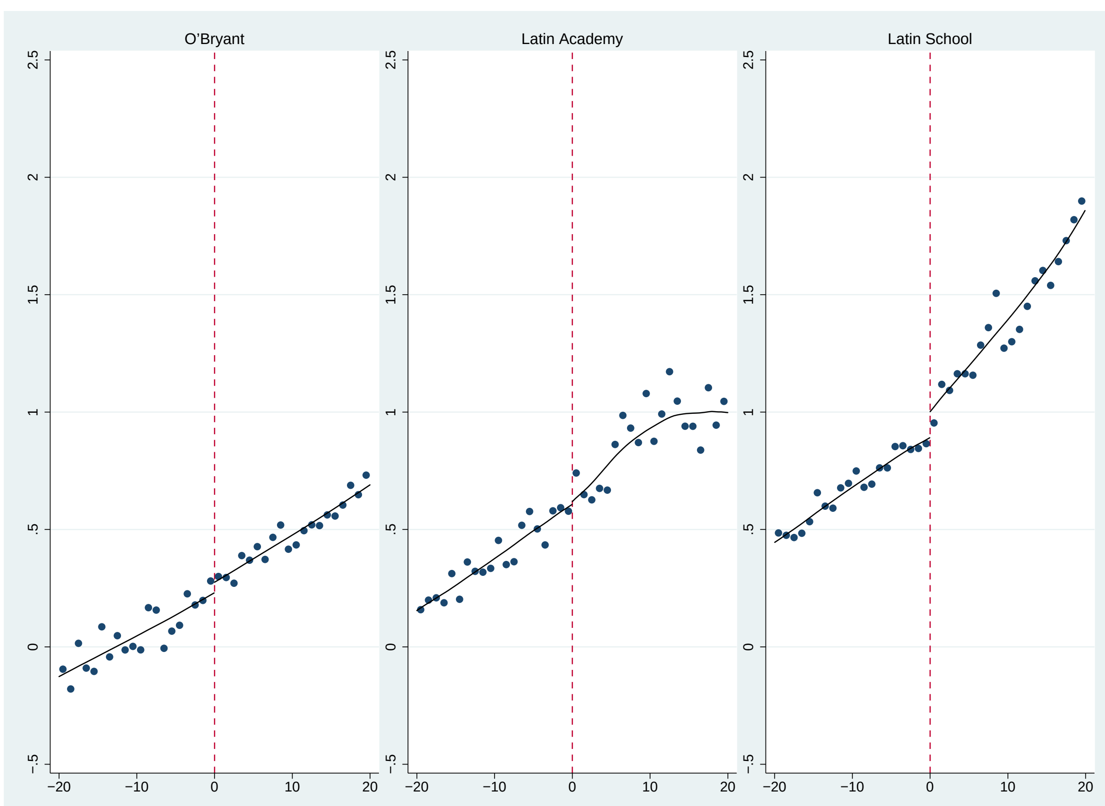

Figure 15. PSAT Scores for 7th (2000-2005) and 9th (2002-2007) Grade Applicants in Boston

