

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

BY ATILA ABDULKADIROĞLU, JOSHUA ANGRIST, AND PARAG PATHAK<sup>1</sup>

Parents gauge school quality in part by the level of student achievement and a school's racial and socioeconomic mix. The importance of school characteristics in the housing market can be seen in the jump in house prices at school district boundaries where peer characteristics change. The question of whether schools with more attractive peers are really better in a value-added sense remains open, however. This paper uses a fuzzy regression-discontinuity design to evaluate the causal effects of peer characteristics. Our design exploits admissions cutoffs at Boston and New York City's heavily over-subscribed exam schools. Successful applicants near admissions cutoffs for the least selective of these schools move from schools with scores near the bottom of the state SAT score distribution to schools with scores near the median. Successful applicants near admissions cutoffs for the most selective of these schools move from above-average schools to schools with students whose scores fall in the extreme upper tail. Exam school students can also expect to study with fewer nonwhite classmates than unsuccessful applicants. Our estimates suggest that the marked changes in peer characteristics at exam school admissions cutoffs have little causal effect on test scores or college quality.

KEYWORDS: Peer effects, school choice, deferred acceptance, selective education.

### 1. INTRODUCTION

A three bedroom house on the northern edge of Newton, Massachusetts costs \$412,000 (in 2008 dollars), while across the street, in Waltham, a similar place can be had for \$316,000.<sup>2</sup> Black (1999) attributed this and many similar Massachusetts contrasts to differences in perceived school quality. Indeed, 92 percent of Newton's high school students are graded proficient in math, while only 78 percent are proficient in Waltham. These well-controlled com-

<sup>1</sup>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 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 are grateful for comments from participants in the June 2010 Tel Aviv Frontiers in the Economics of Education conference, the Summer 2011 NBER Labor Studies workshop, and the December 2011 Hong Kong Human Capital Symposium. Thanks also go to Jonah Rockoff for comments and data on teacher tenure in NYC. We are also grateful to Daron Acemoglu, Gary Chamberlain, Yingying Dong, Guido Imbens, and especially to Glenn Ellison for helpful discussions. Alex Bartik, Weiwei Hu, and Miikka Rokkanen provided superb research assistance. We thank the Institute for Education Sciences for financial support under Grant R305A120269. Pathak also thanks the Graduate School of Business at Stanford University, where parts of this work were completed, and the NSF for financial support; and Abdulkadiroglu acknowledges an NSF-CAREER award.

<sup>2</sup>These are average prices of 42 three bedroom units in Newton and 27 units in Waltham, separated by 0.1 miles or less, as quoted on Greater Boston's Multiple Listing Service for transactions between 1998 and 2008.

parisons suggest *something* changes at school district boundaries. Parents looking for a home are surely aware of achievement differences between Newton and Waltham, and many are willing to pay a premium to see their children attend what appear to be better schools. At the same time, it is clear that differences in achievement levels can be a highly misleading guide to value-added, a possibility suggested by theoretical and empirical analyses in Rothstein (2006), Hastings, Kane, and Staiger (2009), and MacLeod and Urquiola (2009), among others.

Similar observations can be made regarding the relationship between racial composition and home prices. For over a half-century, American education policy has struggled with the challenge of racial integration. The view that racial mixing contributes to learning motivates a range of social interventions ranging from within-district busing and court supervision of school assignment, to Boston's iconic Metco program, which sends minority children to mostly white suburban districts. In this context as well, home-buying parents vote with their housing dollars—typically for more white classmates—as shown recently by Boustan (2012) using cross-border comparisons in the spirit of Black (1999).<sup>3</sup>

An ideal experiment designed to reveal causal effects of peer characteristics would randomly assign the opportunity to attend schools with high-achieving peers and fewer minority classmates. The subjects of such a study should be a set of families likely to take advantage of the opportunity to attend schools that differ from their default options. Imagine sampling parents found in suburban Boston real estate offices, as they choose between homes in Newton and Waltham. We might randomly offer a subset of those who settle for Waltham a voucher that entitles them to send their children to Newton schools in spite of their choice of a Waltham address. This manipulation bears some resemblance to the Moving to Opportunity (MTO) experiment, which randomly allocated housing vouchers valid only in low-poverty neighborhoods. MTO was a complicated intervention, however, that did not manipulate the school environment in isolation (see Kling, Liebman, and Katz (2007) and Sanbonmatsu, Ludwig, Katz, Gennetian, Duncan, Kessler, McDade, and Lindau (2011)).

While a perfect peer characteristics experiment is hard to engineer, an important set of existing educational institutions induces quasi-experimental variation that comes close to the ideal experiment. A network of selective public schools in Boston and New York known as exam schools offer public school students the opportunity to attend schools with much higher achieving peers. Moreover, in these mostly nonwhite districts, exam schools have a markedly higher proportion of white classmates than do the public schools that applicants are otherwise likely to attend. Of course, exam school admissions are not

<sup>3</sup>Guryan (2004) found that court-order integration schemes increase nonwhite high school graduation rates without hurting whites, but evidence on the achievement consequences of busing for racial balance is mixed (see, e.g., Hoxby (2000) and Angrist and Lang (2004)).

made by random assignment; rather, students are selected by an admissions test with sharp cutoffs for each school and cohort. This paper exploits these admissions cutoffs in a fuzzy regression discontinuity (RD) design that identifies causal effects of peer achievement and racial composition for applicants to the six traditional exam schools operating in Boston and New York. The application of RD methods in this context generates a number of challenges related to the real-world messiness of school assignment and the exclusion restrictions needed to interpret two-stage least squares (2SLS) estimates. Solutions for these problems are detailed in the sections that follow.<sup>4</sup>

## 2. INSTITUTIONAL BACKGROUND

Boston's three exam schools span grades 7–12. The best-known is the Boston Latin School, which enrolls about 2,400 students. Seen by many 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)).<sup>5</sup> Boston Latin School is a model for other exam schools, including the recently opened Brooklyn Latin School 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 now 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 20th century and share a number of features with Boston's exam schools. Stuyvesant and Bronx Science appear on Newsweek's list of elite public high schools, and all three have been high 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 Tech has about 4,500 students. New York opened three new exam schools in 2002: 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

<sup>4</sup>Neighborhoods and schools are not the only settings in which peer effects might arise, but these are among the most commonly encountered contexts for peer effects in social science research. A voluminous literature, summarized in a recent survey by Sacerdote (2011), reveals a strong association between the performance of students and their classmates.

<sup>5</sup>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.

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 the new schools from this study because they are not as well established as New York's traditional exam schools, and some have unusual characteristics such as small enrollment.<sup>6</sup>

Boston Public Schools span a range of peer achievement that may be unique among American urban districts. Like many urban students elsewhere in the United States, Boston exam school applicants who fail to enroll in an exam school end up at schools with average SAT scores well below the state average, in this case, at schools close to the 5th percentile of the distribution of school averages in the state. By contrast, O'Bryant's average SAT falls at the 40th percentile of the state distribution of averages, a big step up from the overall BPS average, but not elite in an absolute sense. Successful Boston Latin Academy applicants find themselves at a school with average SATs around the 80th percentile of the distribution of school means, while the average SAT score at the Boston Latin School is the fourth highest among public schools in the state.

Data from New York's exam schools enrich this picture by allowing us to evaluate the impact of exposure to extremely high-achieving peers. The least selective of New York's three traditional exam schools, Brooklyn Tech, is attended by students with average SAT scores found around the 99th percentile of the distribution of average scores in New York State, a level comparable to the Boston Latin School. Successful applicants to Brooklyn Tech typically move from schools where peer achievement is around the 30th percentile of the school average SAT distribution. Students at the two most selective New York exam schools are exposed to the brightest of classmates, with the Bronx Science average SAT falling at percentile 99.9, while Stuyvesant has the highest average SAT scores in New York State, placing it among the top five public schools nationwide.

As far as we know, ours is one of two RD analyses of achievement effects at highly selective U.S. exam schools. In independent contemporaneous work, [Dobbie and Fryer \(2013\)](#) estimated the reduced-form impact of admissions offers at New York exam schools; their analysis showed no impact on college enrollment or quality. Selective high schools have also been studied elsewhere. [Pop-Eleches and Urquiola \(2013\)](#) estimated 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 reported large score gains for those attending a selective school in

<sup>6</sup>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 City Department of Education.

Trinidad and Tobago. On the other hand, [Clark \(2008\)](#) found only modest score gains at selective U.K. schools. Likewise, using admissions lotteries to analyze the consequences of selective middle school attendance in China, [Zhang \(2010\)](#) found no achievement gains for students randomly offered seats at a selective school. In contrast with our work, none of these studies interpret the reduced-form impact of exam school offers as operating through specific causal channels for which there is a clear first stage.<sup>7</sup>

Selective institutions are more commonly found in American higher education than at the secondary level. [Dale and Krueger \(2002\)](#) compared students who were accepted by the same sets of colleges but made different choices in terms of selectivity. 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 \(2009\)](#) reported 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 [Hoxby \(2000\)](#), [Hanushek, Kain, Markman, and Rivkin \(2003\)](#), [Angrist and Lang \(2004\)](#), [Hoxby and Weingarth \(2006\)](#), [Lavy, Silva, and Weingarth \(2012\)](#), [Ammermueller and Pischke \(2009\)](#), [Imberman, Kugler, and Sacerdote \(2012\)](#), and [Carrell, Sacerdote, and West \(2012\)](#). Findings in the voluminous education peer effects literature are mixed and not easily summarized. It seems fair to say, however, that the likelihood of 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 contradictory evidence on the latter ([Duflo, Dupas, and Kremer \(2011\)](#)).

The exam schools of interest here are also associated with marked changes in peers' racial mix. In our fuzzy RD setup, which uses exam school admissions offers to construct instrumental variables for peer characteristics, enrollment compliers at Boston Latin Academy are exposed to a peer group that falls from two-thirds to 40 percent black and Hispanic. The proportion minority falls by half, from 40 to 20, for Latin School compliers.

Changes in peer composition are not necessarily the only component of the education production function associated with changes in attendance at the exam schools in our sample. Still, our research design holds many potential confounders fixed, including family background, ability, and residential sorting. The principal sources of omitted variables bias, in our setup interpreted here as violating an exclusion restriction, are changes in resources or curriculum. We argue that bias from omission of these factors is likely to be positive,

<sup>7</sup>[Pop-Eleches and Urquiola \(2013\)](#) reported a peer achievement first stage in their analysis of Romanian selective schools, but the effect of a Romanian exam school offer on peer composition is small and, as the authors noted, unlikely to explain their findings.

reinforcing our interpretation of the findings as offering little evidence for peer achievement or racial composition effects on state test scores; PSAT, SAT, and AP scores; or college quality. As a theoretical matter, we also show that 2SLS estimates are free of omitted variables bias if resource and curriculum changes are themselves a consequence of peer composition. Importantly, most of the 2SLS estimates reported here are reasonably precise; we can rule out relatively modest peer composition effects.

The next section describes Boston data and school assignment. A complication here is Boston's deferred acceptance (DA) assignment algorithm. As a preliminary to the estimation of causal effects, we develop an empirical strategy that embeds DA in an RD framework.

### 3. BOSTON DATA AND ADMISSIONS PROCESS

#### 3.1. *Data*

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

BPS demographic and registration information was merged with Massachusetts Comprehensive Assessment System (MCAS) scores using student identification numbers.<sup>8</sup> The MCAS database contains raw scores for math, English Language Arts (ELA), Writing, and Science. MCAS tests are taken each spring, typically in grades 3–8 and 10. 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 tests. Baseline English scores for 9th grade applicants come from 8th grade math and 7th grade English tests (the 8th grade English exam was introduced in 2006). We lose some applicants with missing baseline scores. Other outcomes examined here include scores on the Preliminary SAT (PSAT), the SAT, and Advanced Placement (AP) exams from the College Board. For the purposes of our analysis, MCAS, PSAT, and SAT scores were standardized by subject, grade, and year to have mean zero and unit variance in the BPS population. Data on college enrollment come from the National Student Clearinghouse, as reported to BPS for their students.

Our analysis file combines student registration, test scores, and college outcome files with the BPS exam school applicant file. The exam school applicant file records grade, year, sending school, applicants' preference ranking of exam schools, applicants' Independent Schools Entrance Exam (ISEE) test scores,

<sup>8</sup>The MCAS is a state-mandated series of achievement tests that includes a high-stakes exit exam in 10th grade.

and each exam school's ranking of its applicants on the basis of ISEE scores and grades. This ranking variable determines exam school admissions decisions.

Our analysis sample includes BPS-enrolled students who applied for exam school seats in 7th grade from 1997 to 2008 or in 9th grade from 2001 to 2007. We focus on applicants enrolled in BPS at the time of application because we expect the peer experiment to be most dramatic for this group. Moreover, 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. Table A.I in Appendix A reports additional information on demographic characteristics and baseline scores for all BPS students and for Boston exam school applicants and those enrolled in exam schools. Exam school applicants are clearly a select group, with markedly higher baseline scores than other BPS students. For example, grade 7 applicants' 4th grade math scores are more than  $0.7\sigma$  higher than those of a typical BPS student. Enrolling students are even more positively selected. The Boston data appendix explains the analysis file further, and describes test coverage and application timing in detail.

### 3.2. Exam School Admissions

Boston exam school admissions are based on the student-proposing DA algorithm, which matches students to schools on the basis of student preferences and schools' rankings of their applicants. DA complicates RD because it loosens the direct link between the running variable and school admissions offers. Our econometric strategy therefore begins by constructing analysis samples that restore a direct link, so that offers are sharp around cutoffs. This approach seems likely to be useful elsewhere, since DA is now used for school assignment in Chicago, Denver, New York City, Newark, and England (Abdulkadiroğlu, Pathak, and Roth (2009), Pathak and Sönmez (2008, 2013)), as well as in Boston.

Boston residents interested in an exam school seat take the ISEE in the fall of the school year before they would like to transfer. We focus on those applying for seats in 7th and 9th grade (O'Bryant also accepts a handful of 10th graders). Successful 7th grade applicants transfer out of middle school, while 9th grade applicants are picking a high school. Exam school applicants also submit an official GPA report, based on their grades through the most recent fall term. Finally, exam school applicants are asked to rank up to three exam schools. Each exam school running variable is a composite constructed as a weighted average of applicants' standardized math and English GPA, along with standardized scores on the four parts of the ISEE (verbal, quantitative, reading, and math).



Let  $p_{ik}$  denote student  $i$ 's  $k$ th choice and represent  $i$ 's preference list by  $\mathbf{p}_i = (p_{i1}, p_{i2}, p_{i3})$ , where  $p_{ik} \equiv 0$  if the student's rank order list is incomplete. Applicants are ranked only for schools to which they have applied, so applicants with the same GPA and ISEE scores might be ranked differently at different schools depending on where they fall in each school's applicant pool.<sup>9</sup> Let  $c_{ik}$  denote student  $i$ 's school- $k$  specific ranking as determined by his or her composite score (where we adopt the convention that a higher number is better) and write the vector of ranks as  $\mathbf{c}_i = (c_{i1}, c_{i2}, c_{i3})$ , where  $c_{ik}$  is missing if student  $i$  did not rank school  $k$ .

Assignment is determined by the student-proposing DA with student preferences over the three schools, school capacities, and students' (rank-ordered) school-specific composites as parameters. The algorithm works as follows:

ROUND 1: Each student applies to her first choice school. Each school rejects the lowest-ranking students in excess of its capacity, with the rest provisionally admitted (students not rejected at this step may be rejected in later steps).

ROUND  $\ell > 1$ : Students rejected in Round  $\ell - 1$  apply to their next most preferred school (if any). Each school considers these students *and* provisionally admitted students from the previous round, rejecting the lowest-ranking students in excess of capacity, producing a new provisional admit list (again, students not rejected at this step may be rejected in later steps).

The algorithm terminates when either every student is matched to a school or every unmatched student has been rejected by every school he has ranked.

Let  $\tau_k$  denote the rank of the lowest ranked student offered a seat at school  $k$ . We center and scale school-specific composite ranks around this cutoff value using

$$(3.1) \quad r_{ik} = \frac{100}{N_k} \times (c_{ik} - \tau_k),$$

where  $N_k$  is the number of students who ranked school  $k$ . These standardized school-specific ranks equal zero at the cutoff for school  $k$ , with nonnegative values indicating students who ranked and qualified for admission at that school. Absent centering, standardized ranks give applicants percentile position in the distribution of applicants to school  $k$ . A dummy variable,  $q_i(k) = \mathbf{1}\{c_{ik} \geq \tau_k\}$ , indicates that student  $i$  qualified for school  $k$  by clearing  $\tau_k$  (when  $k$  is not ranked by  $i$ ,  $q_i(k)$  is zero).

<sup>9</sup>School-specific running variables arise because schools standardize GPA and ISEE scores among only their applicants, implicitly generating school-specific weights in the composite formula.



Students who ranked and qualified for a school are not offered a seat at that school if they obtain an offer at a more preferred school. With three schools ranked, applicant  $i$  gets an offer at school  $k$  in one of three ways:

- The applicant ranks school  $k$  as his top choice and qualifies:  $(\{p_{i1} = k\} \cap \{q_i(k) = 1\})$ .
- The applicant does not qualify for his top choice, ranks school  $k$  as his second choice, and qualifies there:  $(\{q_i(p_{i1}) = 0\} \cap \{p_{i2} = k\} \cap \{q_i(k) = 1\})$ .
- The applicant does not qualify at his top two choices, ranks school  $k$  as his third choice, and qualifies there:  $(\{q_i(p_{i1}) = q_i(p_{i2}) = 0\} \cap \{p_{i3} = k\} \cap \{q_i(k) = 1\})$ .

To summarize these relationships, let  $O_i$  denote the identity of student  $i$ 's offer, with the convention that  $O_i = 0$  means the student receives no offer.<sup>10</sup> DA then produces the following offer rule:

$$O_i = \sum_{j=1}^J p_{ij} q_i(p_{ij}) \left[ \prod_{\ell=1}^{j-1} (1 - q_i(p_{i\ell})) \right].$$

The sample for whom offers at school  $k$  are sharp in the sense of being deterministically linked with  $k$ 's running variable—a group we refer to as the *sharp sample* for school  $k$ —is the union of three sets of applicants:

- applicants who rank  $k$  first, so  $(p_{i1} = k)$ ,
- applicants who did not qualify for their top choice and rank  $k$  second, so  $(\{q_i(p_{i1}) = 0\} \cap \{p_{i2} = k\})$ ,
- applicants who did not qualify for their top two choices and rank  $k$  third, so  $(\{q_i(p_{i1}) = q_i(p_{i2}) = 0\} \cap \{p_{i3} = k\})$ .

Applicants can be in multiple sharp samples. For example, a student who ranked Boston Latin first, but did not qualify there, is also in the sharp sample for Latin Academy if Latin Academy is her second choice.

An offer dummy,  $Z_{ik}$ , indicates applicants who clear the admissions cutoff at school  $k$ , defined separately for each school and sharp sample. This is the instrumental variable in the fuzzy RD strategy used here. Note that  $Z_{ik} = 0$  for a student who qualifies at  $k$ , but is not in the  $k$  sharp sample. Within sharp samples, the *discontinuity sample* consists of applicants ranked in the interval  $[-20, +20]$ . Applicants outside this “Boston window” are well below or well above the relevant cutoffs. At the same time, the  $[-20, +20]$  window is wide enough to allow for reasonably precise inference.