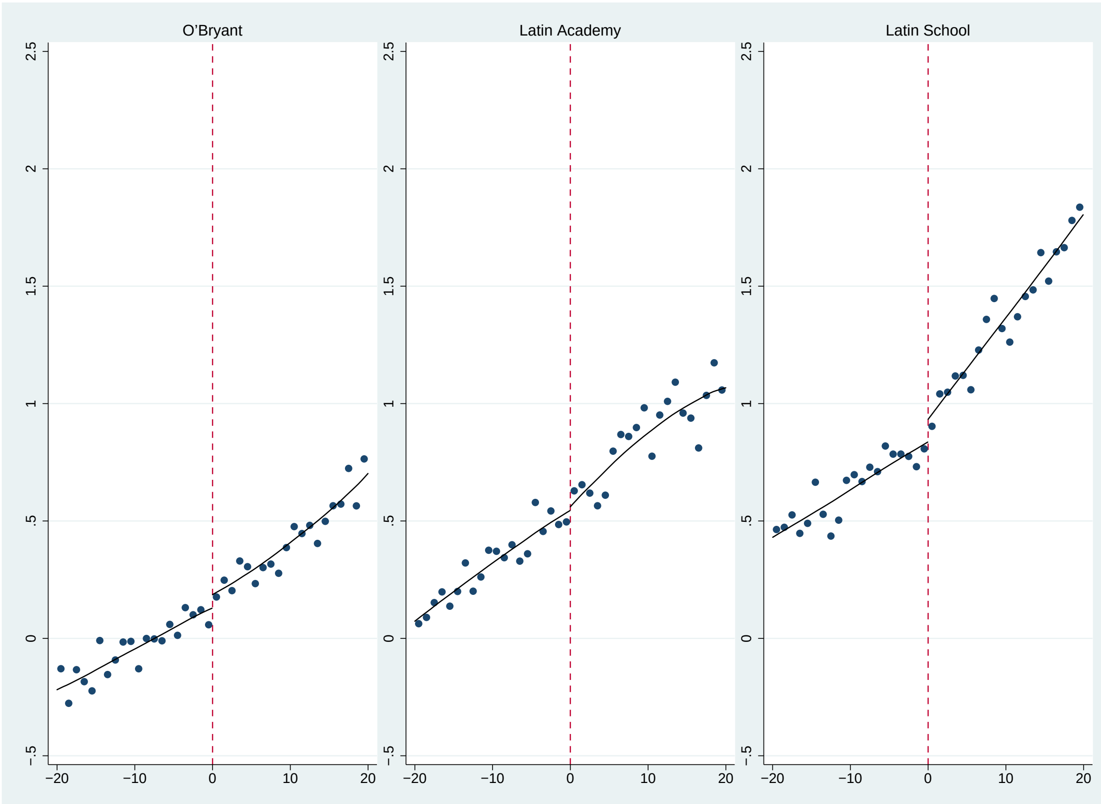

Figure 16. SAT Scores for 7th (2000-2005) and 9th (2001-2006) Grade Applicants in Boston

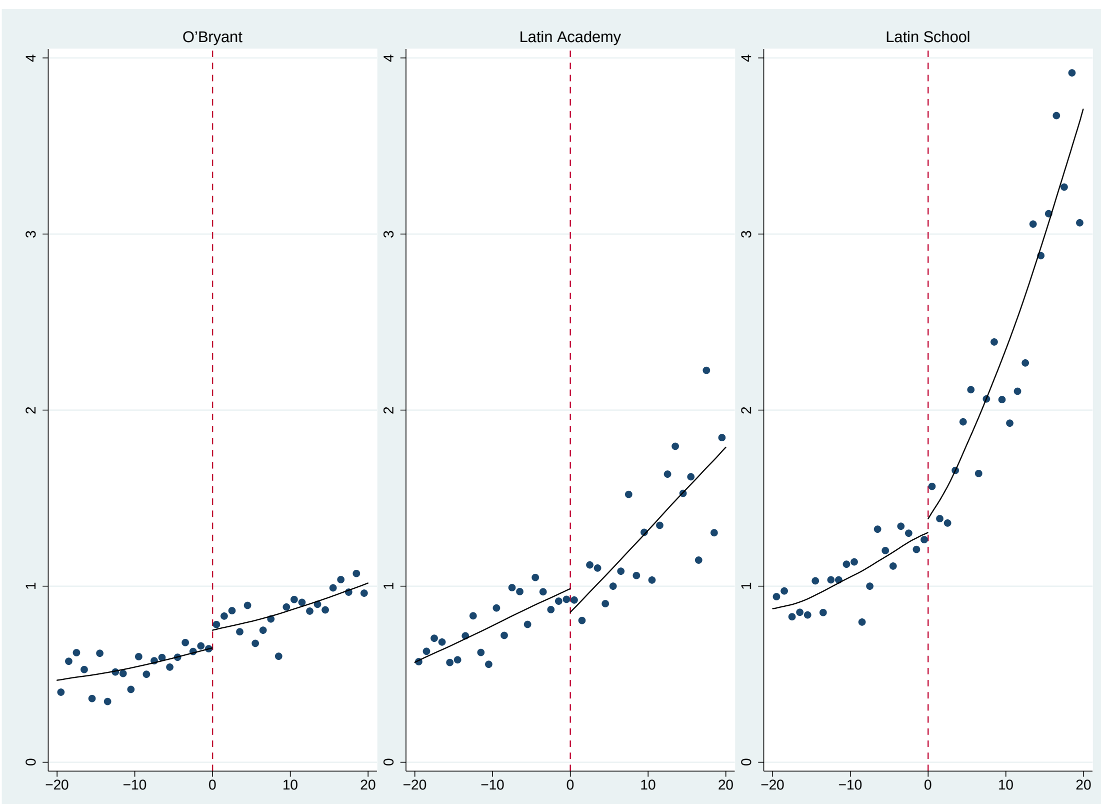

Figure 17. Number of AP Classes for 7th (1999-2004) and 9th (2001-2006) Grade Applicants in Boston