A possible drawback in the sharp sample estimation strategy arises from the fact that the sharp sample itself may change discontinuously at the cutoff.<sup>11</sup> Suppose, for example, two schools have the same admissions cutoff and use a common running variable to select students. Some students rank school 2

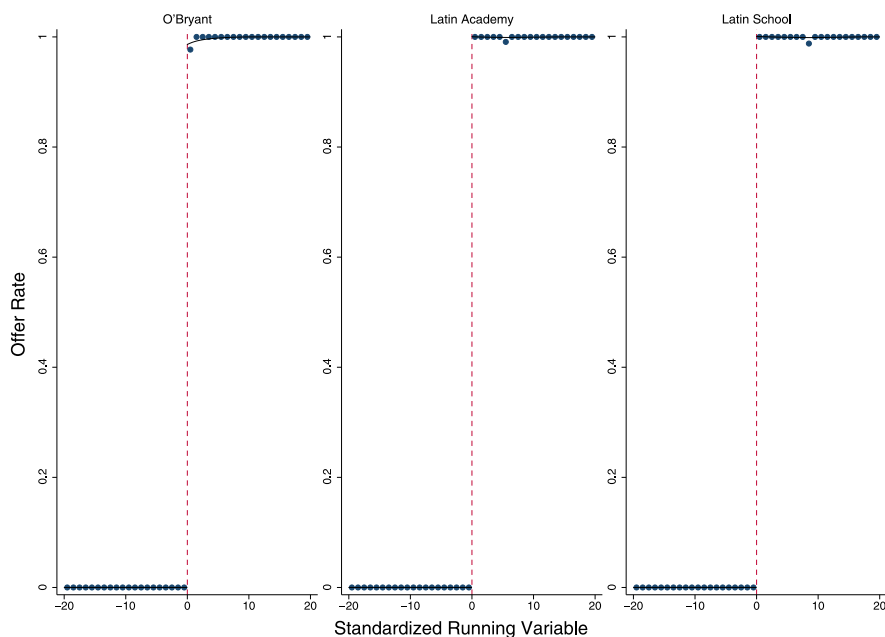
<sup>10</sup>We also adopt the convention that  $\prod_{\ell=1}^0 a_\ell = 1$ .

<sup>11</sup>Our thanks to a referee for pointing this out.

ahead of school 1 and some rank school 1 ahead of school 2. The sharp sample for school 1 includes those who rank 1 first, as well as those who rank 2 first but are disqualified there. This second group appears only to the left of the common cutoff, changing the composition of the sharp sample at the cutoff. In practice, this is unlikely to be a problem because cutoffs are well-separated. Moreover, in Boston, running variables are school-specific. Not surprisingly, therefore, we find no evidence of discontinuities in sharp sample membership at each cutoff in our data. Nevertheless, as insurance against bias of this sort, the estimating equation includes a full set of dummies for application risk sets. That is, estimating equations include dummies for the full interaction of application cohort and applicant preferences. By construction, estimates that condition on applicant preferences are immune to changes in preferences at the cutoff.

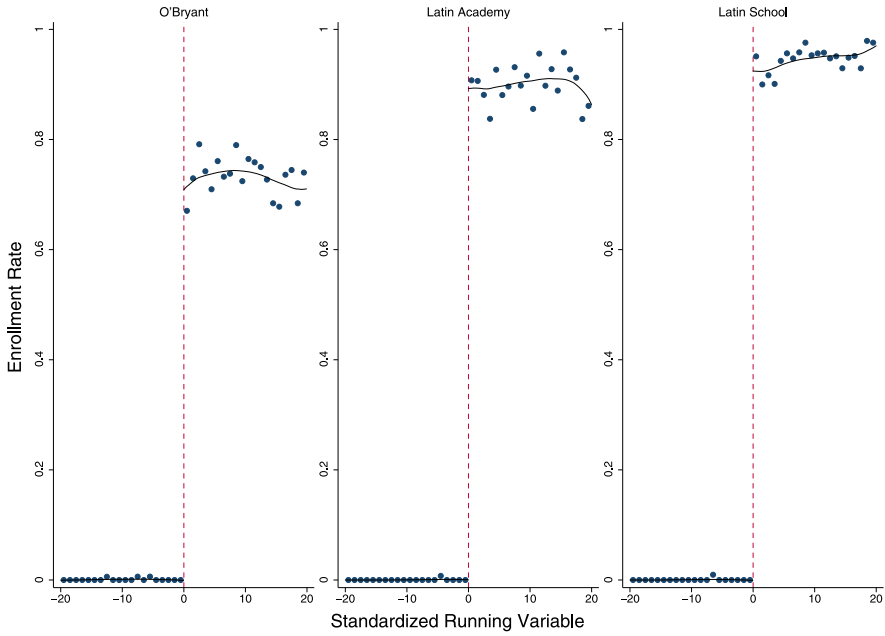
### *Offers, Enrollment, and Schools in Sharp Samples*

Figure 1(a) plots offers as a function of standardized composite ranks in sharp samples, confirming the sharpness of offers in these samples. Plotted points are conditional means for all applicants in a one-unit binwidth, sim-

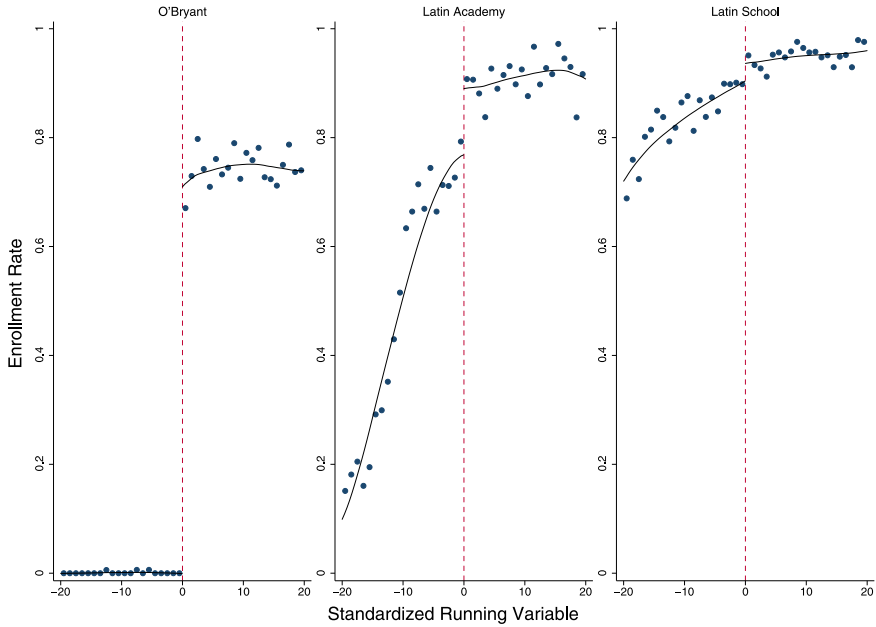


(a) Offers at each Boston exam school

FIGURE 1.—This figure shows offers (a) and enrollment (b) at each Boston exam school, as well as enrollment at any Boston exam school (c), plotted against school-specific standardized running variables.



(b) Enrollment at each Boston exam school



(c) Enrollment at any Boston exam school

FIGURE 1.—Continued.

ilar to the empirical conditional mean functions reported in [Lee, Moretti, and Butler \(2004\)](#). The plots also show estimated conditional mean functions smoothed using local linear regression (LLR). Specifically, for school  $k$ , data in the Boston window were used to construct estimates of  $E[y_i|r_{ik}]$ , where  $y_i$  is the dependent variable and  $r_{ik}$  is the running variable. The LLR smoother uses the edge kernel,

$$K_h(r_{ik}) = \mathbf{1}\left\{\left|\frac{r_{ik}}{h}\right| \leq 1\right\} \cdot \left(1 - \left|\frac{r_{ik}}{h}\right|\right),$$

where  $h$  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 the [DesJardins and McCall \(2008\)](#) bandwidth, studied by [Imbens and Kalyanaraman \(2012\)](#) (IK), who derived optimal bandwidths for sharp RD using a mean squared error loss function with a regularization adjustment (hereafter, DM). This DM smoother (which generates somewhat more stable estimates in our data than the bandwidth IK prefer) is also used to construct nonparametric RD estimates, below.

In sharp samples, offers are determined by the running variable, but exam school enrollment remains probabilistic. Specifically, not all offers are accepted. [Figure 1\(b\)](#) shows that applicants scoring just above admissions cut-offs are much more likely to enroll in a given school than are those just below, but enrollment rates among the offered are below 1. Enrollment rates at other schools also change around each school-specific cutoff. [Figure 1\(c\)](#) puts these pieces together by plotting the probability of enrollment in *any* exam school. Overall 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 of this cutoff are very likely to enroll in either O'Bryant or Latin Academy.

The effect of qualification on enrollment is detailed further in [Table I](#). This table reports LLR estimates of school-specific enrollment rates in the neighborhood of each school's cutoff. Among qualifying 7th grade applicants in the O'Bryant sharp sample, 72% enroll in O'Bryant, while the remaining 28% enroll in a regular BPS school. Ninety-one percent of those qualifying at Latin Academy enroll there the following fall, while 92% qualifying at Latin School enroll there. Many of those not offered seats at one exam school end up in another, mostly the next school down in the hierarchy of school selectivity.

Our fuzzy RD strategy uses exam school offer dummies as instruments for exam school exposure. Specifically, we assume exam school offers affect test scores and other outcomes solely by virtue of changing peer composition. A prerequisite for this change in peer exposure is exam school enrollment. [Table I](#) therefore also describes destination schools in the relevant subpopulation of compliers. Here, compliers are defined as applicants to school  $k$  who enroll there when offered, but go elsewhere otherwise. Complier enrollment outcomes are estimated using the IV strategy described in [Abadie \(2003\)](#), where

TABLE I  
BOSTON AND NEW YORK SCHOOL CHOICES<sup>a</sup>

	All Applicants						Compliers		
	Z = 0 (1)	Z = 1 (2)	Z = 0 (3)	Z = 1 (4)	Z = 0 (5)	Z = 1 (6)	Z = 0 (7)	Z = 0 (8)	Z = 0 (9)
<b>Panel A. Boston 7th Grade Applicants</b>									
	O'Bryant	Latin Academy		Latin School		O'Bryant	Latin Academy	Latin School	
Traditional Boston public schools	1.00	0.28	0.22	0.09	0.08	0.06	1.00	0.15	0.03
O'Bryant	...	0.72	0.77	...	0.06	...	...	0.84	0.05
Latin Academy	...	...	...	0.91	0.86	0.01	...	...	0.93
Latin School	...	...	...	...	...	0.92	...	...	...
<b>Panel B. Boston 9th Grade Applicants</b>									
	O'Bryant	Latin Academy		Latin School		O'Bryant	Latin Academy	Latin School	
Traditional Boston public schools	1.00	0.32	0.27	0.14	0.14	0.04	1.00	0.13	0.10
O'Bryant	...	0.68	0.73	...	0.01	...	...	0.85	...
Latin Academy	...	...	...	0.87	0.86	0.02	...	...	0.91
Latin School	...	...	...	-0.01	...	0.94	...	0.01	...
<b>Panel C. NYC 9th Grade Applicants</b>									
	Brooklyn Tech		Bronx Science		Stuyvesant		Brooklyn Tech	Bronx Science	Stuyvesant
Traditional NYC public schools	0.74	0.36	0.49	0.22	0.15	0.09	0.86	0.72	0.12
Brooklyn Tech	0.10	0.54	0.39	0.30	0.25	0.08	...	0.23	0.32
Bronx Science	0.02	...	0.02	0.39	0.43	0.17	0.04	...	0.50
Stuyvesant	0.03	0.01	...	...	0.08	0.63	0.04	...	...

<sup>a</sup>This table describes the destination schools of exam school applicants in Boston and New York. Columns (1)–(6) show enrollment rates to the left ( $Z = 0$ ) and right ( $Z = 1$ ) of each exam school cutoff. Enrollment rates are measured in the fall following exam school application and estimated using local linear smoothing. Columns (7)–(9) show enrollment destinations when not offered a seat, for enrollment compliers only. Enrollment compliers are applicants who attend the target exam school when offered a seat but not otherwise. Panels A and B report distributions for Boston applicants in 7th and 9th grade. Panel C reports distributions for 9th grade applicants to NYC schools. The Boston 7th grade sample includes students who applied for admission from 1999–2008. The Boston 9th grade sample includes students who applied for admission from 2001–2007. The NYC sample includes students who applied for admission from 2004–2007. Boston calculations are for the sharp sample of applicants who are offered a seat at the target school when they qualify. Entries of ... indicate no enrollment.

a school-specific enrollment dummy is the endogenous variable.<sup>12</sup> Column (7) of Table I shows that the counterfactual for all O’Bryant compliers is regular public school. Among Latin Academy compliers, the counterfactual school is mostly O’Bryant, while among Latin School compliers, the counterfactual school is most often Latin Academy. This serves to highlight the progressive nature of the Boston exam school “experiment”: only among O’Bryant compliers do we get to compare exam school and traditional public schools directly. At the same time, as we show below for New York as well as Boston, movement up the ladder of exam school selectivity is associated with dramatic changes in peer composition.

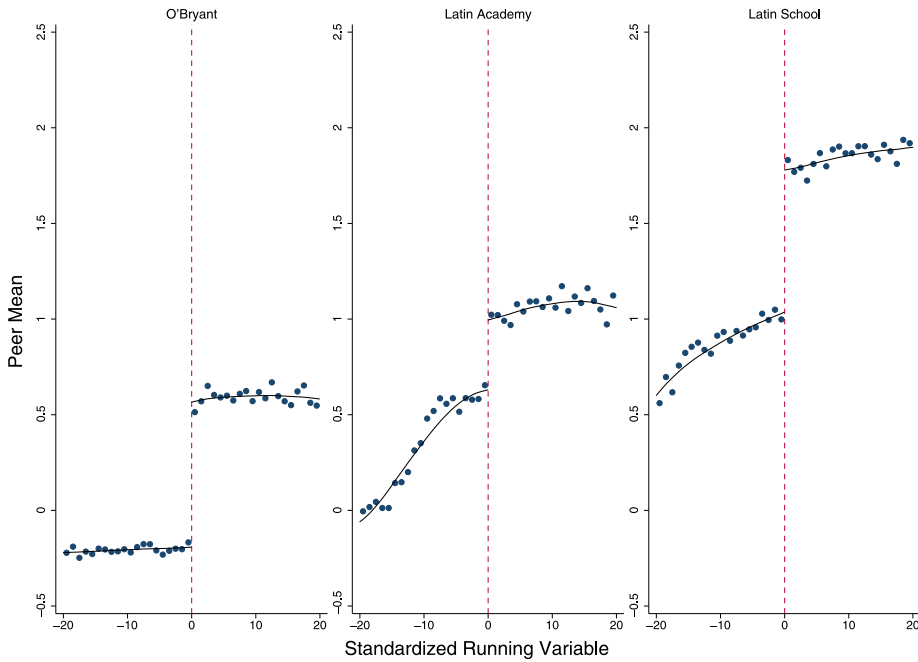
### 3.3. *The Exam School Environment*

The peer achievement first stage that lies behind our fuzzy RD identification strategy is described in Figure 2(a). This figure plots peer mean math scores for 7th and 9th grade applicants in the sharp sample who are on either side of admissions cutoffs. Peer means are defined as the average baseline score of same-grade schoolmates in the year following exam school application. Baseline peer means jump by roughly half a standard deviation at each admissions cutoff. The jump in peer mean English scores (not shown) is similar to that for math.

The proportion nonwhite among exam school students has often been a lightning rod for controversy. Beginning in the 1970s, Boston’s court-mandated desegregation plan maintained the proportion black and Hispanic in exam schools at roughly 35%. Racial preferences were challenged in 1996, however, and Boston exam school admissions have ignored race since 1999. In our sample, drawn from years after racial preferences were abandoned, the proportion of black and Hispanic peers drops sharply at exam school cutoffs, a fact documented in Figure 2(b). The proportion nonwhite falls by about 10 percentage points at the O’Bryant cutoff, with even larger drops at the Latin Academy and Latin School cutoffs.

Additional features of the exam school environment are summarized in Table II, focusing on enrollment compliers as in columns (7)–(9) of Table I. Table II documents the marked shifts in peer achievement and racial composition captured graphically in Figure 2. Other contrasts between the exam school environment and regular public schools are less systematic. Class sizes for middle school applicants tend to be larger at exam schools, but differences in size shrink in grade 9 and change little at the Latin School cutoff. Exam school

<sup>12</sup>Specifically, compliers are defined as follows. Let  $D_{1i}$  denote exam school enrollment status when the instrument  $Z_i$  is switched on and  $D_{0i}$  denote exam school enrollment status when the instrument  $Z_i$  is switched off. Compliers have  $D_{1i} = 1$  and  $D_{0i} = 0$ . Although the compliant population cannot be enumerated, characteristics of this population are nonparametrically identified and easily estimated.



(a) Baseline peer math score at Boston exam schools for 7th and 9th grade applicants

FIGURE 2.—This figure shows average baseline peer math scores (a) and proportion of peers that are black or Hispanic (b) among 7th and 9th grade applicants to Boston exam schools, plotted against school-specific standardized running variables.

teachers tend to be older than regular public school teachers, as can be seen at the O’Bryant cutoff, but teacher age changes little at the Latin Academy and Latin School cutoffs.

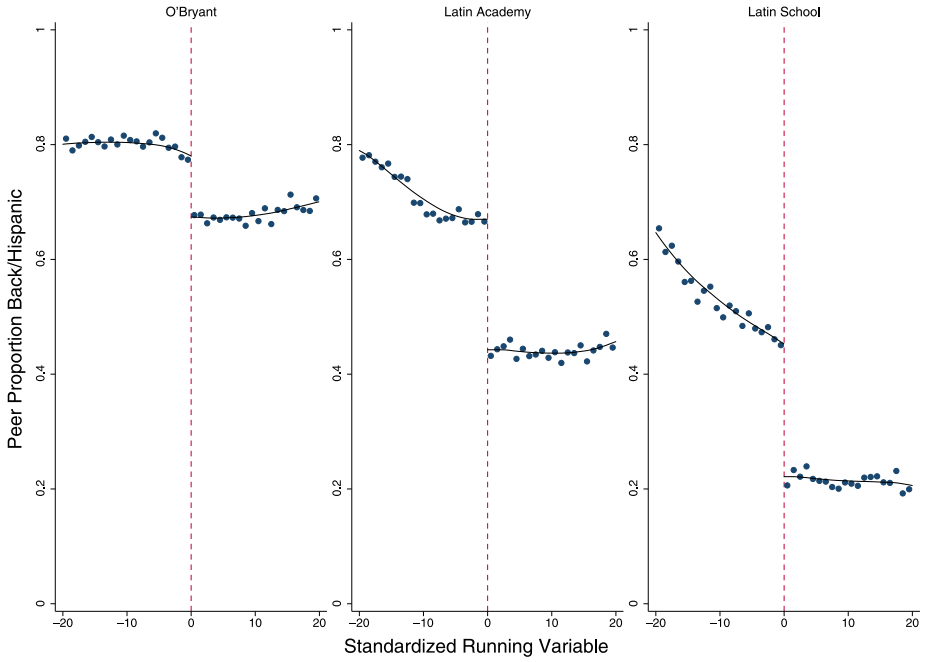
The large and systematic changes in peer composition at each cutoff and entry grade motivate our focus on peers as the primary mediator of the exam school treatment. Before turning to a 2SLS analysis that treats peer composition as the primary causal channel for exam school effects, however, we begin with reduced-form estimates.