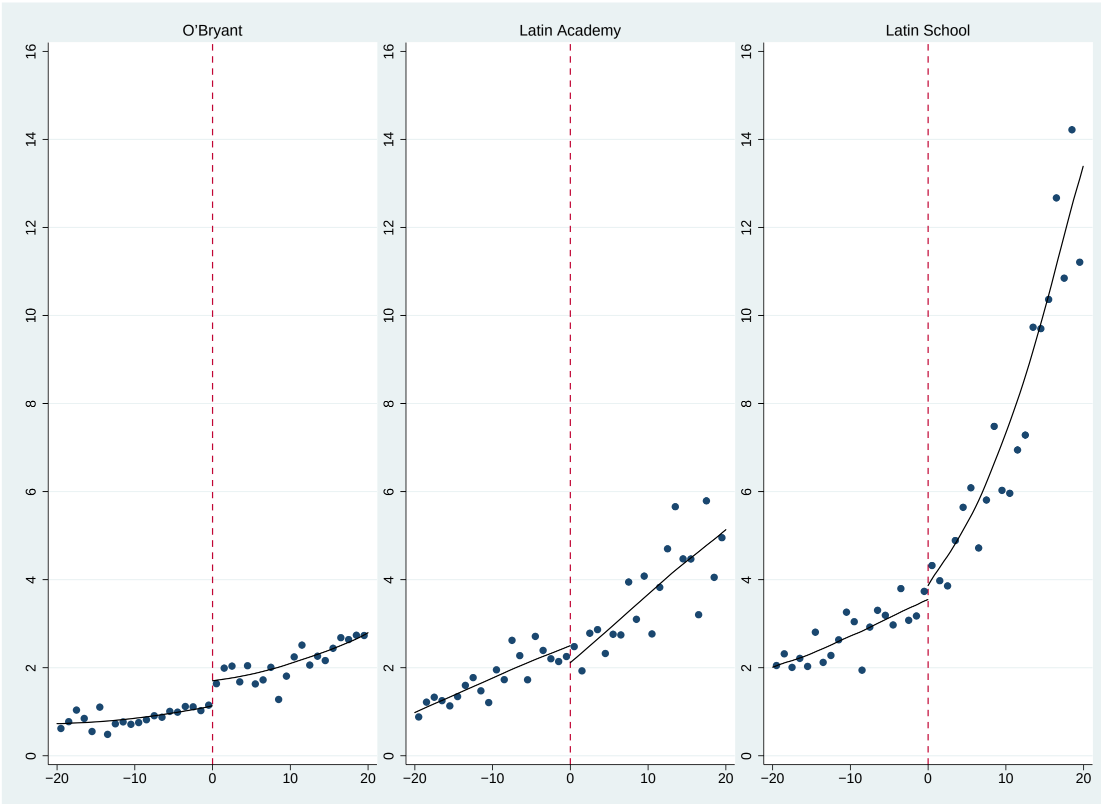

Figure 18. Sum of AP Scores for 7th (1999-2004) and 9th (2001-2006) Grade Applicants in Boston