#### 4. REDUCED-FORM ACHIEVEMENT EFFECTS

##### 4.1. *Boston Estimates*

We constructed parametric and nonparametric RD estimates of the effect of an exam school offer using the standardized composite rank (3.1) as the running variable. We refer to this initial set of estimates as “reduced form” because these estimates capture the overall effect of an exam school offer, without adjusting for the relationship between offers and mediating variables. As noted in





(b) Proportion black or Hispanic at Boston exam schools

FIGURE 2.—Continued.

the recent survey by Lee and Lemieux (2010), parametric and nonparametric RD estimates are complementary, providing a mutually reinforcing specification check.

The parametric estimating equation for applicants in the sharp sample at school  $k$  is

$$(4.1) \quad y_{itk} = \alpha_{tk} + \sum_j \delta_{jk} d_{ij} + (1 - Z_{ik}) f_{0k}(r_{ik}) + Z_{ik} f_{1k}(r_{ik}) + \rho_k Z_{ik} + \eta_{itk},$$

where  $y_{itk}$  is an outcome variable for student  $i$ , observed in year  $t$ , who applied to school  $k$ ;  $Z_{ik}$  indicates an offer at school  $k$ , and the coefficient of interest is  $\rho_k$ . Equation (4.1) controls for test year effects at school  $k$ , denoted  $\alpha_{tk}$ , and for the full interaction of application cohort and applicant preferences, indicated by dummies,  $d_{ij}$ . (These are included for consistency with some of the over-identified 2SLS models discussed below.)<sup>13</sup> The effects of the running variable

<sup>13</sup>The over-identified 2SLS models discussed in Section 5 use interactions between exam offers and applicant cohort dummies as additional instruments.

TABLE II  
BOSTON SCHOOL CHARACTERISTICS FOR COMPLIERS<sup>a</sup>

	O'Bryant		Latin Academy		Latin School	
	Z = 0 (1)	Z = 1 (2)	Z = 0 (3)	Z = 1 (4)	Z = 0 (5)	Z = 1 (6)
Panel A. 7th Grade Applicants						
Baseline peer mean in math	-0.15	0.84	0.68	1.20	1.15	1.97
Baseline peer mean in English	-0.15	0.80	0.65	1.11	1.06	1.78
Proportion black or Hispanic	0.78	0.63	0.65	0.40	0.43	0.18
Proportion qualifying for a free lunch	0.78	0.66	0.66	0.46	0.48	0.28
Proportion female	0.46	0.57	0.56	0.56	0.56	0.55
Student/teacher ratio	12.1	19.7	19.7	21.2	21.5	22.0
Proportion of teachers						
licensed to teach assignment	0.90	0.97	0.96	0.95	0.96	0.96
Proportion of teachers						
highly qualified in core subject	0.91	0.93	0.94	0.95	0.95	0.95
Proportion of teachers 40 and older	0.39	0.63	0.66	0.52	0.56	0.53
Proportion of teachers 48 and older	0.26	0.51	0.54	0.38	0.42	0.41
Proportion of teachers 56 and older	0.10	0.27	0.29	0.19	0.21	0.21
Panel B. 9th Grade Applicants						
Baseline peer mean in math	-0.32	0.87	0.75	1.02	0.94	1.76
Baseline peer mean in English	-0.22	0.72	0.61	0.99	0.89	1.40
Proportion black or Hispanic	0.82	0.67	0.69	0.42	0.46	0.18
Proportion free lunch	0.67	0.59	0.58	0.45	0.47	0.26
Proportion female	0.48	0.58	0.58	0.57	0.58	0.55
Student/teacher ratio	16.5	19.8	18.2	21.2	21.2	22.1
Proportion of teachers						
licensed to teach assignment	0.88	0.98	0.97	0.96	0.94	0.96
Proportion of teachers						
highly qualified in core subject	0.86	0.94	0.91	0.95	0.94	0.95
Proportion of teachers 40 and older	0.26	0.65	0.66	0.54	0.52	0.54
Proportion of teachers 48 and older	0.19	0.52	0.53	0.40	0.40	0.42
Proportion of teachers 56 and older	0.06	0.27	0.26	0.18	0.21	0.21

<sup>a</sup>This table shows descriptive statistics for Boston enrollment compliers to the left ( $Z = 0$ ) and right ( $Z = 1$ ) of Boston exam school cutoffs. Student-weighted average characteristics of teachers and schools were constructed from data posted at [http://profiles.doe.mass.edu/state\\_report/teacherdata.aspx](http://profiles.doe.mass.edu/state_report/teacherdata.aspx). *Teachers licensed to teach assignment* gives the percent of teachers at the school attended who are licensed with Provisional, Initial, or Professional licensure to teach in the subject(s) in which they are posted. *Proportion of teachers highly qualified in core subject* gives the percent of teachers of core subjects (ELA, Math, and science, among others) at the school attended that were taught by teachers holding a Massachusetts teaching license and demonstrating subject matter competence in the areas they teach. Teacher data are for Fall 2003–2008, except information on core academic teachers, which is for Fall 2003–2006, and teacher age, which is for Fall 2007–2008. For middle school applicants, peer baseline means are enrollment-weighted scores on 4th grade MCAS for Fall 2000–2008. Peer baseline for 9th grade applicants comes from 7th and 8th grade MCAS tests taken for Fall 2002–2008.

at school  $k$  are controlled by a pair of third-order polynomials that differ on either side of the cutoff, specifically

$$(4.2) \quad f_{jk}(r_{ik}) = \pi_{jk}r_{ik} + \xi_{jk}r_{ik}^2 + \psi_{jk}r_{ik}^3; \quad j = 0, 1.$$

Nonparametric estimates differ from parametric in three ways. First, they narrow the Boston window when the optimal data-driven bandwidth falls below 20.<sup>14</sup> Second, our nonparametric estimates use a tent-shaped edge kernel centered at admissions cutoffs instead of the uniform kernel implicit in parametric estimation. Finally, nonparametric models control for linear functions of the running variable only, omitting higher-order terms. We can write the nonparametric estimating equation as

$$(4.3) \quad \begin{aligned} y_{itk} &= \alpha_{tk} + \sum_j \delta_{jk}d_{ij} + \gamma_{0k}(1 - Z_{ik})r_{ik} + \gamma_{1k}Z_{ik}r_{ik} + \rho_k Z_{ik} + \eta_{itk} \\ &= \alpha_{tk} + \sum_j \delta_{jk}d_{ij} + \gamma_{0k}r_{ik} + \gamma_k^*Z_{ik}r_{ik} + \rho_k Z_{ik} + \eta_{itk} \end{aligned}$$

for each of the three schools indexed by  $k$ . Nonparametric RD estimates come from a kernel-weighted least squares fit of equation (4.3).

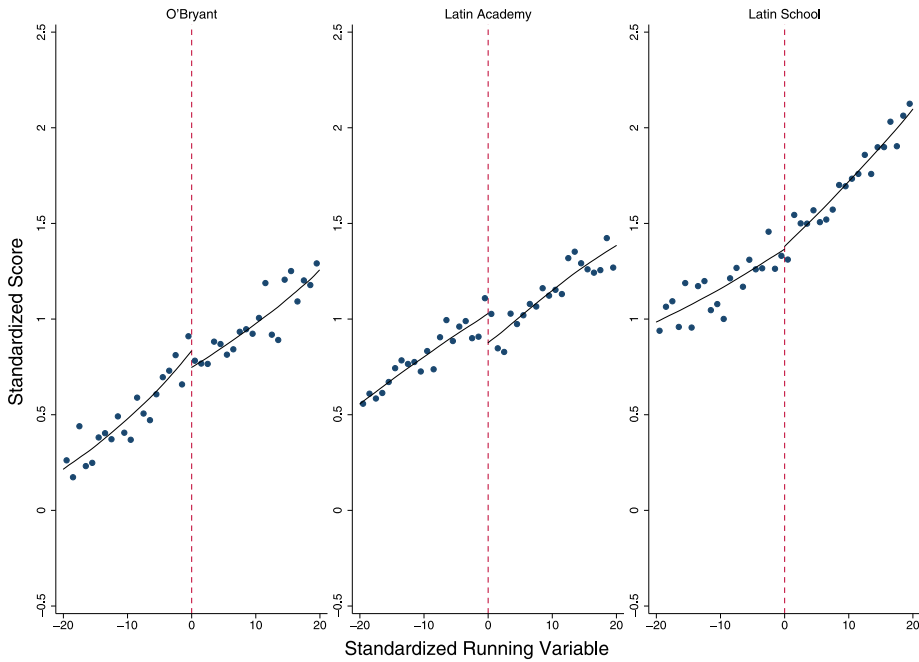
### MCAS Scores

The plots for 10th grade English show jumps at two out of three cutoffs, but other visual reduced forms offer little evidence of marked discontinuities in MCAS scores. This can be seen in Figures 3(a) and 3(b) for middle school and Figures 4(a) and 4(b) for high school. Jumps in smoothed scores at admissions cutoffs constitute nonparametric estimates of the effect of an exam school offer in the sharp sample. The corresponding estimates, reported in Table III, tell the same story. Few of the estimates are significantly different from zero, while some of the significant effects at Latin School are negative (e.g., 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.

In an effort to increase precision, we constructed estimates pooling applicants to all three Boston exam schools. The pooled estimating equations parallel equations (4.1) and (4.3), but with a single offer effect,  $\rho$ . These specifications interact all control variables, including running variables, with application-school dummies.<sup>15</sup> The kernel weight for the stack becomes  $K_{h_k}(r_{ik})$ , where school  $k$ 's bandwidth  $h_k$  is estimated separately in a preliminary step. Because the pooled model includes a full set of main effects and interactions for school-specific subsamples, we can think of the estimate of  $\rho$

<sup>14</sup>The DM bandwidths for Table III range from about 10 to 37.

<sup>15</sup>In the stacked analysis, an observation from the sharp sample for school  $k$  is associated with the running variable for that school. Other running variables are switched off by virtue of the interaction terms included in the stacked model.



(a) 7th and 8th grade math at Boston exam schools for 7th grade applicants

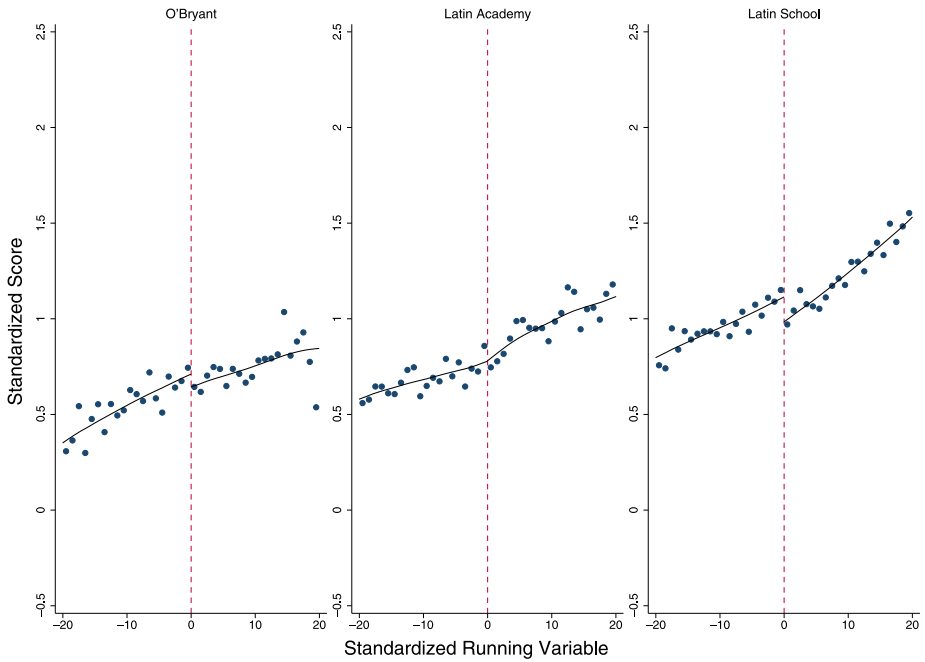
FIGURE 3.—This figure shows the average 7th and 8th grade math (a) and English (b) MCAS scores of 7th grade applicants to Boston exam schools, plotted against a school-specific standardized running variables.

in this stack as a variance-of-treatment-weighted average of school-specific estimates.<sup>16</sup> Note that some students appear in more than one sharp sample; each student contributes up to three observations for each outcome. Our inference framework takes account of this by clustering standard errors by student.

Paralleling the pattern shown in the Boston reduced-form figures, offer effects from the stacked models, reported in columns labeled “All Schools” in Table III, are mostly small, with few significantly different from zero. The large significant estimate for 10th grade English scores, a result generated by both parametric and nonparametric models, is partly offset by marginally significant negative effects on 7th and 8th grade English.<sup>17</sup> When all scores are pooled, the overall estimate is close to zero (models combining years and grades are stacked in much the same way that models combining schools are stacked). Importantly, the combination of school- and score-pooling generates precise

<sup>16</sup>Variance-weighting is a property of regression models with saturated controls; see, for example, Angrist (1998).

<sup>17</sup>Table A.IV in Appendix A reports high school results separately for 7th and 9th grade applicants. This table shows positive English effects for both application grades.



(b) 7th and 8th grade English at Boston exam schools for 7th grade applicants

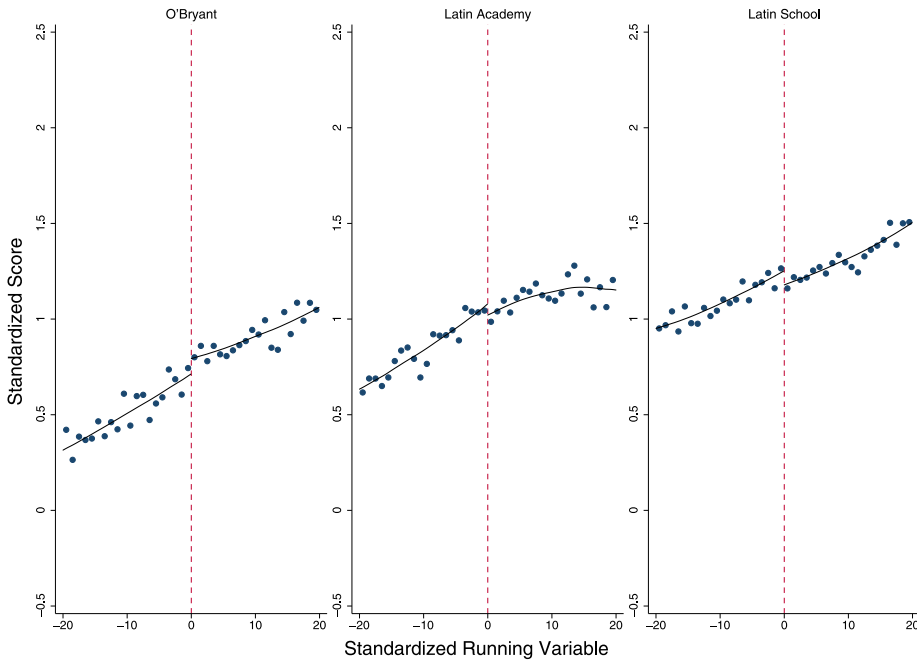
FIGURE 3.—Continued.

estimates, with standard errors on the order of 0.02 for both math and English.

Estimates for black and Hispanic applicants, reported in Table A.V in Appendix A, are in line with the full-sample findings for math and middle school English scores. Also, consistent with the full-sample results for 10th grade English, an exam school education seems especially likely to boost 10th grade English scores for blacks and Hispanics, with an estimated effect of  $0.16\sigma$ , but there are some significant negative estimates as well.

The Appendix reports results from an exploration of possible threats to a causal interpretation of the reduced-form estimates in Table III. Specifically, we look for differential attrition (i.e., missing score data) to the right and left of exam school cutoffs (in Table A.II) and for discontinuities in covariates (in Table A.III). Receipt of an exam school offer makes attrition somewhat less likely, but the gaps here are small and unlikely to impart substantial selection bias in estimates that ignore them.<sup>18</sup> A handful of covariate contrasts also pop up as significantly different from zero, but the spotty nature of these estimates

<sup>18</sup>Lee (2009) bounds on the extent of selection bias confirm this. Also worth noting is the fact that F-tests for the joint significance of differential attrition in all MCAS reduced forms generate  $p$ -values of about 0.2 or higher.



(a) 10th grade math at Boston exam schools for 7th and 9th grade applicants

FIGURE 4.—This figure shows the average 10th grade math (a) and English (b) MCAS scores of 7th and 9th grade applicants to Boston exam schools, plotted against school-specific standardized running variables.

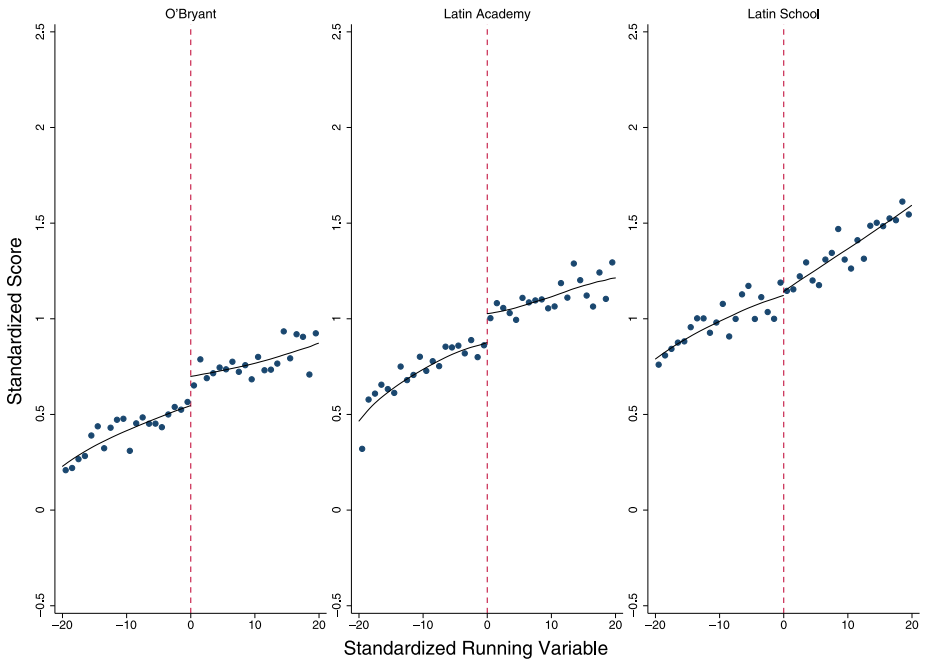
seem consistent with the notion that comparisons to the left and right of exam school admissions cutoffs are indeed a good experiment.

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 less likely to switch schools than are traditional BPS students. Increased persistence in school is especially marked among 7th grade applicants, though this latter increase is due in part to the fact that few exam school students switch schools in grade 9, when most other BPS students transition to a new high school.<sup>19</sup>

### *High Achievers*

To provide additional evidence on effects across quantiles of the applicant ability distribution, we exploit the fact any single test is necessarily a noisy mea-

<sup>19</sup>These estimates come from a nonparametric reduced-form analysis similar to that used to construct the covariate balance and attrition estimates in the [Appendix](#).



(b) 10th grade English at Boston exam schools for 7th and 9th grade applicants

FIGURE 4.—Continued.

sure of ability. Although we cannot construct (nonparametric) 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. This approach operationalizes a suggestion in Lee and Lemieux’s (2010) recent survey of RD methods, which points out that a test score running variable can be seen as a noisy measure of an underlying ability control. Here, we exploit the fact that some in the high-baseline group are ultra-high achievers who earned marginal ISEE scores by chance.

The average baseline score for exam school applicants in the upper half of the baseline MCAS distribution hovers around  $1.2\text{--}1.5\sigma$  in both math and English. Table IV shows that this is close to the average baseline achievement level among students enrolled in exam schools. Importantly, MCAS scores remain informative even for these high achievers: no more than one third top out in the sense of testing at the Advanced (highest) MCAS proficiency level. Likewise, MCAS scores remain informative even for applicants in the upper baseline MCAS quartile, where average baseline scores are  $0.5\text{--}0.6\sigma$  beyond those of the average among exam school 7th graders. Note also that applicants in these high-achieving groups are exposed to almost exactly the same changes in peer composition as applicants in the full sample.



TABLE III  
 BOSTON REDUCED-FORM ESTIMATES: MCAS MATH AND ENGLISH<sup>a</sup>

Application Grade	Test Grade	Parametric Estimates				Nonparametric (DM) Estimates			
		O'Bryant (1)	Latin Academy (2)	Latin School (3)	All Schools (4)	O'Bryant (5)	Latin Academy (6)	Latin School (7)	All Schools (8)
Panel A. Math									
7th	7th and 8th	-0.125 (0.100) 4,047	-0.105 (0.093) 4,208	0.002 (0.099) 3,786	-0.079 (0.054) 12,041	-0.093 (0.071) 3,637	-0.144* (0.074) 4,000	0.012 (0.060) 3,067	-0.086** (0.034) 10,704
7th and 9th	10th	0.066 (0.066) 3,389	-0.097 (0.085) 2,709	-0.056 (0.051) 2,459	-0.018 (0.036) 8,557	0.067 (0.045) 3,083	-0.047 (0.047) 2,027	-0.064** (0.028) 1,827	0.000 (0.026) 6,937
7th and 9th	7th, 8th, and 10th	-0.038 (0.068) 7,436	-0.102 (0.067) 6,917	-0.020 (0.072) 6,245	-0.054 (0.039) 20,598	-0.020 (0.049) 6,720	-0.115** (0.049) 6,027	-0.016 (0.043) 4,894	-0.053** (0.024) 17,641
Panel B. English									
7th	7th and 8th	-0.061 (0.078) 4,151	-0.092 (0.067) 4,316	-0.187*** (0.065) 3,800	-0.110** (0.043) 12,267	-0.062 (0.041) 3,931	0.012 (0.042) 3,762	-0.128*** (0.037) 3,533	-0.063** (0.025) 11,226
7th and 9th	10th	0.108 (0.079) 3,398	0.136 (0.096) 2,715	0.028 (0.085) 2,463	0.095* (0.053) 8,576	0.140*** (0.048) 3,308	0.182*** (0.057) 1,786	-0.002 (0.065) 1,916	0.113*** (0.036) 7,010
7th and 9th	7th, 8th, and 10th	0.014 (0.055) 7,549	-0.001 (0.070) 7,031	-0.106* (0.061) 6,263	-0.026 (0.039) 20,843	0.029 (0.034) 7,239	0.067 (0.042) 5,548	-0.089*** (0.032) 5,449	0.002 (0.023) 18,236

<sup>a</sup>This table reports estimates of the effects of exam school offers on MCAS scores. The sample covers students within 20 standardized units of offer cutoffs. Parametric models include a cubic function of the running variable, allowed to differ on either side of offer cutoffs. Nonparametric estimates use the edge kernel, with bandwidth computed following DesJardins and McCall (2008) and Imbens and Kalyanaram (2012), as described in the text. Optimal bandwidths were computed separately for each school. Robust standard errors, clustered on year and school, are shown in parentheses. Standard errors for the all-schools estimates and for estimates pooling outcomes also cluster on student. Sample sizes are shown below standard errors. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

TABLE IV  
BOSTON REDUCED-FORM ESTIMATES FOR HIGH ACHIEVERS<sup>a</sup>

Application Grade	Test Grade	Baseline Mean for Enrolled (1)	Conditional on Baseline Score			
			Baseline in Upper Half		Baseline in Upper Quartile	
			Baseline Mean (2)	Estimate (3)	Baseline Mean (4)	Estimate (5)
Panel A. Math						
7th	7th and 8th	1.43	1.53	-0.097** (0.042) 5,986	2.13	-0.016 (0.063) 2,906
7th and 9th	10th	1.33	1.38	-0.006 (0.024) 4,196	1.85	0.014 (0.030) 2,459
7th and 9th	7th, 8th, and 10th	1.35	1.461	-0.062** (0.028) 10,182	1.99	-0.002 (0.038) 5,365
Panel B. English						
7th	7th and 8th	1.31	1.42	-0.074** (0.029) 6,926	1.85	-0.091** (0.042) 3,038
7th and 9th	10th	1.21	1.25	0.048 (0.039) 4,399	1.58	0.094* (0.051) 2,468
7th and 9th	7th, 8th, and 10th	1.23	1.35	-0.028 (0.028) 11,325	1.72	-0.009 (0.038) 5,506

<sup>a</sup>This table reports nonparametric reduced-form estimates of the all-schools model for Boston exam school applicants with high baseline MCAS scores. The baseline mean reported in column (1) is the average baseline MCAS score for enrolled applicants from study cohorts. Conditional-on-baseline estimates are nonparametric estimates in upper-half and upper-quartile subsamples, with bandwidth computed as for the all-schools results reported in Table III. Robust standard errors, clustered on year and school, are shown in parentheses. Standard errors also cluster on student. Sample sizes are shown below standard errors. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

Perhaps surprisingly, nonparametric RD estimates of the All Schools model for applicants in the upper half and upper quartile of the baseline score distribution come out similar to those for the full sample. These results, reported in columns (3) and (5) of Table IV, are mostly negative, with few significantly different from zero. The exception again is a significant positive effect for 10th grade English. At the same time, the sample of high achievers generates a significant negative estimate of effects on middle school English—an effect of roughly the same magnitude as the positive estimate for 10th graders. Thus, even in a sample of ultra-high (baseline) achievers, there is little evidence of a consistent exam school boost.

### *PSAT, SAT, and AP Exam Scores*

With the exception of the 10th grade test that serves as an exit exam, MCAS scores are only indirectly linked to educational attainment. We therefore look at other indicators of human capital and learning. The first of these is the PSAT, which serves as a warmup for the SAT and is used in the National Merit Scholarship program; the second is the SAT.<sup>20</sup>

SAT and PSAT tests are usually taken toward the end of high school, so scores are unavailable for the youngest applicant cohorts in our sample (Table C.II in the Supplemental Material (Abdulkadiroğlu, Angrist and Pathak (2014)) lists the cohorts contributing each outcome). 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 for both SAT and PSAT tests. The average PSAT score for exam school applicants in the Boston window 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. As with MCAS outcomes, PSAT and SAT scores are standardized to have mean zero and unit variance among all test-takers in our data in a given year.

About 70–80 percent of exam school applicants take the PSAT. O’Bryant offers are estimated to increase PSAT taking by about 6 points, but the corresponding estimate from a model that pools applicants to different schools is small and not significantly different from zero. These results can be seen in Panel A of Table V. Panel B of this table shows that exam school offers have little effect on the likelihood that applicants take the SAT. Selection bias in the sample of test-takers therefore seems unlikely to be a concern. Consistent with the MCAS results, exam school offers generate little apparent gain in either PSAT or SAT scores for test-takers near admissions cutoffs.

Motivated by the prevalence of AP courses in the Boston exam school curriculum, we estimated exam school offer effects on AP participation rates and

<sup>20</sup>The correlation between 10th grade MCAS math scores and PSAT and SAT math scores is about 0.7; the correlation for English is similar. These estimates come from models that control for application cohort and grade, test year, and demographics (race, gender, free lunch).

TABLE V  
BOSTON REDUCED-FORM ESTIMATES: PSAT, SAT, AND AP SCORES<sup>a</sup>

Dependent Variable	O'Bryant (1)	Latin Academy (2)	Latin School (3)	All Schools (4)	O'Bryant (5)	Latin Academy (6)	Latin School (7)	All Schools (8)
	Probability Tested				Test Score for Takers			
PSAT	0.062** (0.025) 2,728	0.017 (0.023) 2,058	-0.071** (0.034) 1,790	0.012 (0.015) 6,576	0.038 (0.043) 2,670	-0.016 (0.053) 1,433	0.006 (0.054) 1,351	0.017 (0.031) 5,454
SAT	-0.024 (0.031) 2,518	0.036 (0.030) 1,731	0.018 (0.028) 1,540	0.003 (0.019) 5,789	0.043 (0.038) 2,364	-0.006 (0.046) 1,429	0.074 (0.063) 1,516	0.040 (0.031) 5,309
	Number of Exams				Sum of Scores			
AP—All Exams	0.075 (0.106) 2,654	-0.172 (0.230) 1,735	0.067 (0.178) 1,827	0.008 (0.068) 6,216	0.761*** (0.238) 2,651	-0.573 (0.653) 1,641	0.146 (0.614) 1,404	0.271 (0.203) 5,696
AP—Exams With 500+ Takers	-0.032 (0.096) 2,681	-0.204 (0.202) 1,719	-0.228* (0.130) 1,993	-0.132** (0.063) 6,393	0.405** (0.201) 2,628	-0.693 (0.563) 1,624	-0.349 (0.478) 1,535	-0.067 (0.170) 5,787

<sup>a</sup>This table reports nonparametric RD estimates of effects of exam school offers on PSAT, SAT, and AP test taking and scores for pooled 7th and 9th grade applicant samples. Panel D results are for AP tests with 500 or more takers (Calculus AB/BC, Statistics, Biology, Chemistry, Physics B/C, English Language and Composition, English Literature and Composition, European History, U.S. Government and Politics, U.S. History, Microeconomics, and Macroeconomics). Outcome-specific nonparametric estimates, bandwidths, and standard errors were computed as for Table III. Robust standard errors, clustered on year and school, are shown in parentheses. Standard errors also cluster on student when schools are stacked. Sample sizes are shown below standard errors. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

scores in a pooled sample of 7th and 9th grade applicants for whom the AP is relevant. As with the PSAT/SAT analysis, younger cohorts are excluded since these tests are usually taken in grades 11–12 (again, Table C.II in the Supplemental Material gives details). AP tests are scored on a scale of 1–5, with some colleges granting credit for subjects in which an applicant scores at least 3 or 4. At the high end of the distribution of the number of tests taken, Latin School students take an average of three to four AP exams.

Table V reports nonparametric RD estimates of AP effects on scores summed over all AP exams, as well as on the sum of scores 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 widely assessed subjects like math, science, English, history, and economics, but omits music and art.<sup>21</sup> Exam school offers fail to increase the number of tests taken, though the sum of scores goes up at O’Bryant. The sum of scores on the most commonly taken, and probably the most substantively important, tests increases as a result of an O’Bryant offer, but is unaffected by offers from Latin School and Latin Academy.

### *Post-Secondary Outcomes*

BPS matches data on high school seniors to National Student Clearinghouse (NSC) files, which record information on enrollment at over 90 percent of American 4-year colleges and universities. We used the BPS-NSC match to look at college attendance, excluding post-secondary institutions that focus on technical and vocational training. The sample here includes 7th and 9th grade exam school applicant cohorts for whom college outcomes are relevant. (For details, see Tables C.II and C.V in the Supplemental Material.) Most Boston exam school applicants go to college: roughly 60 percent to the left of the O’Bryant cutoff, and 90 percent to the right of the Latin School cutoff. At the same time, Table VI, which reports estimated effects on post-secondary outcomes, shows little evidence of a positive exam school treatment effect on college enrollment or quality.<sup>22</sup> Despite the positive grade 10 English and AP score results for some O’Bryant applicants, applicants who clear the O’Bryant cutoff appear to be less likely to attend a competitive college or highly competitive college than they otherwise would have been, though only the nonparametric estimates here are significant.

<sup>21</sup>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, U.S. Government and Politics, U.S. History, Microeconomics, and Macroeconomics.

<sup>22</sup>Selectivity is defined by Barron’s. Boston University and Northeastern University are examples of “Highly Competitive” schools. The University of Massachusetts—Boston and Emmanuel College are “Competitive.”

TABLE VI  
BOSTON REDUCED-FORM ESTIMATES: POST-SECONDARY OUTCOMES<sup>a</sup>

Dependent Variable	Parametric Estimates				Nonparametric (DM) Estimates			
	O'Bryant (1)	Latin Academy (2)	Latin School (3)	All Schools (4)	O'Bryant (5)	Latin Academy (6)	Latin School (7)	All Schools (8)
Attended Any College	-0.039 (0.052) 2,793	0.047 (0.066) 2,217	0.045 (0.051) 2,039	0.010 (0.032) 7,049	0.007 (0.031) 2,721	0.048 (0.031) 1,926	0.062* (0.035) 1,277	0.031 (0.019) 5,924
Attended 4 Year College	-0.053 (0.070) 2,793	-0.007 (0.081) 2,217	0.102 (0.069) 2,039	0.003 (0.041) 7,049	-0.056 (0.044) 2,769	0.060* (0.035) 2,131	0.101** (0.040) 1,401	0.013 (0.026) 6,301
Attended Competitive College	-0.082 (0.078) 2,793	-0.011 (0.087) 2,217	0.104 (0.089) 2,039	-0.011 (0.051) 7,049	-0.100** (0.045) 2,631	0.052 (0.044) 2,217	0.101** (0.049) 1,302	-0.004 (0.029) 6,150
Attended Highly Competitive College	-0.062 (0.049) 2,793	0.028 (0.054) 2,217	0.036 (0.080) 2,039	-0.009 (0.032) 7,049	-0.062*** (0.022) 2,770	0.048 (0.032) 2,047	-0.002 (0.044) 1,450	-0.014 (0.017) 6,267

<sup>a</sup>This table reports nonparametric RD estimates of the effects of exam school offers on college enrollment using data from the National Student Clearinghouse. College selectivity is as classified by Barron's. Each panel shows estimates for pooled 7th and 9th grade applicants. Outcome-specific nonparametric estimates, bandwidths, and standard errors were computed as for Table III. Robust standard errors, clustered on year and school, are shown in parentheses. Standard errors also cluster on student when schools are stacked. Sample sizes are shown below standard errors. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

#### 4.2. *New York Estimates*

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

The New York exam school admissions process is simpler than the Boston process because selection is based solely 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 in 9th grade take the SHSAT and submit an application listing school preferences. Students are ordered by SHSAT 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. There is no corresponding sharp sample for New York exam school applicants, because New York applicants rank many schools, both exams and others, and we have no information on applicant preferences beyond the fact of an exam school application.<sup>23</sup>

Stuyvesant is the most competitive New York City exam school, so the minimum score needed to obtain an offer there exceeds the minimum required at Bronx Science and Brooklyn Technical. As in equation (3.1), school-specific standardized running variables equal zero at each cutoff, with positive values indicating applicants offered a seat. Also as in Boston, applicants might qualify for placement at one school, but rank a less competitive school first and get an offer at that school instead. New York admissions cutoffs are typically separated by six standardized rank units, so the estimation window for each of the New York schools is set at  $[-6, +6]$ . The New York window is narrower than the Boston window of  $+/-20$  but still includes many more applicants.

Figure 5(a) shows how New York offers are related to the running variable. Here, the dots indicate averages in half-unit bins, while the smoothed line was constructed using LLR with the DM bandwidth generated by the estimation sample. Own-school offers jump at each cutoff. Unlike in Boston, however, offer rates among qualified applicants are less than 1 because the sample here is not sharp; that is, some New York applicants who qualify at the target school in each panel have ranked another school at which they qualify higher. Five or six points to the right of the Brooklyn Tech and Bronx Science cutoffs, offers at the next most selective exam school replace those at the target school.

<sup>23</sup>The NYC exam school assignment mechanism is a serial dictatorship with students ordered by SHSAT score. Students apply for exam schools at the same time that they rank regular New York City high schools, and may receive offers from both. Abdulkadiroğlu, Pathak, and Roth (2009) described how exam school admissions interact with admissions at regular high schools in



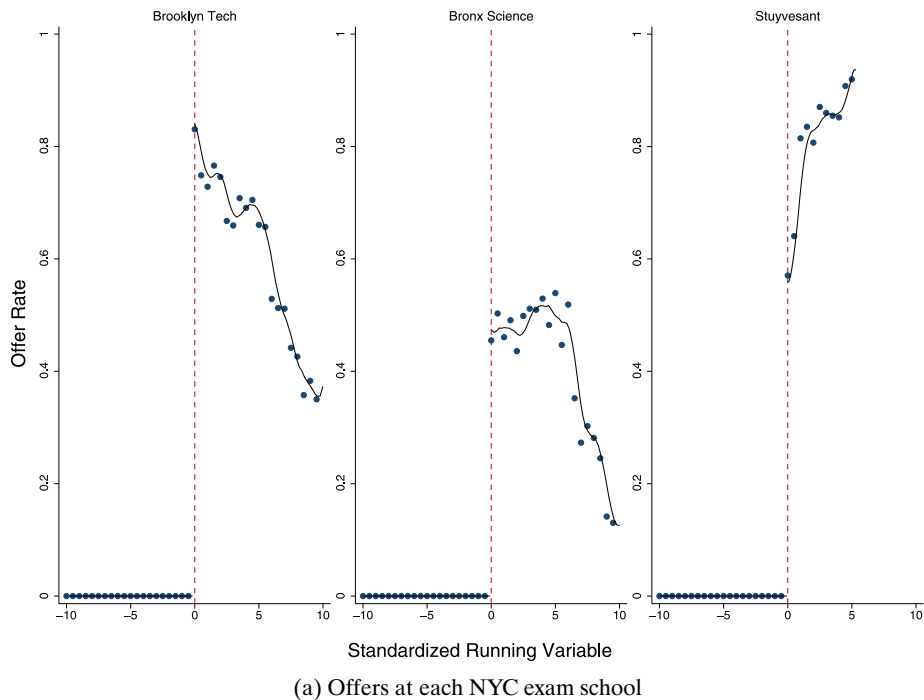


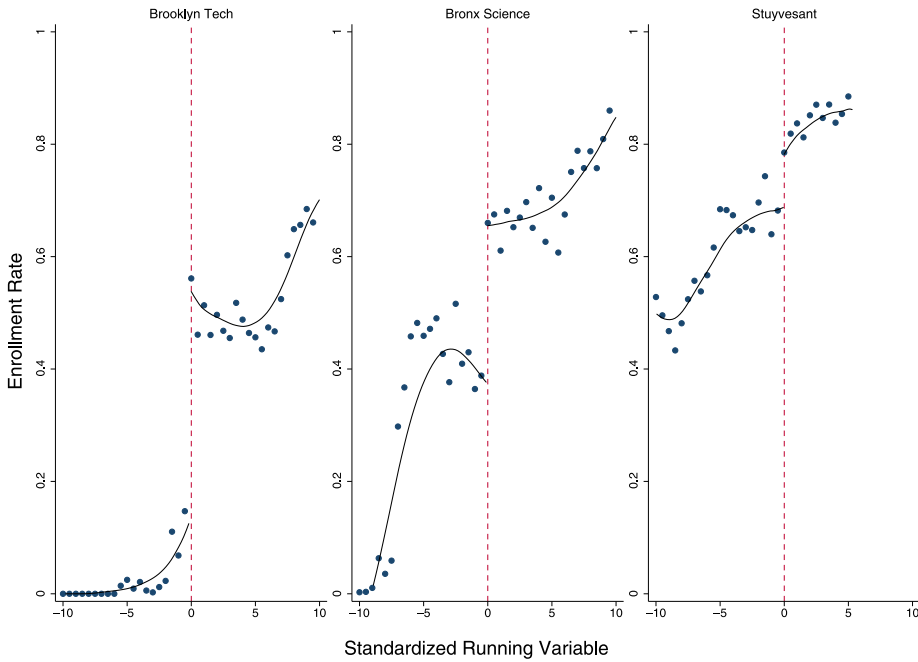
FIGURE 5.—This figure shows offers (a) at each NYC exam school and enrollment (b) at any NYC exam school, plotted against school-specific standardized running variables.