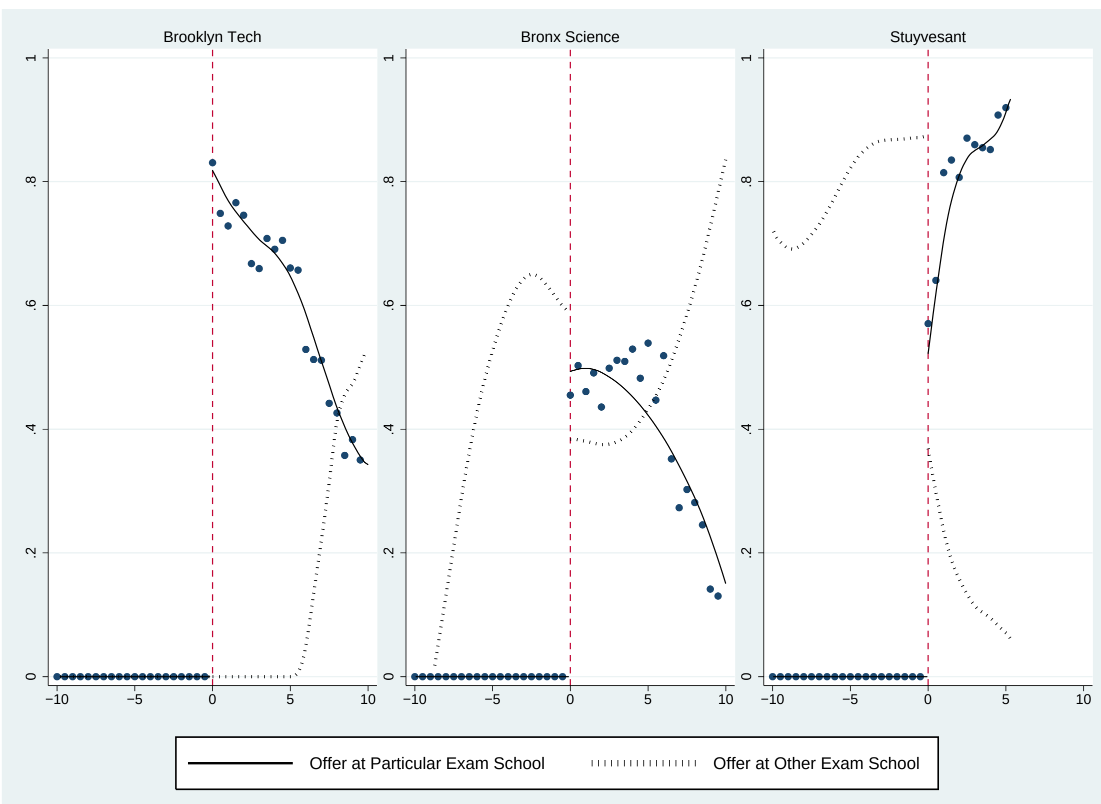

Figure 19. Offers at Each NYC Exam School for 9th Grade Applicants (2004-2007)

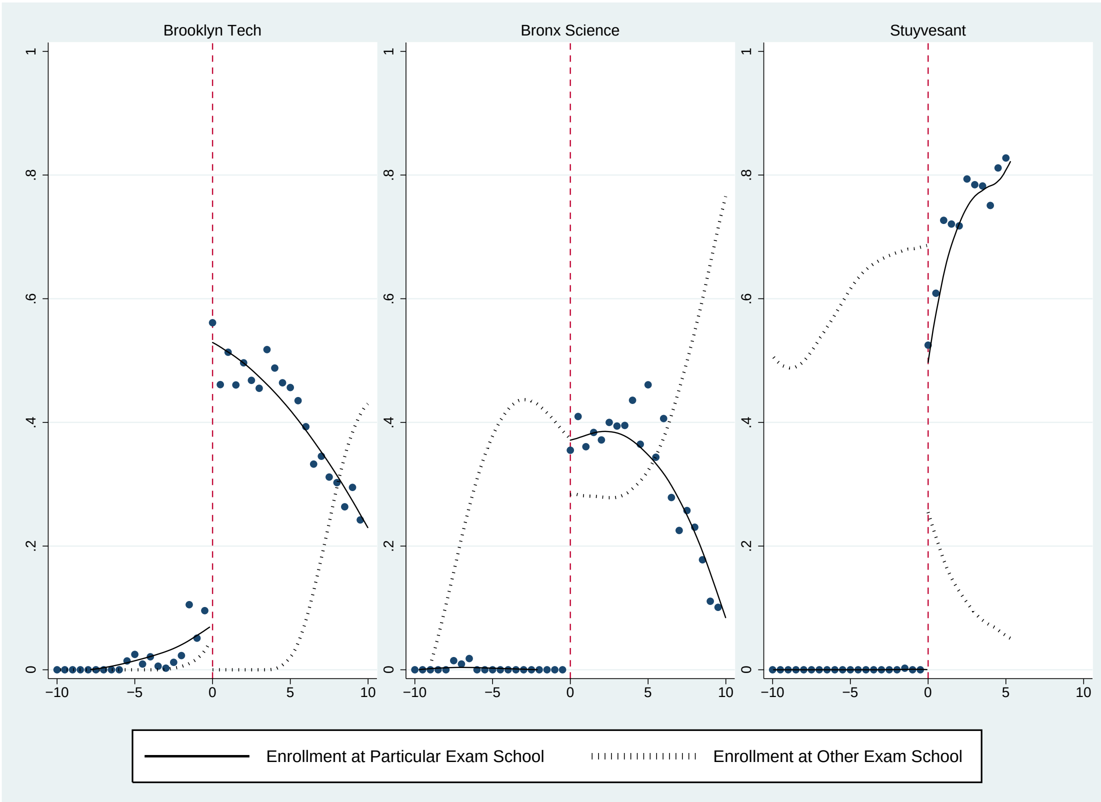

Figure 20. Enrollment at Each NYC Exam School for 9th Grade Applicants (2004-2007)

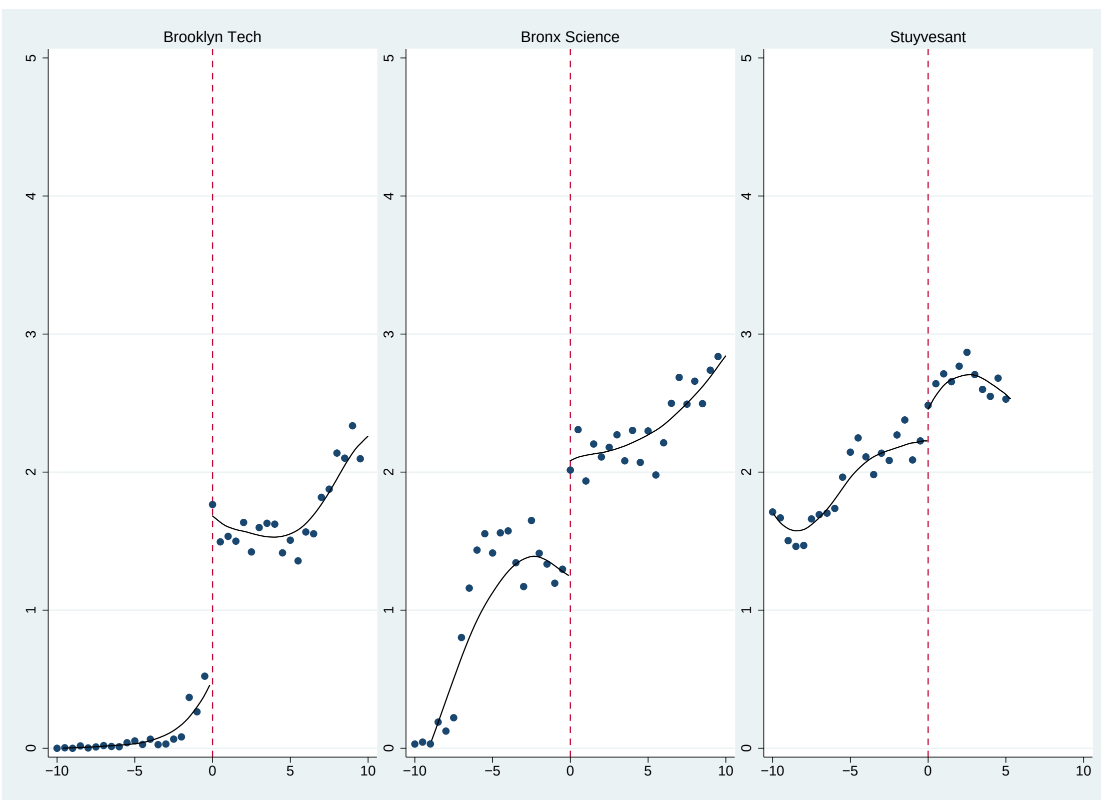

Figure 21. Years at Any Exam School for 9th Grade Applicants (2004-2007)