Offers at each exam school boost rates of overall exam school enrollment, defined here as enrollment at any of the three schools in our New York sample. This pattern is documented in Figure 5(b). The any-school enrollment jumps seen at the Brooklyn Tech and Bronx Science cutoffs are larger than the corresponding jump at Stuyvesant, implying that many students not offered a seat at Stuyvesant are offered and attend one of the other two exam schools.

New York has considerable school choice, with other selective public schools outside the set of traditional exam schools. Admission to one of the three traditional exam schools is nevertheless associated with a sharp jump in peer achievement, as can be seen in Figure 6(a). The average baseline math score of peers increases by about  $0.5\sigma$  at the Brooklyn Tech cutoff. The peer quality jump is smaller for Bronx Science and Stuyvesant, though still substantial at about  $0.2\sigma$ . Peer averages for English move similarly. As at Boston's exam school cutoffs, qualification for a New York exam school seat induces a sharp drop in the proportion of peers who are nonwhite. This can be seen in Fig-

---

New York. In the notation introduced in Section 3, the information available for New York is  $Z_{ik}$ , but the underlying preference orderings,  $p_i$ , are not.



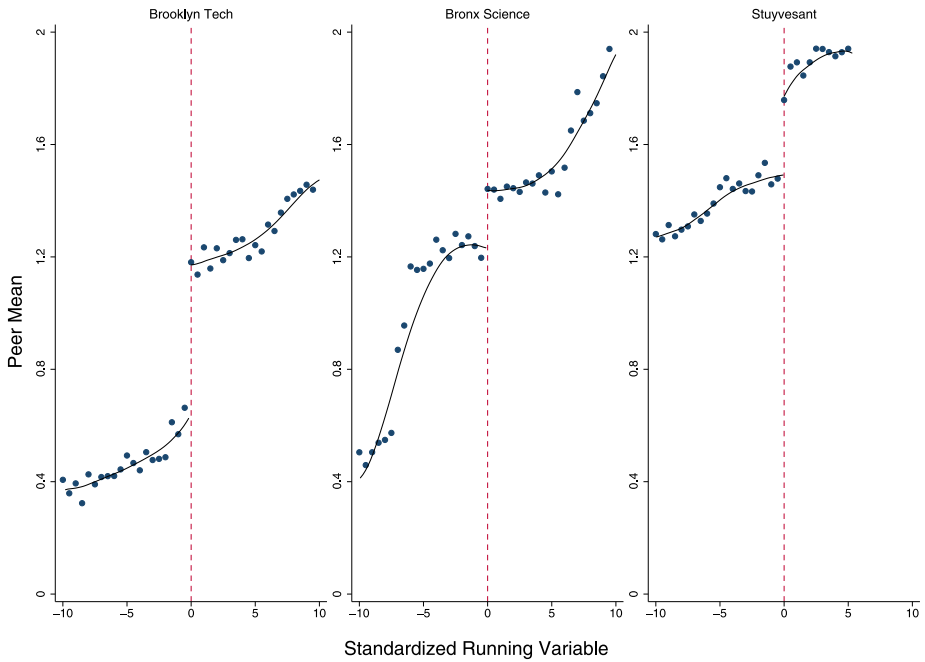
(b) Enrollment at any NYC exam school

FIGURE 5.—Continued.

ure 6(b), which shows a 20 percentage point drop in the proportion of peers nonwhite at Brooklyn Tech, and 10 percentage point drops at Bronx Science and Stuyvesant.

New York's exam schools expose successful applicants to a number of changes in school environment, but here, too, the largest and most consistent changes involve peer achievement and race. This can be seen in Table VII, which characterizes the changes in school environment experienced by New York exam school enrollment compliers. Class size changes less at New York exam school cutoffs than at Boston's. Differences in teacher experience across New York cutoffs are small.

Finally, as for Boston, reduced-form estimates for New York offer little evidence that exam schools boost achievement. This is apparent in Figures 7(a) and 7(b), which plot performance on the Advanced Math and English components of the New York Regents exam against standardized New York running variables. Table VIII reports the corresponding parametric and nonparametric estimates of offer effects on Advanced Math and English scores, as well as estimates for other Regents outcomes. The estimates here come from equations similar to (4.1) and (4.3), fit to samples of New York applicants in a  $[-6, +6]$  interval. These estimates are precise enough to rule out even modest score



(a) Baseline peer math score at NYC exam schools

FIGURE 6.—This figure shows the average peer math score (a) and the proportion of peers black or Hispanic (b) for NYC exam school applicants, plotted against school-specific standardized running variables.

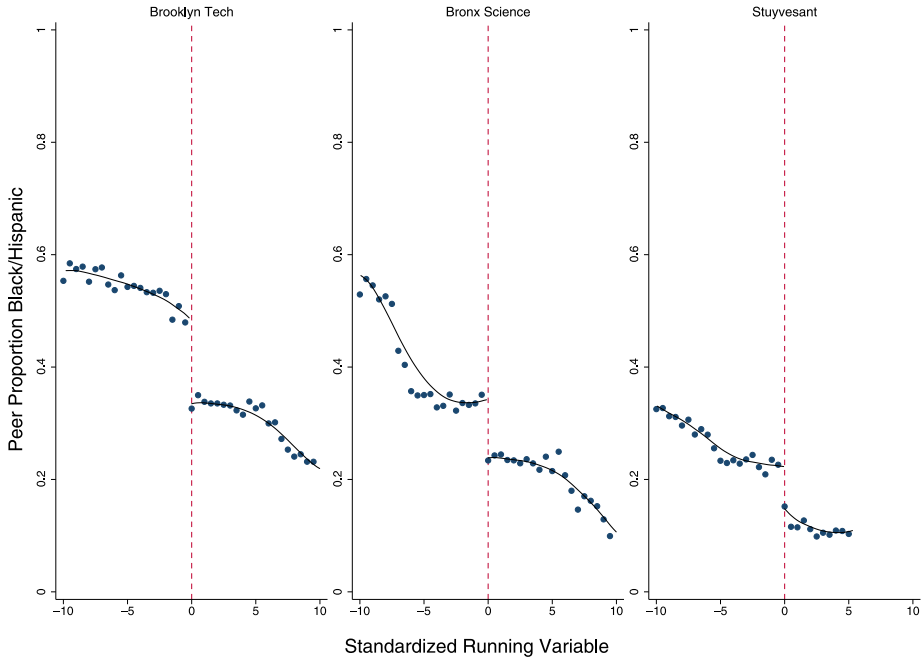
gains. For example, the nonparametric estimate of the effect on English scores in the stacked sample is  $0.01\sigma$ , with a standard error also around  $0.01$ . The few significant pooled estimates in Table VIII are negative.<sup>24</sup>

## 5. PEERS IN EDUCATION PRODUCTION

The reduced-form estimates reported here show remarkably little evidence of an exam school offer effect on test scores and post-secondary outcomes. These findings are relevant for policy questions related to exam school expansion, including contemporary proposals to lower admissions cutoffs and increase the number of exam school seats.<sup>25</sup> At the same time, we are also interested in the general lessons that might emerge from an exam school analysis.

<sup>24</sup>Appendix B reports additional descriptive statistics, results, and specification checks for New York.

<sup>25</sup>Vaznis (2009) discussed efforts to add 6th grade cohorts at Boston exam schools, while Hernandez (2008) reported on proposals to increase minority representation at New York's exam schools. Further afield, Lutton (2012) described a proposed exam school expansion in Chicago.



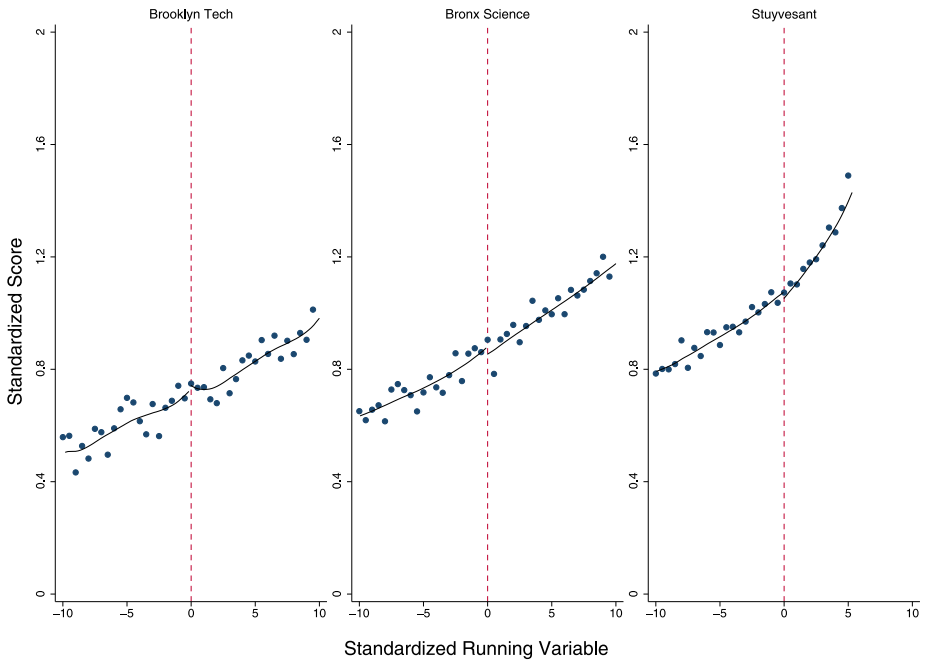
(b) Proportion black or Hispanic at NYC exam schools

FIGURE 6.—Continued.

TABLE VII  
NYC COMPLIER CHARACTERISTICS<sup>a</sup>

	Brooklyn Tech		Bronx Science		Stuyvesant	
	Z = 0 (1)	Z = 1 (2)	Z = 0 (3)	Z = 1 (4)	Z = 0 (5)	Z = 1 (6)
Baseline peer mean in math	0.35	1.60	1.21	1.75	1.58	2.12
Baseline peer mean in English	0.31	1.44	1.21	1.69	1.51	2.08
Proportion black or Hispanic	0.57	0.23	0.39	0.12	0.20	0.05
Proportion qualifying for a free lunch	0.65	0.61	0.60	0.65	0.64	0.68
Proportion female	0.53	0.41	0.52	0.45	0.46	0.43
Average English class size	29.3	31.8	27.1	31.8	31.0	29.1
Average math class size	29.0	31.1	27.8	31.6	30.9	33.0
Proportion of teachers fully licensed	0.96	0.97	0.97	0.97	0.97	0.99
Proportion of teachers highly educated	0.45	0.59	0.51	0.60	0.60	0.64
Proportion of teachers with less than 3 years experience	0.13	0.07	0.12	0.12	0.10	0.07

<sup>a</sup>This table shows descriptive statistics for NYC exam school enrollment compliers to the left ( $Z = 0$ ) and right ( $Z = 1$ ) of admission cutoffs, using data on applicants for admission from 2004–2007. Student-weighted average characteristics are reported for teachers and schools. *Fully licensed teachers* are those who have Provisional, Initial, or Professional licenses to teach in their subject(s). *Highly educated teachers* have Masters or other graduate degrees.

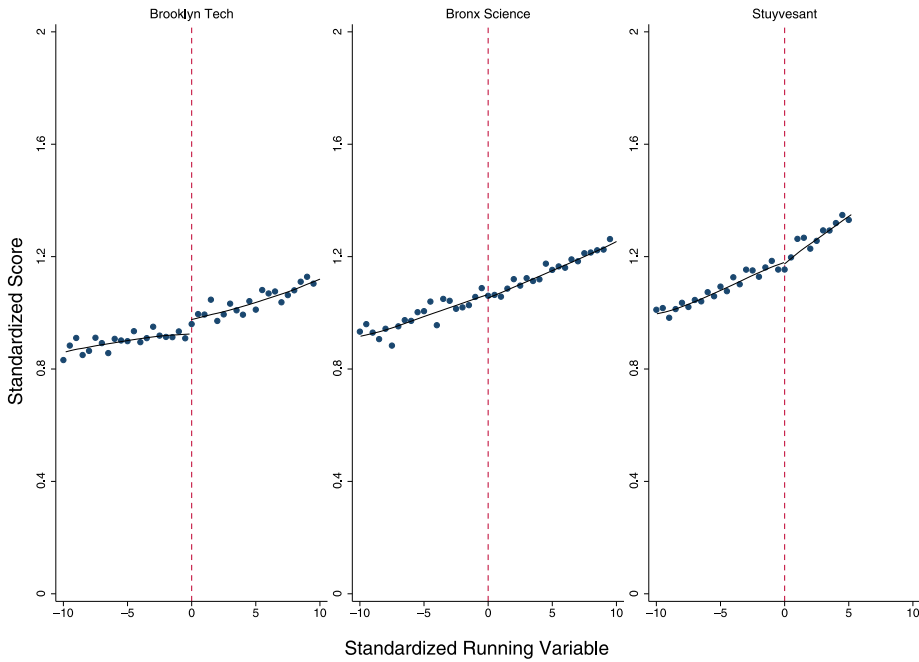


(a) Regents Advanced Math at NYC exam schools

FIGURE 7.—This figure shows average Regents Advanced Math (a) and English (b) scores for applicants to NYC exam schools, plotted against school-specific standardized running variables.

What is the exam school treatment? An overall change in school quality is hard to document or even define, but it is clear that exam school students gain the opportunity to study with high-achieving peers. The peer achievement changes documented here emerge at each exam school admissions cutoff. In other words, each admissions cutoff induces a substantial “peer achievement experiment,” in spite of the fact that overall exam school admission probabilities jump markedly only at cutoffs for the least selective schools (O’Bryant in Boston and Brooklyn Tech in New York). Jumps in peer achievement allow us to identify causal peer effects. Moreover, because the six exam school cutoffs under consideration intersect the applicant ability distribution over a wide range, the resulting estimates reflect peer effects for exceptionally high-ability as well as moderate-ability students.

In addition to manipulating peer achievement, admissions cutoffs induce a sharp change in racial composition, with large shifts at each cutoff. The exam school racial mix partly reflects the selective admissions policies that drive peer achievement: Because white applicants have higher test scores than do nonwhites (in this case, black and Hispanic applicants), the enrolled population is disproportionately white. Successful exam school applicants therefore receive



(b) Regents English at NYC exam schools

FIGURE 7.—Continued.

the same sort of treatment generated by our imaginary voucher experiment for Waltham homeowners on the Newton town line: the opportunity to attend school with fewer minority as well as higher-achieving classmates. These observations motivate models that specify peer race and ability characteristics as the primary causal channels mediating exam school offer effects.

The 2SLS estimates of peer achievement and racial composition effects reported here come from specifications and samples paralleling those used for the pooled reduced-form estimates reported in Tables III and VIII (pooling applicant grades and test years, as well as schools). All control variables, including year and grade of test, application cohort effects, and own- and other-school running variable controls, are subsumed in a vector  $X_{it}$ , with conformable coefficient vector  $\Gamma$ . The 2SLS second stage can then be written

$$(5.1) \quad y_{it} = \Gamma' X_{it} + \psi a_{it} + \varepsilon_{it},$$

where  $a_{it}$  is a vector of endogenous variables to be instrumented,  $\psi$  is the causal effect of interest, and  $\varepsilon_{it}$  is the 2SLS residual. The corresponding first-stage equations include the same controls plus offer dummies as excluded instruments. We estimated both (5.1) and the first-stage equations using the non-

TABLE VIII  
 NYC REDUCED-FORM ESTIMATES: REGENTS EXAMS<sup>a</sup>

	Parametric Estimates				Nonparametric (DM) Estimates			
	Brooklyn Tech (1)	Bronx Science (2)	Stuyvesant (3)	All Schools (4)	Brooklyn Tech (5)	Bronx Science (6)	Stuyvesant (7)	All Schools (8)
Math	0.048 (0.060) 5,116	-0.105* (0.054) 4,479	-0.05 (0.044) 4,259	-0.032 (0.027) 13,854	0.01 (0.039) 3,990	-0.132*** (0.033) 4,479	-0.045 (0.039) 2,915	-0.062*** (0.018) 11,384
Advanced Math	-0.081 (0.072) 6,758	-0.040 (0.062) 6,605	-0.023 (0.040) 7,308	-0.046 (0.038) 20,671	-0.013 (0.047) 4,859	-0.053 (0.040) 6,605	-0.026 (0.026) 5,350	-0.034 (0.024) 16,814
English	0.030 (0.051) 5,926	-0.042 (0.038) 5,506	-0.020 (0.033) 5,693	-0.011 (0.025) 17,125	0.048 (0.038) 5,926	-0.011 (0.021) 5,506	-0.005 (0.022) 5,693	0.012 (0.013) 17,125
Global History	-0.112** (0.048) 7,540	-0.039 (0.036) 7,103	-0.008 (0.036) 7,635	-0.051** (0.023) 22,278	-0.060* (0.031) 6,920	-0.013 (0.027) 7,103	0.014 (0.024) 5,918	-0.020 (0.014) 19,941
U.S. History	-0.100*** (0.037) 5,316	-0.012 (0.030) 5,139	0.032 (0.032) 5,486	-0.024 (0.023) 15,941	-0.036 (0.022) 3,886	-0.015 (0.021) 5,139	0.038 (0.023) 3,913	-0.006 (0.014) 12,938
Living Environment	-0.077** (0.038) 6,980	0.069* (0.037) 6,575	-0.061* (0.034) 6,991	-0.024 (0.020) 20,546	-0.078*** (0.022) 6,980	0.057** (0.024) 5,665	-0.032 (0.020) 6,991	-0.024** (0.012) 19,636

<sup>a</sup>This table reports estimates of the effect of New York exam school offers on Regents scores. The discontinuity sample includes applicants 5 standardized units from the cutoff. Model parameterization and estimation procedures parallel those 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%.

parametric bandwidth chosen for the associated single-school reduced forms in each city.