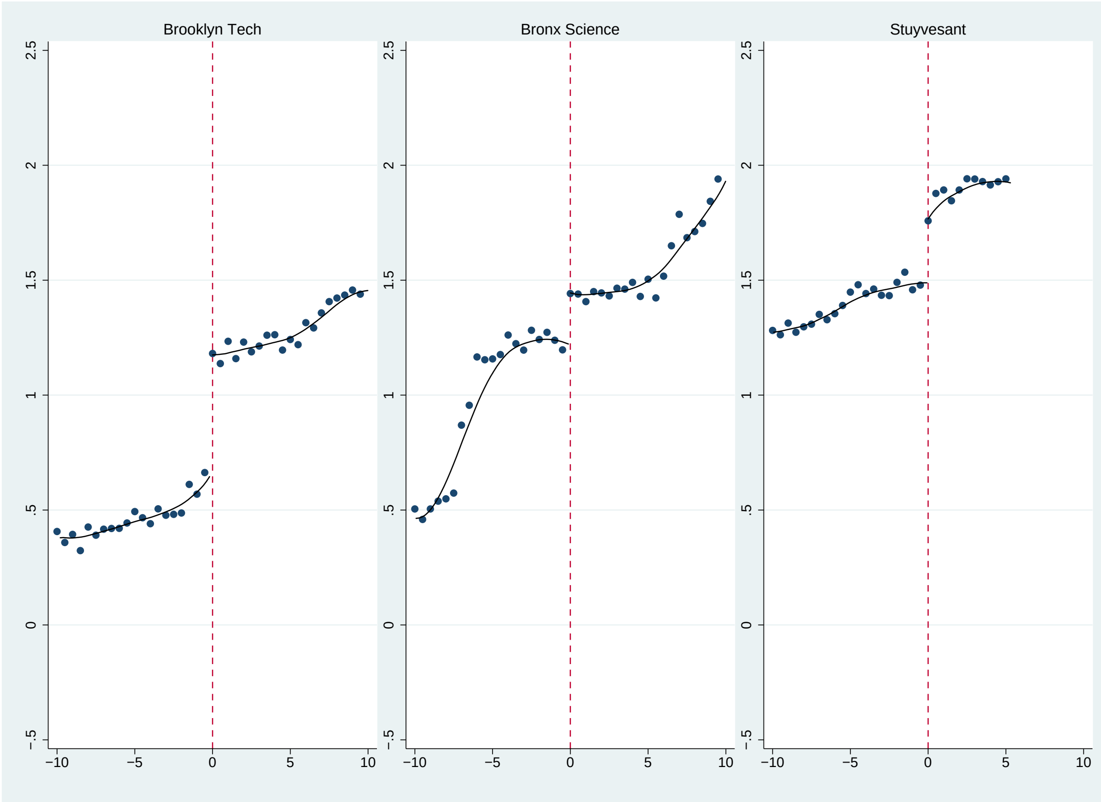

Figure 22. Average Baseline Math Score of Peers for 9th Grade Applicants (2004-2007) in NYC

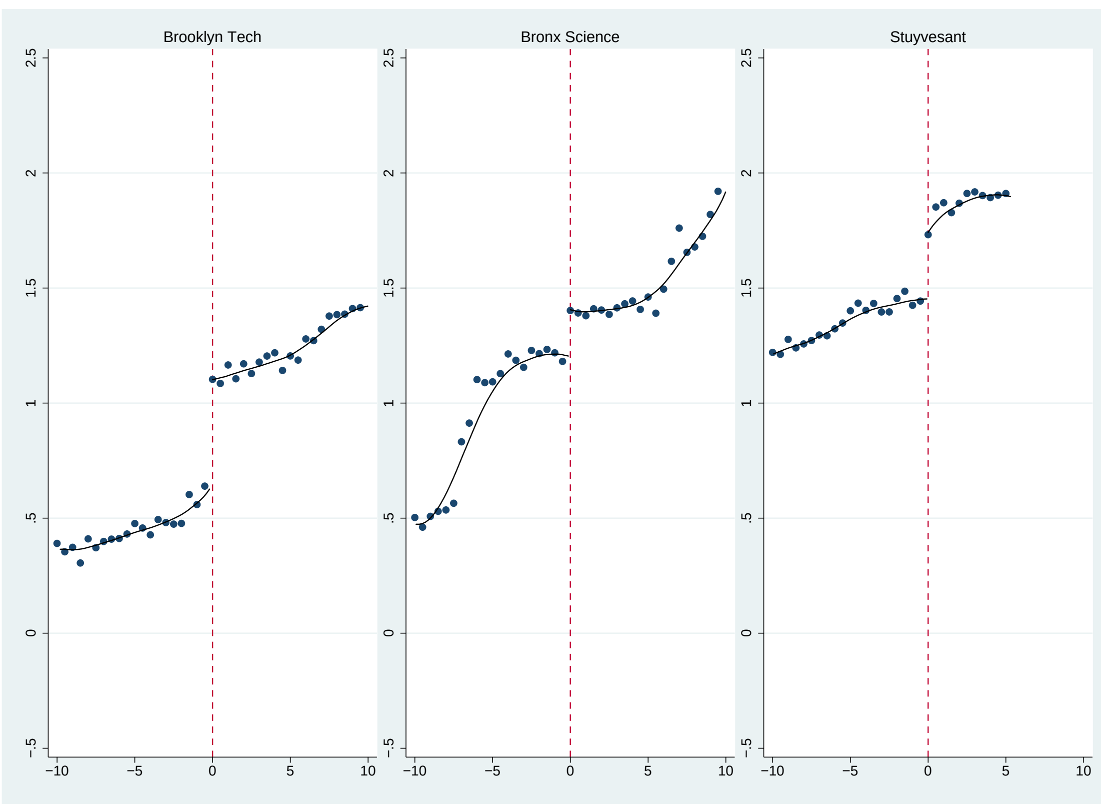

Figure 23. Average Baseline English Score of Peers for 9th Grade Applicants (2004-2007) in NYC

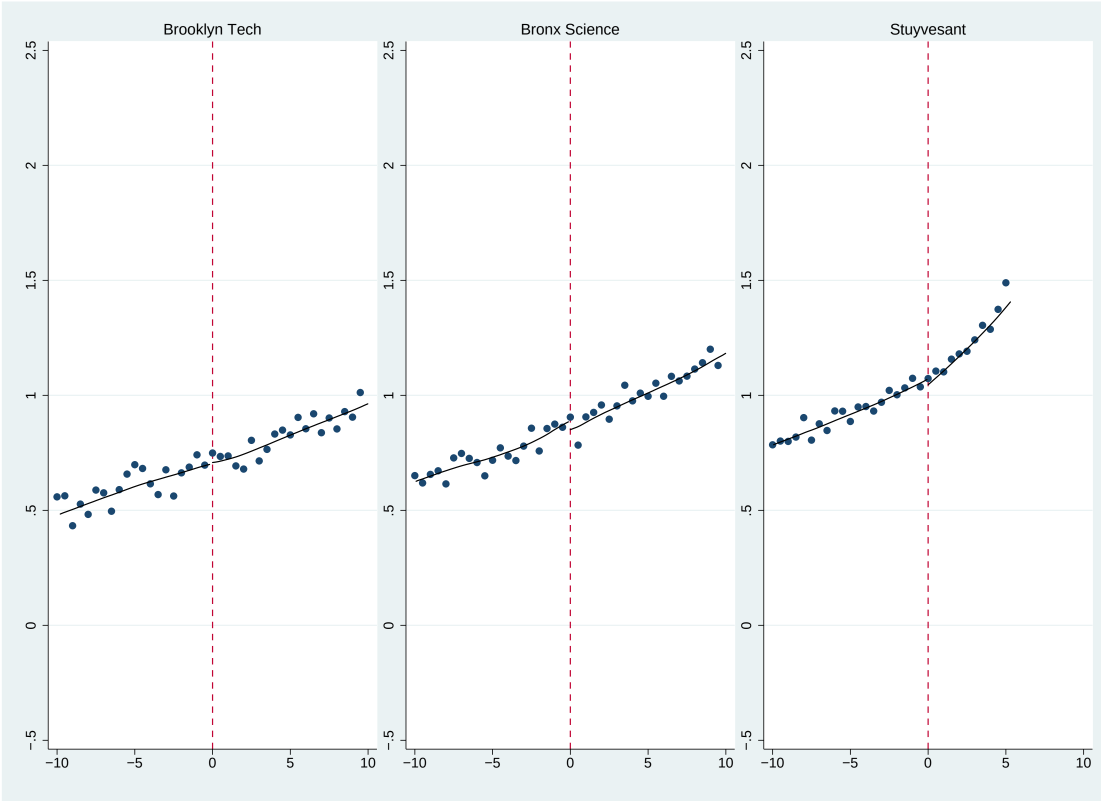

Figure 24. Advanced Math Regents Scores for 9th Grade Applicants (2004-2007) in NYC