A simple causal model of education production facilitates interpretation of 2SLS estimates of equation (5.1). Ignoring the time dimension, a vector  $m_i$  is assumed to contain education inputs measured in the exam school entry grade. These inputs include peer achievement and race, measures of school quality, and teacher effects. Our goal is to identify the causal effects of variation in a subset of these inputs at a specific point in the education profile, holding earlier inputs and family background fixed.<sup>26</sup> A parsimonious representation of the education production function linking contemporaneous inputs with achievement is

$$y_i = \pi' m_i + \eta_i,$$

where  $\eta_i$  is the random part of potential outcomes revealed under alternative assignments of the input bundle,  $m_i$ . We partition  $m_i$  into observed peer achievement and racial composition,  $a_i$ , and unobserved inputs,  $w_i$ . That is,

$$m_i = [a_i' w_i']',$$

with conformable vectors of coefficients, “beta” and “gamma,” so that we can write the structural education production function as

$$(5.2) \quad y_i = \beta' a_i + \gamma' w_i + \eta_i,$$

where  $w_i$  is defined so that  $\gamma$  is positive.

The instrument vector in this context,  $z_i$ , indicates exam school offers. Offers are assumed to be independent of potential outcomes (i.e., independent of  $\eta_i$ ), without necessarily satisfying an exclusion restriction. In other words, exam school offers, taken to be as good as randomly assigned in a nonparametric RD setup, lead to exam school enrollment, which in turn changes peer characteristics and perhaps other features of the school environment, denoted by  $w_i$ . We capture these changes in the following first-stage relationships:

$$a_i = \theta_1' z_i + \nu_{1i},$$

$$w_i = \theta_2' z_i + \nu_{2i},$$

where first-stage residuals are orthogonal to the instruments by construction, but possibly correlated with  $\eta_i$ . The proposition below characterizes the causal effects identified by 2SLS given this structure:

<sup>26</sup>Todd and Wolpin (2003) discussed the conceptual distinction between this type of interruption-based causal relationship and a complete cumulative education production function.



PROPOSITION 1: 2SLS estimates using  $z_i$  as an instrument for  $a_i$  in (5.2), omitting  $w_i$ , identify  $\beta + \delta'\gamma$ , where  $\delta$  is the population 2SLS coefficient vector from a regression of  $w_i$  on  $a_i$ , using  $z_i$  as instruments.

This is a 2SLS version of the omitted variables bias formula (see, e.g., Angrist and Krueger (1992)). Proposition 1 implies that if  $\delta$  is positive (because exam schools have better unmeasured inputs), 2SLS estimates of peer effects omitting  $w_i$  tend to be too big. The notion that omitted variables bias is likely to be positive seems reasonable in this context; among other distinctions, Boston and New York exam schools feature, to varying degrees, a rich array of course offerings, relatively modern facilities, and a challenging curriculum meant to prepare students for college.

An alternative interpretation under somewhat stronger assumptions is based on the notion that any input correlated with exam school offers is itself caused by  $a_i$ . In other words, the relationship between  $w_i$  and exam school offers is a *consequence* of the effect of exam school attendance on peer characteristics (exam school curricula are challenging because exam school students are high-achieving; the prevalence of nonwhite students affects course content). Suppose the causal effect of  $a_i$  on  $w_i$  is described by a linear constant effects model with coefficient vector  $\lambda$ . Then we have

$$(5.3) \quad w_i = \lambda a_i + \xi_i, \\ E[z_i \xi_i] = 0.$$

This assumption generates a triangular structure in which 2SLS estimates combine both the direct and indirect effects of peers.<sup>27</sup> When other inputs are causally downstream to peer characteristics, 2SLS estimates of peer effects omitting  $w_i$  capture the total impact of randomly assigning  $a_i$ . In other words, in this scenario, 2SLS identifies  $\beta + \lambda'\gamma$ .

### 5.1. Estimates

To maximize precision and facilitate exploration of models with multiple endogenous variables, we constructed 2SLS estimates using a combined Boston and New York sample, with six offer dummies as instruments. The 2SLS specifications parallel those used to construct the single-city stacked reduced-form estimates, except that here the stack includes six schools. In addition to estimates using one offer dummy for each school as instruments, we also report 2SLS estimates from more heavily over-identified models adding interac-

<sup>27</sup>A referee points out that the list of omitted variables affected by exam school enrollment might include parental behavior such as help with homework or the provision of tutors.

tions between offer dummies and applicant cohort dummies to the instrument list.

Table IX reports first-stage estimates and the associated F statistics (adjusted, where appropriate, for multiple endogenous variables), as well as second-stage estimates. Consistent with the figures, the first-stage estimates show large, precisely estimated offer effects on peer achievement and racial composition. For example, an O'Bryant offer increases average baseline math peer scores by over two-thirds of a standard deviation, while the math peer achievement gain is about  $0.36\sigma$  at the Latin Academy cutoff, and  $0.77\sigma$  at the Latin School cutoff. Peer achievement also shifts sharply at New York cutoffs, though less than in Boston. First stages for racial composition show that offers induce a 12–23 percentage point reduction in the proportion nonwhite at each Boston cutoff, and a 6–15 percentage point reduction at cutoffs in New York.

Consistent with the reduced-form offer estimates discussed in the previous section, 2SLS estimates treating peer achievement as a single endogenous variable show no evidence of a statistically significant peer effect. These results appear in columns (1) and (6) of Table IX. Importantly, the 2SLS estimates and the associated standard errors, on the order of 0.03, also provide a basis for comparisons. For example, these estimates allow us to reject the corresponding large positive OLS estimates of peer effects reported as a benchmark in our working paper (Abdulkadiroğlu, Angrist, and Pathak (2011)). The small peer effects in Table IX are also significantly different from estimates of conceptually similar education peer effects reported elsewhere. Examples include Hoxby (2000) (with effects on the order of  $0.3$ – $0.5\sigma$ ), Hanushek et al. (2003) (effects on the order of  $0.15$ – $0.24\sigma$ ), and estimates from many other studies summarized in Sacerdote's (2011) recent survey.

2SLS estimates of racial composition effects, reported in columns (2) and (7) of Table IX, likewise show no statistically significant evidence of a substantial impact, though these are less precise than the corresponding peer achievement effects. At the same time, we can easily rule out large negative effects of proportion nonwhite. (Compare, e.g., estimates reported in Hoxby (2000) ranging from  $-1$  to  $-2$  for black and Hispanic third graders.)

Models with two endogenous variables capture pairs of causal effects at the same time. These models, identified by variation at six admissions cutoffs, allow for the possibility that different sorts of causal effects are reinforcing or offsetting. We also introduce a secular exam school effect, parameterized as operating through years of exam school enrollment. This provides a simple adjustment for possible violations of the exclusion restriction in models with specific causal channels. Results from models with multiple endogenous variables are naturally less precise than the estimates generated by models with a single channel. Except possibly for a large positive effect of the proportion nonwhite on applicants' math scores in column (3) of Table IX, multiple-endogenous-variable estimates are consistent with those generated by models allowing only a single causal channel.

TABLE IX  
2SLS ESTIMATES FOR BOSTON AND NEW YORK<sup>a</sup>

	Math					English				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	2SLS Estimates (Models With Cohort Interactions)									
Peer mean	-0.038 (0.032)		0.064 (0.080)	-0.035 (0.044)		0.006 (0.030)		0.044 (0.064)	-0.047 (0.051)	
Proportion nonwhite		0.145 (0.110)	0.421 (0.279)		0.160 (0.137)		-0.014 (0.102)	0.141 (0.218)		0.063 (0.134)
Years in exam school				-0.003 (0.036)	0.006 (0.030)				0.045 (0.034)	0.027 (0.025)
	First-Stage F-Statistics (Models With Cohort Interactions)									
Peer mean	65.8		9.1	50.0		39.8		5.7	22.8	
Proportion nonwhite		65.8	17.6		60.0		52.3	12.4		41.2
Years in exam school				12.0	16.2				10.6	15.8
N	31,911	33,313	31,911	31,911	33,313	31,222	32,185	31,222	31,222	32,185

(Continues)

TABLE IX—Continued

	Math					English				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
First-Stage Estimates (Models Without Cohort Interactions)										
Panel A. Boston										
O'Bryant	0.763*** (0.071)	-0.119*** (0.013)				0.702*** (0.064)	-0.118*** (0.013)			
Latin Academy	0.355*** (0.073)	-0.210*** (0.014)				0.355*** (0.063)	-0.212*** (0.014)			
Latin School	0.769*** (0.037)	-0.225*** (0.011)				0.632*** (0.033)	-0.216*** (0.011)			
Panel B. NYC										
Brooklyn Tech	0.486*** (0.074)	-0.137*** (0.024)				0.517*** (0.058)	-0.152*** (0.022)			
Bronx Science	0.174*** (0.067)	-0.101*** (0.031)				0.158** (0.074)	-0.098*** (0.034)			
Stuyvesant	0.264*** (0.076)	-0.066*** (0.022)				0.255*** (0.096)	-0.060** (0.029)			

<sup>a</sup>This table reports two-stage least squares (2SLS) estimates of the effects of peer characteristics on test 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 nonparametric estimates using bandwidths computed one school at a time. The 2SLS estimates and first-stage F-statistics reported in the upper half of the table are from models that interact exam school offers with application cohort dummies. The first-stage coefficient estimates shown in the lower half of the table are from models without these interactions. Robust standard errors, clustered on year and school, are shown in parentheses. Standard errors also cluster on student. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

## 6. SUMMARY AND CONCLUSIONS

The results reported here suggest that an exam school education produces only scattered gains for applicants, even among students with baseline scores close to or above the mean in the target school. Because the exam school experience is associated with sharp increases in peer achievement, these results weigh against the importance of peer effects in the education production function. Our results also fail to uncover systematic evidence of racial composition effects. The outcome most strengthened by exam school attendance appears to be the 10th grade English score, a result driven partly by gains for minorities. Given the history of racial preferences (and their more recent elimination) in Boston's exam schools, this finding seems worth further exploration. Overall, however, while the exam school students in our samples typically have good outcomes, most of these students would likely have done well without the benefit of an exam school education.

Of course, test scores and peer effects are only part of the exam school story. It may be that preparation for exam school entrance is itself worthwhile. The RD design captures the impact of peer composition and possibly other changes at admissions cutoffs, while ignoring effects common to applicants on both sides. Likewise, unique features of an exam school education may boost achievement in specific subject areas. Students who attend Boston Latin School almost certainly learn more Latin than they would have otherwise. The many clubs and activities found at some exam schools may expose students to ideas and concepts not easily captured by achievement tests or our post-secondary outcomes. It is also possible that exam school graduates earn higher wages, a question we plan to explore in future work. Still, the estimates reported here suggest that any labor market gains are likely to come through channels other than peer composition and increased cognitive achievement.

Can these results be reconciled with those of other studies reporting a strong association between group averages and individual outcomes? Every context is different, and the absence of peer effects in one setting does not prove that such effects are unimportant elsewhere. At the same time, it is worth emphasizing the high risk of specification error in peer analysis. As noted by [Manski \(2000\)](#) and [Angrist and Pischke \(2009\)](#), among others, the relationship between an individual-level variable and any group average of this variable is essentially mechanical. Likewise, any regression of one outcome variable on the group average of another outcome variable is biased toward a finding of peer effects due to the presence of common shocks. Random assignment to groups fails to solve either of these problems. Finally, any regression of an individual outcome on a combination of observationally varying individual- and group-level treatments is almost sure to produce something that looks like a peer effect for reasons related to the information content of the average (see, e.g., [Acemoglu and Angrist \(2000\)](#) and [Moffitt \(2001\)](#)). It is perhaps unsurprising, therefore, that experimental or quasi-experimental manipulation of predetermined peer

characteristics is less likely to uncover a peer effect than are other study designs. Estimates from the Moving to Opportunity demonstration, for example, which manipulated peer composition by randomly assigning housing vouchers, are consistent with the results reported here in offering little evidence of causal peer effects (see, e.g., Kling, Liebman, and Katz (2007)).

Our results are also relevant to the economic debate around school quality and school choice, as developed in analyses by Kane and Staiger (2002), Hastings, Kane, and Staiger (2009), Hsieh and Urquiola (2006), Rothstein (2006), and MacLeod and Urquiola (2009), among others. As with the jump in house prices at school district boundaries, heavy rates of exam school over-subscription suggest that parents believe peer composition matters a great deal for their children's welfare. The fact that we find little support for causal peer effects suggests that parents either mistakenly equate attractive 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 school choice in and of itself has strong salutary demand-side effects in education production.

Finally, our study makes a number of methodological contributions. As school choice has proliferated across districts, so, too, has the use of sophisticated assignment mechanisms such as deferred acceptance. We have shown how to craft a sharp regression discontinuity design from a deferred-acceptance match of students to schools. In the spirit of a suggestion by Lee and Lemieux (2010), we have also shown how information on a second variable with content similar to the running variable facilitates an exploration of the external validity of regression discontinuity estimates. Recent and ongoing work by Angrist and Rokkanen (2012) and Rokkanen (2013) pursues these ideas further.

#### APPENDIX A: ADDITIONAL RESULTS FOR BOSTON

Table A.I compares Boston exam school applicants with the general BPS population. Non-exam BPS students are mostly nonwhite 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. 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.

Students near admissions cutoffs should be similar at the time of application. Even so, subsequent attrition may lead to differences in the follow-up sample, unless the attrition process is also random. In other words, a threat to our research design is differential and selective attrition by exam offer status. For instance, students just below the cutoff may be less likely to be found than students above the cutoff if students below the cutoff leave the public school system when they do not obtain an exam offer. Differential attrition may generate selection bias. A simple test for selection bias looks at the effect offers

TABLE A.I  
 ADDITIONAL DESCRIPTIVE STATISTICS FOR BOSTON EXAM SCHOOL APPLICANTS<sup>a</sup>

	BPS (1)	O'Bryant			Latin Academy			Latin School		
		Applicants (2)	Enrolled (3)	Compliers (4)	Applicants (5)	Enrolled (6)	Compliers (7)	Applicants (8)	Enrolled (9)	Compliers (10)
Panel A. 7th Grade Applicants										
Female	0.478	0.539	0.582	0.543	0.540	0.559	0.626	0.541	0.561	0.577
Black	0.471	0.377	0.381	0.399	0.373	0.271	0.302	0.370	0.124	0.194
Hispanic	0.314	0.214	0.231	0.205	0.210	0.185	0.194	0.211	0.094	0.166
Free lunch	0.750	0.754	0.805	0.812	0.745	0.697	0.741	0.744	0.489	0.647
Limited English Proficient	0.203	0.144	0.146	0.156	0.142	0.122	0.138	0.142	0.083	0.047
Special Education <sup>b</sup>	0.234	0.047	0.010	0.009	0.047	0.008	0.010	0.047	0.007	0.007
Baseline Math <sup>c</sup>	-0.019	0.746	0.835	0.693	0.761	1.201	0.929	0.764	1.986	1.461
Baseline English <sup>c</sup>	-0.023	0.711	0.799	0.664	0.725	1.110	0.911	0.729	1.791	1.195
Panel B. 9th Grade Applicants										
Female	0.476	0.540	0.608	0.519	0.540	0.591	0.592	0.541	0.559	0.547
Black	0.494	0.479	0.410	0.388	0.480	0.311	0.299	0.479	0.189	0.205
Hispanic	0.341	0.258	0.226	0.230	0.255	0.206	0.274	0.256	0.170	0.257
Free lunch	0.762	0.807	0.808	0.766	0.805	0.815	0.854	0.805	0.648	0.908
Limited English Proficient	0.182	0.132	0.140	0.185	0.131	0.115	0.099	0.131	0.111	0.088
Special Education <sup>b</sup>	0.248	0.080	0.018	0.010	0.080	0.020	0.048	0.079	0.015	0.023
Baseline Math <sup>c</sup>	-0.313	0.230	0.848	0.685	0.230	1.423	1.284	0.229	1.717	1.402
Baseline English <sup>c</sup>	-0.246	0.276	0.702	0.638	0.278	1.085	0.842	0.278	1.269	1.053

<sup>a</sup>This table reports descriptive statistics for 7th grade applicants from 1999–2008 and for 9th grade applicants from 2001–2007. Column (1) reports descriptive statistics for 6th and 8th grade students in Boston public schools who had not previously enrolled in any exam school. Columns (2), (5), and (8) report descriptive statistics for applicants to each Boston exam school. Columns (3), (6), and (9) report descriptive statistics for students who enroll in each exam school the following fall. Columns (4), (7), and (10) report statistics for enrollment compliers at each admissions cutoff. Baseline math and English scores for 7th grade applicants are from 4th grade. Baseline scores for 9th grade applicants are from middle school.

<sup>b</sup>Information on special education status is available only for 1999–2004.

<sup>c</sup>Baseline scores are available from 2000 onward for 6th grade and from 2002 onward for grade 8.

have on the likelihood that an applicant contributes MCAS scores to our sample. If differences in follow-up rates across cutoffs are small, then selection bias from differential attrition is also likely to be modest.

Between 76% and 89% of applicants contribute MCAS outcomes to the discontinuity sample. This relatively high follow-up rate, documented in Table A.II, is likely due to the fact that the analysis here is limited to students who were enrolled in BPS at baseline. Follow-up differentials are estimated using models that parallel those used to construct the estimates in Table III. Most of the estimated differentials for math and English are small and not significantly different from zero. While the parametrically estimated follow-up differential is significant when schools are stacked, the estimated difference (on the order of 3 percent) seems too small to impart substantial bias.

A second potential threat to the validity of our research design is some sort of sorting behavior that changes in a discontinuous manner at admissions cutoffs. The fact that exam school admissions decisions are made in the BPS central office suggests that it is unlikely that schools have much discretion in selecting which applicants obtain offers at particular schools. Nonetheless, discontinuities in the characteristics of applicants may arise in situations where the admissions process is compromised.

Table A.III briefly examines this possibility. This table shows estimates from models that parallel the reduced-form models discussed in the text, but the dependent variables here are covariates. These results reveal little evidence of covariate imbalance across admissions cutoffs. Joint tests suggest that the few significant differences in covariates seen in the table are chance findings.

As background for the high-achievers estimation strategy reported on in Table IV, we explored the correlation between ISEE (exam school admissions) scores and MCAS scores. The two scores are correlated, but far from perfectly. For instance, the correlation between the verbal ISEE and grade 7 English is 0.62, while the correlation between verbal ISEE and grade 10 English is 0.58. The correlation between math ISEE and grade 8 Math is 0.76, while the correlation between math ISEE and grade 10 Math is 0.66.