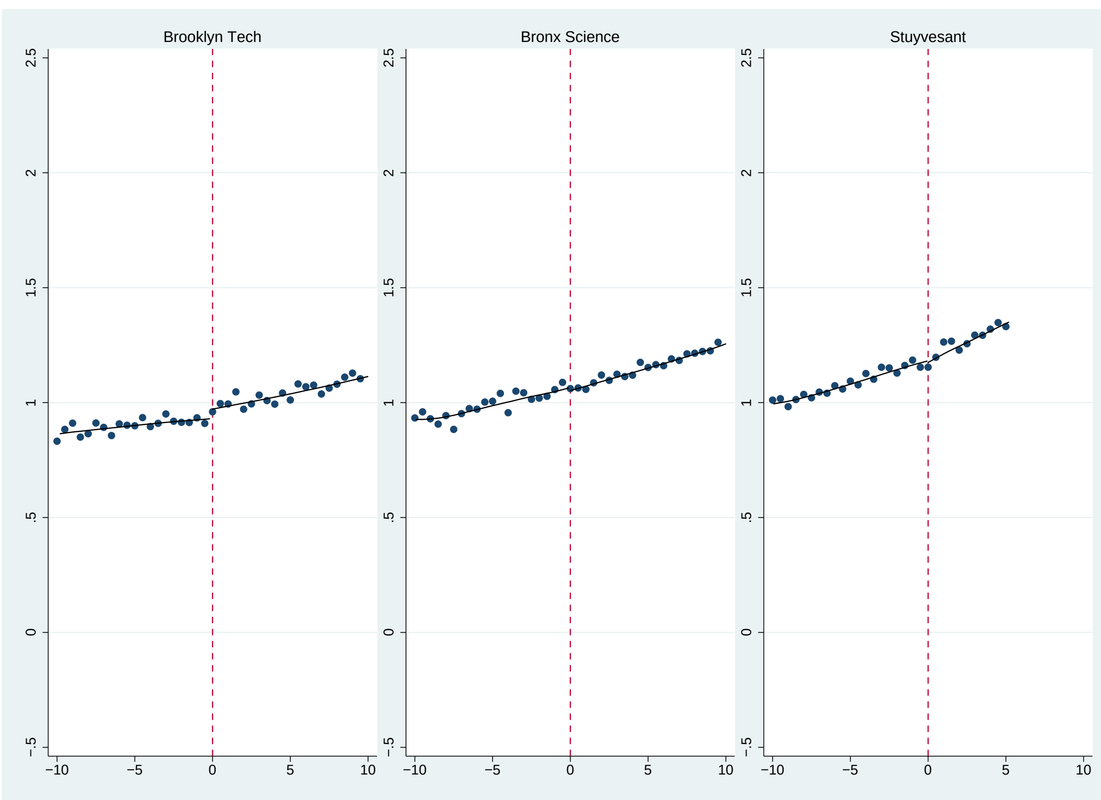

Figure 25. English Regents Scores for 9th Grade Applicants (2004-2007) in NYC

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): "Explaining Charter School Effectiveness," NBER Working Paper 17332.

Angrist, J. D., and J.-S. Pischke (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

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.

DOBBIE, W., AND R. FRYER (2011a): "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.

DOBBIE, W., AND R. G. FRYER (2011b): "Exam High Schools and Academic Achievement: Evidence from New York City," NBER Working Paper 17286.

DONG, Y., AND A. LEWBEL (2011): "Regression Discontinuity Marginal Threshold Treatment Effects," Working paper, Boston College.

DUFLO, E., P. DUPAS, AND M. KREMER (2011): "Peer Effects and the Impacts of Tracking: Evidence from a Randomized Evaluation in Kenya," *American Economic Review*, 101(5), 1739–1774.

ELLISON, G. (2011): "A Model of Curriculum Design and Student Achievement," Working paper, MIT.

ELLISON, G., AND A. SWANSON (2010): "The Gender Gap in Secondary School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions," *Journal of Economic Perspectives*, 24(2), 109–128.

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.

Hsieh, C.-T., and M. Urquiola (2006): "The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program," *Journal of Public Economics*, 90, 1477–1503.

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*, 120, 1399–1429.

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.

Li, Q., and J. Racine (2007): *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, Princeton, New Jersey.

MacLeod, B., and M. Urquiola (2009): "Anti-Lemons: School Reputation and Educational Quality," NBER Working Paper, 15112.

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.

Pathak, P. A. (2011): "The Mechanism Design Approach to Student Assignment," *Annual Reviews*, 3, 513–536.

Pop-Eleches, C., and M. Urquiola (2010): "Going to a Better School: Effects and Behavioral Responses," Working paper, Columbia University.

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*.

Stone, C. J. (1982): "Optimal Global Rates of Convergence for Nonparametric Regression," *Annals of Statistics*, 10(4), 1040–1053.

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