Table A.IV reports high school MCAS estimates for grade 7 and grade 9 applicants separately. This breakdown is motivated in part by the fact that 7th grade applicants who enroll in an exam school have longer exam school exposure than 9th grade applicants who do so. As it turns out, however, the group for which evidence of an exam school gain is strongest consists of 9th grade applicants, especially for English scores.

Table A.V reports detailed MCAS results for the sample of Boston minority applicants. These results show no clear pattern, though, as noted in the text, there is reasonably strong evidence here for a gain in 10th grade English scores. There are also some significant negative effects.

Figure A.1 reports on the distance to school faced by students on either side of exam school admissions cutoffs. Travel distance was computed using straight line distance as measured by ArcGIS. Students offered a seat at O'Bryant



TABLE A.II  
BOSTON ATTRITION DIFFERENTIALS<sup>a</sup>

Application Grade	Test Grade	Fraction With Followup	Parametric Estimates				Nonparametric (DM) Estimates			
			O'Bryant (1)	Latin Academy (2)	Latin School (3)	All Schools (4)	O'Bryant (5)	Latin Academy (6)	Latin School (7)	All Schools (8)
Panel A. Math										
7th	7th and 8th	0.875	0.046 (0.049) 4,775	0.072* (0.041) 4,845	-0.023 (0.041) 4,187	0.034 (0.026) 13,807	0.043 (0.028) 4,258	0.023 (0.021) 4,609	0.032 (0.025) 3,384	0.032** (0.014) 12,251
7th and 9th	10th	0.759	0.072 (0.050) 4,652	0.002 (0.056) 3,613	0.005 (0.057) 3,058	0.034 (0.031) 11,323	0.022 (0.029) 4,243	0.004 (0.034) 2,673	0.055 (0.036) 2,254	0.025 (0.020) 9,170
7th and 9th	7th, 8th, and 10th	0.823	0.059 (0.041) 9,427	0.041 (0.039) 8,458	-0.011 (0.040) 7,245	0.034 (0.024) 25,130	0.032 (0.024) 8,501	0.017 (0.021) 7,282	0.041* (0.025) 5,638	0.029** (0.014) 21,421
Panel B. English										
7th	7th and 8th	0.891	0.046 (0.049) 4,786	0.081* (0.043) 4,883	-0.011 (0.040) 4,164	0.041 (0.026) 13,833	0.036 (0.027) 4,519	0.032 (0.024) 4,250	0.018 (0.021) 3,860	0.029** (0.014) 12,629
7th and 9th	10th	0.761	0.080 (0.050) 4,652	-0.021 (0.056) 3,613	0.020 (0.056) 3,058	0.034 (0.031) 11,323	0.027 (0.028) 4,515	-0.022 (0.037) 2,351	0.056 (0.035) 2,361	0.022 (0.020) 9,227
7th and 9th	7th, 8th, and 10th	0.832	0.063 (0.040) 9,438	0.036 (0.039) 8,496	0.002 (0.037) 7,222	0.038* (0.023) 25,156	0.031 (0.022) 9,034	0.012 (0.023) 6,601	0.031 (0.021) 6,221	0.026** (0.013) 21,856

<sup>a</sup>This table reports estimates of the effects of exam school offers on an indicator for non-missing outcome scores. The specification and estimation procedures are the same as used to construct the estimates in Table III. The fraction with MCAS is the probability an MCAS score is observed for applicants who appear in any school-specific discontinuity sample. Robust standard errors, clustered on year and school, are shown in parentheses. Standard errors for the all-schools estimates also cluster on student. Sample sizes are shown below standard errors. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

TABLE A.III  
BOSTON COVARIATE DISCONTINUITIES<sup>a</sup>

Covariate	Mean	Parametric Estimates				Nonparametric (DM) Estimates			
		O'Bryant (1)	Latin Academy (2)	Latin School (3)	All Schools (4)	O'Bryant (5)	Latin Academy (6)	Latin School (7)	All Schools (8)
Panel A. 7th Grade Applicants									
Female	0.567	-0.060 (0.061) 3,696	-0.024 (0.065) 3,728	-0.015 (0.070) 3,236	-0.035 (0.037) 10,660	-0.026 (0.036) 3,368	0.061 (0.041) 2,760	0.071 (0.043) 2,456	0.028 (0.024) 8,584
Black	0.307	0.063 (0.058) 3,677	0.023 (0.060) 3,709	0.026 (0.056) 3,211	0.039 (0.033) 10,597	-0.024 (0.035) 3,330	0.047 (0.031) 3,709	0.031 (0.029) 3,211	0.020 (0.018) 10,250
Hispanic	0.199	-0.052 (0.052) 3,677	-0.045 (0.055) 3,709	-0.048 (0.054) 3,211	-0.048 (0.031) 10,597	-0.017 (0.028) 3,677	-0.029 (0.028) 3,709	-0.017 (0.030) 3,145	-0.021 (0.017) 10,531
Free lunch	0.724	0.035 (0.048) 3,696	-0.108** (0.055) 3,728	-0.116* (0.065) 3,236	-0.056* (0.032) 10,660	0.010 (0.032) 2,862	-0.068** (0.029) 3,728	-0.037 (0.034) 3,236	-0.038** (0.019) 9,826
Limited English Proficient	0.126	-0.020 (0.042) 3,696	-0.021 (0.042) 3,728	-0.131*** (0.037) 3,236	-0.053** (0.024) 10,660	0.012 (0.023) 3,696	-0.023 (0.026) 2,931	-0.076*** (0.021) 3,135	-0.023* (0.014) 9,762
Special Education	0.017	-0.020 (0.021) 3,696	0.004 (0.013) 3,728	0.000 (0.013) 3,236	-0.006 (0.010) 10,660	-0.018 (0.011) 3,415	0.004 (0.007) 3,402	0.005 (0.007) 2,763	-0.004 (0.005) 9,580
Joint <i>p</i> -value		0.559	0.525	0.010	0.107	0.629	0.053	0.005	0.123

(Continues)

TABLE A.III—Continued

Covariate	Mean	Parametric Estimates				Nonparametric (DM) Estimates			
		O'Bryant (1)	Latin Academy (2)	Latin School (3)	All Schools (4)	O'Bryant (5)	Latin Academy (6)	Latin School (7)	All Schools (8)
Panel B. 9th Grade Applicants									
Female	0.602	−0.065 (0.083) 1,978	−0.025 (0.117) 978	−0.018 (0.172) 754	−0.047 (0.064) 3,710	0.024 (0.048) 1,875	−0.055 (0.065) 978	−0.059 (0.100) 501	−0.012 (0.036) 3,354
Black	0.409	−0.006 (0.085) 1,967	0.063 (0.108) 971	−0.046 (0.138) 750	0.008 (0.061) 3,688	−0.023 (0.046) 1,967	0.028 (0.084) 500	−0.087 (0.086) 487	−0.023 (0.037) 2,954
Hispanic	0.237	−0.049 (0.072) 1,967	0.075 (0.103) 971	0.013 (0.131) 750	−0.005 (0.055) 3,688	0.008 (0.044) 1,686	0.063 (0.069) 643	0.027 (0.093) 363	0.024 (0.034) 2,692
Free lunch	0.789	−0.044 (0.069) 1,978	0.039 (0.088) 978	−0.278* (0.150) 754	−0.051 (0.052) 3,710	0.013 (0.043) 1,561	0.095* (0.051) 978	0.037 (0.087) 396	0.046 (0.031) 2,935
Limited English Proficient	0.120	−0.002 (0.057) 1,978	−0.096 (0.070) 978	0.038 (0.098) 754	−0.023 (0.042) 3,710	0.022 (0.034) 1,889	−0.032 (0.041) 913	0.047 (0.060) 389	0.009 (0.025) 3,191
Special Education	0.031	0.013 (0.025) 1,978	0.061 (0.047) 978	−0.072** (0.035) 754	0.015 (0.020) 3,710	−0.014 (0.014) 1,824	0.024 (0.025) 793	−0.008 (0.028) 455	−0.003 (0.011) 3,072
Joint $p$ -value		0.940	0.589	0.240	0.906	0.906	0.354	0.887	0.840

<sup>a</sup>This table reports discontinuities in covariates estimated using models like those used to construct the reduced-form estimates in Table III. Robust standard errors, clustered on year and school, are shown in parentheses. Standard errors for the all-schools estimates also cluster on student. Sample sizes are shown below standard errors. The joint  $p$ -value in the bottom row is for tests looking at all covariate discontinuities at once. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

TABLE A.IV  
 ADDITIONAL BOSTON ESTIMATES FOR 10TH GRADERS: MCAS MATH AND ENGLISH<sup>a</sup>

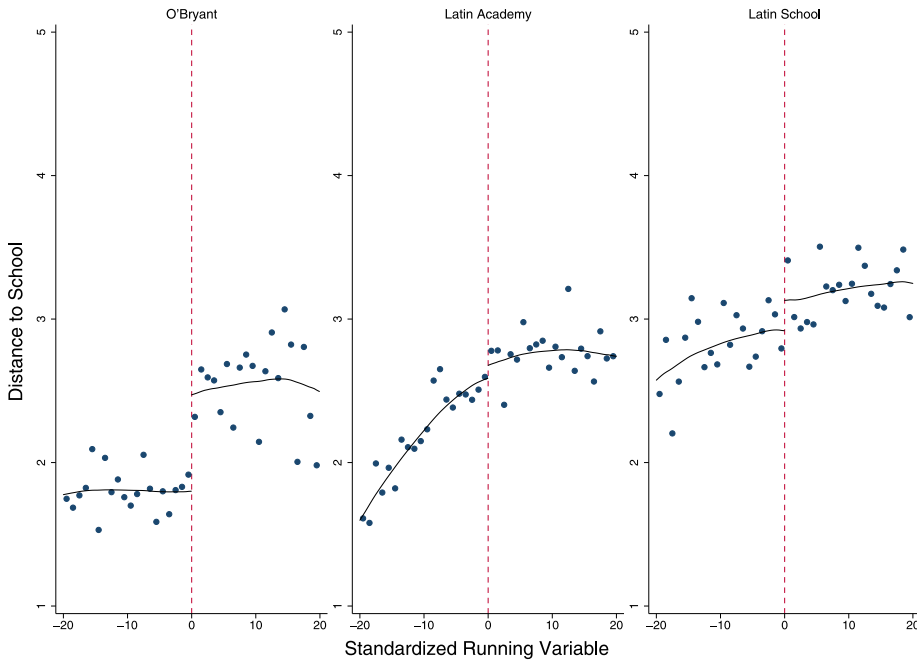
Application Grade	Test Grade	Parametric Estimates				Nonparametric (DM) Estimates			
		O'Bryant (1)	Latin Academy (2)	Latin School (3)	All Schools (4)	O'Bryant (5)	Latin Academy (6)	Latin School (7)	All Schools (8)
Panel A. Math									
7th	10th	-0.015 (0.092)	-0.190* (0.113)	-0.053 (0.055)	-0.085** (0.042)	0.027 (0.052)	-0.118* (0.066)	-0.055* (0.032)	-0.042 (0.031)
		1,832	1,920	1,854	5,606	1,699	1,423	1,467	4,589
9th	10th	0.159 (0.099)	0.109 (0.081)	-0.072 (0.114)	0.112 (0.069)	0.118* (0.064)	0.097* (0.053)	-0.107* (0.058)	0.083** (0.040)
		1,557	789	605	2,951	1,384	604	360	2,348
Panel B. English									
7th	10th	0.035 (0.097)	0.144 (0.108)	0.015 (0.082)	0.065 (0.055)	0.117** (0.055)	0.178*** (0.063)	-0.008 (0.069)	0.094** (0.040)
		1,836	1,925	1,857	5,618	1,778	1,325	1,459	4,562
9th	10th	0.192 (0.118)	0.118 (0.184)	0.081 (0.179)	0.154* (0.088)	0.168** (0.067)	0.195 (0.124)	0.021 (0.089)	0.150*** (0.050)
		1,562	790	606	2,958	1,530	461	457	2,448

<sup>a</sup>This table reports estimates of the effects of exam school offers on 10th grade MCAS scores separately by application grade. Models and methods parallel those used to construct the estimates in Table III. Robust standard errors, clustered on year and school, are shown in parentheses. Standard errors for the all-school estimates also cluster on student. The sample size is reported below standard errors. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

TABLE A.V  
 ADDITIONAL BOSTON REDUCED-FORM ESTIMATES FOR BLACKS AND HISPANICS: MCAS MATH AND ENGLISH<sup>a</sup>

Application Grade	Test Grade	Parametric Estimates				Nonparametric (DM) Estimates			
		O'Bryant (1)	Latin Academy (2)	Latin School (3)	All Schools (4)	O'Bryant (5)	Latin Academy (6)	Latin School (7)	All Schools (8)
Panel A. Math									
7th	7th and 8th	-0.275** (0.128) 2,554	-0.213** (0.108) 2,196	-0.135 (0.151) 1,325	-0.222*** (0.073) 6,075	-0.153* (0.079) 2,185	-0.135 (0.086) 1,991	-0.069 (0.093) 1,258	-0.127*** (0.044) 5,434
7th and 9th	10th	0.029 (0.098) 2,237	-0.190 (0.116) 1,441	-0.043 (0.074) 907	-0.054 (0.067) 4,585	0.024 (0.066) 1,824	-0.103 (0.071) 1,073	-0.075 (0.052) 687	-0.034 (0.044) 3,584
7th and 9th	7th, 8th, and 10th	-0.133 (0.090) 4,791	-0.203** (0.083) 3,637	-0.101 (0.109) 2,232	-0.150*** (0.054) 10,660	-0.076 (0.061) 4,009	-0.124** (0.053) 3,064	-0.071 (0.069) 1,945	-0.092*** (0.032) 9,018
Panel B. English									
7th	7th and 8th	-0.051 (0.092) 2,616	-0.187** (0.093) 2,246	-0.245** (0.102) 1,330	-0.142** (0.059) 6,192	-0.056 (0.047) 2,616	-0.013 (0.067) 2,130	-0.113** (0.052) 1,330	-0.055 (0.035) 6,076
7th and 9th	10th	0.127 (0.099) 2,247	0.218* (0.112) 1,443	-0.131 (0.134) 909	0.110 (0.072) 4,599	0.162*** (0.063) 2,091	0.238*** (0.072) 961	0.039 (0.091) 668	0.160*** (0.045) 3,720
7th and 9th	7th, 8th, and 10th	0.031 (0.078) 4,863	-0.021 (0.086) 3,689	-0.205** (0.096) 2,239	-0.036 (0.055) 10,791	0.034 (0.043) 4,707	0.065 (0.061) 3,091	-0.067 (0.052) 1,998	0.022 (0.032) 9,796

<sup>a</sup>This table reports additional estimates of the effects of exam school offers on MCAS scores for minority applicants. Models and methods parallel those used to construct the estimates in Table III. Robust standard errors, clustered on year and school are shown in parentheses. Standard errors for the all-school estimates also cluster on student. The number of observations is reported below standard errors. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.



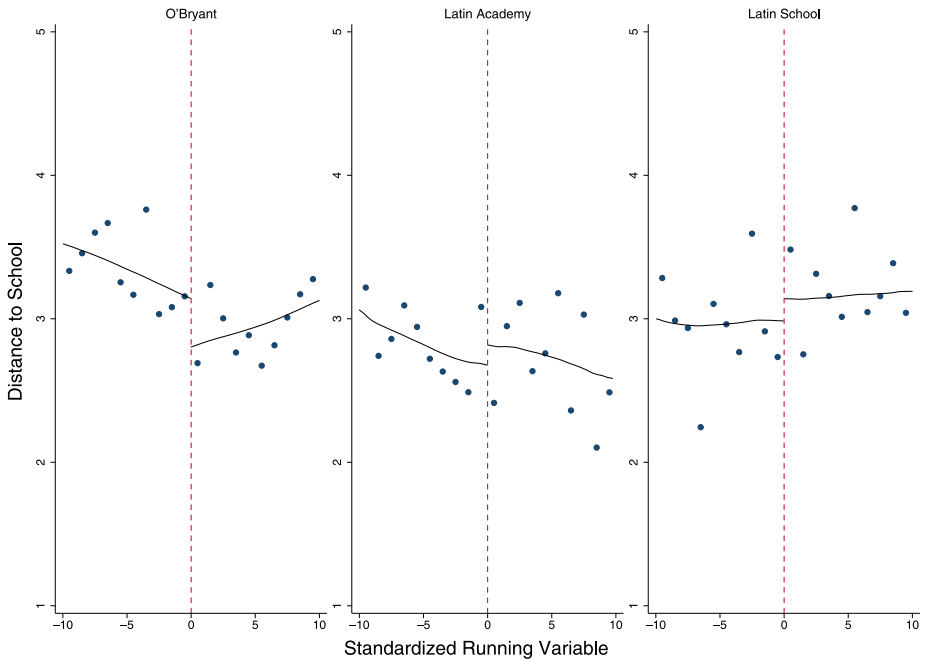
(a) Distance to school for 7th grade applicants to Boston exam schools

FIGURE A.1.—This figure shows the average distance in miles between students' home addresses and the school attended in the year after application for 7th grade (a) and 9th grade (b) applicants to Boston exam schools, plotted against a school-specific standardized running variables.

travel about 0.5 miles farther in grade 7, but are 0.2 miles closer in grade 9. The figure shows modest differences in commuting distance across the Latin Academy and Latin School cutoffs for both grade 7 and grade 9.

#### APPENDIX B: ADDITIONAL RESULTS FOR NEW YORK

As in Boston, New York exam school applicants are positively selected relative to the population of New York 8th graders. This is documented in Table B.I, which reports descriptive statistics for New York 8th graders, exam school applicants, exam school students, and exam school enrollment compliers. Applicants' baseline scores exceed those of other 8th graders by about  $0.7\text{--}0.8\sigma$ , while the score gap for enrolled students is over twice as large. 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, fewer than 15% of those enrolled in New York exam schools are black or Hispanic.



(b) Distance to next school for 9th grade applicants to Boston exam schools

FIGURE A.1.—Continued.

Table B.II reports estimates of NYC follow-up rates and attrition differentials in a format paralleling that of Table A.II. Scores are most often missing for math since many applicants take Regents exams in these subjects before 9th grade. For the other subjects, the follow-up rates range from 79% to 87%. For instance, 80% of students in at least one discontinuity sample contribute Advanced Math scores, while 87% contribute English scores. Although some of the estimated attrition differentials are significantly different from zero when estimated in the All Schools model, they are mostly small and seem unlikely to be an important source of bias. For Math, Advanced Math, and English scores, for example, the largest estimated differential is a marginally significant 3.6% when estimated using a parametric specification.

As in Boston, the NYC admissions process is run in the central office, suggesting limited scope for school discretion or parent sorting in making assignments. In support of this claim, we document covariate balance for New York exam schools in Table B.III. With the possible exception of the parametric estimates for Stuyvesant reported in column (3), the scattered significant estimates seem likely to be chance findings, a conclusion supported by joint test results reported in the bottom row. Nonparametric estimates for Stuyvesant are small and not significantly different from zero.

TABLE B.I  
 ADDITIONAL DESCRIPTIVE STATISTICS FOR NYC EXAM SCHOOL APPLICANTS<sup>a</sup>

	NYC (1)	Brooklyn Tech			Bronx Science			Stuyvesant		
		Applicants (2)	Enrolled (3)	Compliers (4)	Applicants (5)	Enrolled (6)	Compliers (7)	Applicants (8)	Enrolled (9)	Compliers (10)
Female	0.487	0.497	0.415	0.434	0.501	0.443	0.474	0.505	0.429	0.441
Black	0.336	0.306	0.133	0.181	0.291	0.040	0.042	0.296	0.019	0.022
Hispanic	0.377	0.224	0.089	0.139	0.253	0.070	0.124	0.243	0.030	0.046
Free lunch <sup>b</sup>	0.667	0.685	0.664	0.651	0.677	0.682	0.680	0.683	0.706	0.641
Limited English proficient	0.125	0.037	0.007	-0.002	0.038	0.003	0.002	0.037	0.003	0.002
Special Education	0.089	0.005	0.000	0.000	0.005	0.000	0.000	0.005	0.000	0.000
Baseline Math	-0.004	0.809	1.619	1.350	0.817	1.771	1.601	0.802	2.119	1.892
Baseline English	-0.005	0.734	1.426	1.333	0.755	1.666	1.453	0.730	2.047	1.851

<sup>a</sup>This table reports additional descriptive statistics for 2004–2007. Column (1) reports descriptive statistics for 8th grade students in NYC public schools who had not previously enrolled in any exam school. Columns (2), (5), and (8) report descriptive statistics for applicants to each of the NYC exam schools included in our study. Columns (3), (6), and (9) report descriptive statistics for students who enroll in each exam school included in the study. Columns (4), (7), and (10) report descriptive statistics for enrollment compliers at each admissions cutoff. Baseline math and English scores for applicants are from 8th grade.

<sup>b</sup>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 B.II  
NEW YORK ATTRITION DIFFERENTIALS: REGENTS SCORES<sup>a</sup>

	Fraction With Regents	Parametric Estimates				Nonparametric (DM) Estimates			
		Brooklyn Tech (1)	Bronx Science (2)	Stuyvesant (3)	All Schools (4)	Brooklyn Tech (5)	Bronx Science (6)	Stuyvesant (7)	All Schools (8)
Math	0.535	-0.068 (0.043) 9,181	0.040 (0.045) 8,192	0.042 (0.043) 8,434	0.005 (0.025) 25,807	0.016 (0.026) 7,126	0.016 (0.023) 8,192	0.053* (0.029) 5,757	0.026* (0.015) 21,075
Advanced Math	0.799	0.101*** (0.039) 9,181	0.038 (0.035) 8,192	-0.027 (0.030) 8,434	0.036* (0.020) 25,807	0.074*** (0.024) 6,586	0.022 (0.018) 8,192	0.008 (0.019) 6,141	0.033*** (0.012) 20,919
English	0.871	0.014 (0.035) 7,032	0.054 (0.033) 6,335	0.016 (0.029) 6,309	0.027 (0.018) 19,676	0.026 (0.018) 7,032	0.020 (0.018) 6,335	0.026* (0.015) 6,309	0.024** (0.010) 19,676
Global History	0.862	0.064* (0.034) 9,181	0.052* (0.030) 8,192	0.022 (0.026) 8,434	0.045*** (0.017) 25,807	0.059*** (0.018) 8,407	0.017 (0.015) 8,192	0.019 (0.016) 6,506	0.031*** (0.010) 23,105
U.S. History	0.810	0.029 (0.042) 7,032	0.06 (0.040) 6,335	-0.016 (0.033) 6,309	0.023 (0.022) 19,676	0.044* (0.026) 5,121	0.031 (0.021) 6,335	0.009 (0.021) 4,472	0.029** (0.013) 15,928
Living Environment	0.794	-0.015 (0.037) 9,181	0.041 (0.035) 8,192	0.035 (0.033) 8,434	0.020 (0.020) 25,807	0.032* (0.019) 9,181	0.012 (0.020) 7,017	0.004 (0.016) 8,434	0.016 (0.011) 24,632

<sup>a</sup>This table reports estimates of the effect of exam school offers on indicators for non-missing outcome scores. Models and estimation procedures are the same as used to construct the estimates in Table VIII. The fraction with Regents is the probability a Regents score is observed for applicants who appear in any school-specific discontinuity sample. Robust standard errors, clustered on year and school, are shown in parentheses. Standard errors for the all-schools estimates also cluster on student. Sample sizes are shown below standard errors. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

TABLE B.III  
NEW YORK COVARIATE DISCONTINUITIES<sup>a</sup>

	Mean of Variable	Parametric Estimates				Nonparametric (DM) Estimates			
		Brooklyn Tech (1)	Bronx Science (2)	Stuyvesant (3)	All Schools (4)	Brooklyn Tech (5)	Bronx Science (6)	Stuyvesant (7)	All Schools (8)
Female	0.469	-0.005 (0.044) 9,181	-0.003 (0.045) 8,192	0.012 (0.043) 8,434	0.002 (0.025) 25,807	-0.006 (0.022) 9,181	-0.013 (0.027) 6,768	-0.018 (0.022) 8,434	-0.012 (0.014) 24,383
Black	0.109	-0.046 (0.032) 9,181	-0.002 (0.028) 8,192	0.023 (0.021) 8,434	-0.008 (0.016) 25,807	-0.033* (0.018) 7,699	-0.002 (0.015) 8,192	0.014 (0.011) 7,667	-0.006 (0.009) 23,558
Hispanic	0.107	0.030 (0.031) 9,181	-0.012 (0.028) 8,192	0.055*** (0.020) 8,434	0.025 (0.015) 25,807	0.032** (0.016) 9,181	-0.013 (0.014) 8,192	0.005 (0.011) 7,786	0.008 (0.009) 25,159
Free lunch <sup>b</sup>	0.669	-0.013 (0.041) 9,181	0.058 (0.042) 8,192	-0.091** (0.040) 8,434	-0.018 (0.024) 25,807	-0.008 (0.022) 9,139	0.042* (0.022) 8,192	-0.034 (0.022) 7,944	0.001 (0.013) 25,275
Limited English Proficient	0.005	0.012** (0.006) 9,181	-0.002 (0.004) 8,192	0.001 (0.005) 8,434	0.004 (0.003) 25,807	0.000 (0.003) 8,703	-0.003 (0.003) 6,675	-0.002 (0.002) 8,434	-0.002 (0.002) 23,812
Joint test: <i>p</i> -value		0.228	0.775	0.016	0.393	0.262	0.316	0.384	0.722

<sup>a</sup>This table reports discontinuities in covariates estimated using models like those used to construct the reduced-form estimates in Table VIII. The joint *p*-value in the bottom row is for tests looking at all covariate discontinuities at once. Robust standard errors, clustered on year and school, are shown in parentheses. Standard errors for the all-schools estimates also cluster on student. Sample sizes are shown below standard errors. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

TABLE B.IV  
NEW YORK REDUCED-FORM ESTIMATES FOR SUBGROUPS<sup>a</sup>

	High Baseline Scores								
	Upper Half			Upper Quartile			Black and Hispanic		
	Baseline Mean (1)	Proportion Above 85 on Regents (2)	IK Estimates (3)	Baseline mean (4)	Proportion Above 85 on Regents (5)	IK Estimates (6)	Baseline Mean (7)	Proportion Above 85 on Regents (8)	IK Estimates (9)
Math	1.422	0.869	-0.055*** (0.018) 10,523	1.493	0.925	-0.035* (0.019) 9,138	1.243	0.711	-0.084*** (0.023) 6,189
Advanced Math	0.932	0.567	-0.029 (0.023) 16,168	1.024	0.635	-0.006 (0.020) 14,612	0.59	0.341	-0.048 (0.031) 8,713
English	1.092	0.812	0.011 (0.013) 15,450	1.157	0.866	0.012 (0.017) 10,254	0.969	0.711	0.031 (0.022) 7,146
Global History	1.231	0.837	-0.016 (0.014) 17,569	1.28	0.875	-0.019 (0.014) 13,422	1.076	0.712	0.002 (0.019) 9,856
U.S. History	1.147	0.934	0.003 (0.015) 11,828	1.183	0.953	0.000 (0.013) 9,331	1.029	0.866	-0.001 (0.016) 6,171
Living Environment	1.33	0.706	-0.021* (0.012) 18,928	1.376	0.751	-0.019 (0.014) 15,343	1.175	0.559	-0.025 (0.018) 8,866

<sup>a</sup>This table reports nonparametric reduced-form estimates of the all-schools model for students with high baseline scores and for minorities. Baseline means and the proportion of applicants above 85 are computed for those who belong to at least one discontinuity sample. Math scores are from either Regents Math A (Elementary Algebra and Planar Geometry) or Integrated Algebra I. Advanced Math scores are from either Regents Math B (Intermediate Algebra and Trigonometry) or Geometry. Robust standard errors, clustered on year and school, are shown in parentheses. Standard errors also cluster on student. Sample sizes are shown below standard errors. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

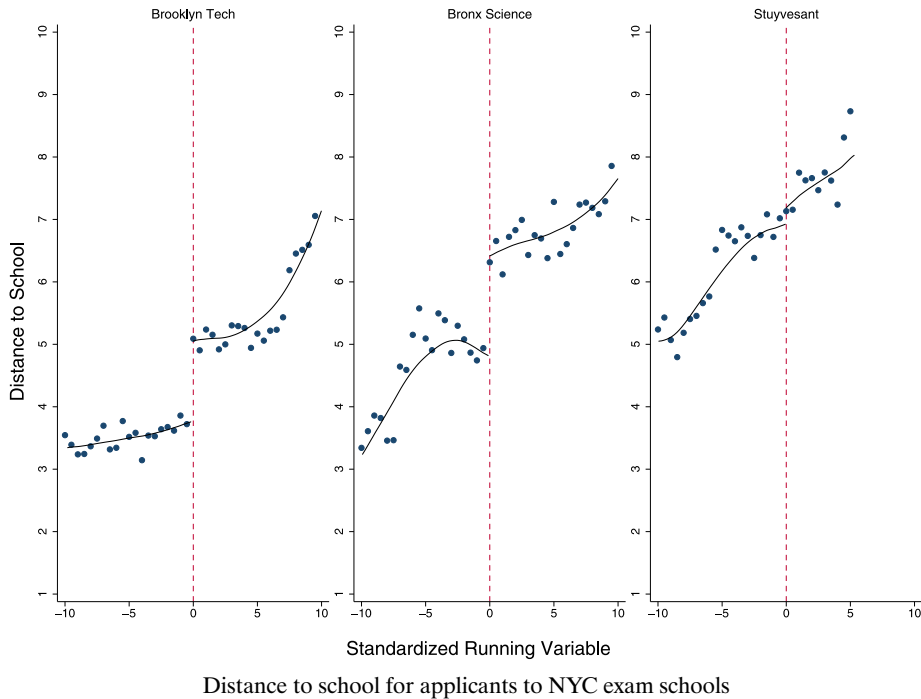


FIGURE B.1.—This figure shows the average distance in miles between students' home addresses and the school they attend in the year after application for applicants to NYC exam schools, plotted against school-specific standardized running variables.

By contrast with Boston, detailed results for black and Hispanic NYC exam school applicants provide little in the way of evidence for minority applicants gain at exam schools. These results can be seen in Table B.IV. For instance, the effect on English for black or Hispanic students is  $0.03\sigma$  (with standard error 0.02). On the other hand, consistent with the Boston findings, Table B.IV shows little in the way of gains for high baseline achievers in New York.

Figure B.1 reports on travel distance for NYC applicants on either side of exam school admissions cutoffs. Students offered a seat at Bronx Science and Brooklyn Tech travel, on average, 1 mile further, while travel appears to be unchanged by offers from Stuyvesant.

#### REFERENCES

- ABADIE, A. (2003): "Semiparametric Instrumental Variables Estimation of Treatment Response Models," *Journal of Econometrics*, 113 (2), 231–263. [148]
- ABDULKADIROĞLU, A., J. ANGRIST, AND P. PATHAK (2011): "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools," Working Paper 17264, NBER. [175]
- (2014): "Supplement to 'The Elite Illusion: Achievement Effects at Boston and New York Exam Schools,'" *Econometrica Supplemental Material*, 82, <http://www.econometricsociety>.

- org/ecta/supmat/10266\_data\_description.pdf; [http://www.econometricsociety.org/ecta/supmat/10266\\_data\\_and\\_programs.zip](http://www.econometricsociety.org/ecta/supmat/10266_data_and_programs.zip). [161]
- 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. [143,165]
- ACEMOGLU, D., AND J. ANGRIST (2000): "How Large Are the Social Returns to Education: Evidence From Compulsory Schooling Laws," in *NBER Macroeconomic Annual*, ed. by B. Bernanke and K. Rogoff. Chicago: University of Chicago Press, 9–59. [178]
- AMMERMUELLER, A., AND J.-S. PISCHKE (2009): "Peer Effects in European Primary Schools: Evidence From the Progress in International Reading Literacy Study," *Journal of Labor Economics*, 27 (3), 315–348. [141]
- 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. [155]
- ANGRIST, J. D., AND A. KRUEGER (1992): "Estimating the Payoff to Schooling Using the Vietnam-Era Draft Lottery," Working Paper 4067, NBER. [174]
- 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. [138,141]
- ANGRIST, J. D., AND J.-S. PISCHKE (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press. [178]
- ANGRIST, J. D., AND M. ROKKANEN (2012): "Wanna Get Away? RD Identification Away From the Cutoff," Working Paper 18662, NBER. [179]
- BLACK, S. (1999): "Do Better Schools Matter? Parental Valuation of Elementary Education," *Quarterly Journal of Economics*, 114 (2), 577–599. [137,138]
- BOUSTAN, L. P. (2012): "School Desegregation and Urban Change: Evidence From City Boundaries," *American Economic Journal: Applied Economics*. [138]
- CARRELL, S., B. SACERDOTE, AND J. WEST (2012): "From Natural Variation to Optimal Policy? An Unsuccessful Experiment in Using Peer Effects Estimates to Improve Student Outcomes," Unpublished Paper, UC Davis. [141]
- CLARK, D. (2008): "Selective Schools and Academic Achievement," Discussion Paper 3182, IZA. [141]
- 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. [141]
- (2011): "Estimating the Return to College Selectivity Over the Career Using Administrative Earnings Data," Industrial Relations Working Paper 563, Princeton University. [141]
- DESJARDINS, S., AND B. MCCALL (2008): "The Impact of Gates Millennium Scholars Program on the Retention, College Finance- and Work-Related Choices, and Future Educational Aspirations of Low-Income Minority Students," Unpublished Manuscript, University of Michigan. [148,159]
- DOBBIE, W., AND R. G. FRYER (2013): "Exam High Schools and Academic Achievement: Evidence From New York City," *American Economic Journal: Applied Economics* (forthcoming). [140]
- 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. [141]
- GOLDIN, C., AND L. F. KATZ (2008): *The Race Between Education and Technology*. Cambridge: Harvard University Press. [139]
- GURYAN, J. (2004): "Desegregation and Black Dropout Rates," *American Economic Review*, 94 (4), 919–943. [138]
- HAHN, J., P. TODD, AND W. VAN DER KLAUW (2001): "Identification and Estimation of Treatment Effects With a Regression Discontinuity Design," *Econometrica*, 69 (1), 201–209. [148]
- HANUSHEK, E. A., J. F. KAIN, J. M. MARKMAN, AND S. G. RIVKIN (2003): "Does Peer Ability Affect Student Achievement?" *Journal of Applied Econometrics*, 18 (5), 527–544. [141,175]

- HASTINGS, J., T. J. KANE, AND D. O. STAIGER (2009): "Heterogeneous Preferences and the Efficacy of Public School Choice," Working Paper, Yale University. [138,179]
- HERNANDEZ, J. (2008): "Racial Imbalance Persists at Elite Public Schools," *New York Times*, N.Y./Region, November 7. [168]
- 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. [141]
- HOXBY, C. (2000): "Peer Effects in the Classroom: Learning From Gender and Race Variation," Working Paper 7867, NBER. [138,141,175]
- 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>. [141]
- 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. [179]
- IMBENS, G., AND K. KALYANARAMAN (2012): "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 79 (3), 933–959. [148,159]
- IMBERMAN, S. A., A. KUGLER, AND B. SACERDOTE (2012): "Katrina's Children: Evidence on the Structure of Peer Effects From Hurricane Evacuees," *American Economic Review*, 102 (5), 2048–2082. [141]
- 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. [140]
- JAN, T. (2006): "Growing a Boston Latin in Brooklyn," *Boston Globe*, Local Desk, March 4. [139]
- KANE, T. J., AND D. O. STAIGER (2002): "The Promise and Pitfalls of Using Imprecise School Accountability Measures," *Journal of Economic Perspectives*, 16 (4), 91–114. [179]
- KLING, J. R., J. B. LIEBMAN, AND L. F. KATZ (2007): "Experimental Analysis of Neighborhood Effects," *Econometrica*, 75 (1), 83–119. [138,179]
- LAVY, V., O. SILVA, AND F. WEINHARDT (2012): "The Good, The Bad, and the Average: Evidence on Ability Peer Effects in Schools," *Journal of Labor Economics*, 30 (2), 367–414. [141]
- 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), 807–859. [148]
- LEE, D. S. (2009): "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *Review of Economic Studies*, 76 (3), 1071–1102. [156]
- LEE, D. S., AND T. LEMIEUX (2010): "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48, 281–355. [152,158,179]
- LUTTON, L. (2012): "Chicago Will Open More Selective High Schools," *WBEZ*, Schools on the Line, April 5. [168]
- MACLEOD, B., AND M. URQUIOLA (2009): "Anti-Lemons: School Reputation and Educational Quality," Working Paper 15112, NBER. [138,179]
- (2000): "Economic Analysis of Social Interactions," *Journal of Economic Perspectives*, 14 (3), 115–136. [178]
- MOFFITT, R. A. (2001): "Policy Interventions, Low-Level Equilibria and Social Interactions," in *Social Dynamics*. Cambridge: MIT Press, 45–82. [178]
- PATHAK, P. A., AND T. SÓNMEZ (2008): "Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism," *American Economic Review*, 98 (4), 1636–1652. [143]
- (2013): "School Admissions Reform in Chicago and England: Comparing Mechanisms by Their Vulnerability to Manipulation," *American Economic Review*, 103 (1), 80–106. [143]
- POP-ELECHES, C., AND M. URQUIOLA (2013): "Going to a Better School: Effects and Behavioral Responses," *American Economic Review*, 103 (4), 1289–1324. [140,141]
- PORTER, J. (2003): "Estimation in the Regression Discontinuity Model," Working Paper, University of Wisconsin. [148]
- ROKKANEN, M. (2013): "Exam Schools, Ability, and the Effects of Affirmative Action: Latent Factor Extrapolation in the Regression Discontinuity Design," Working Paper, MIT. [179]

- 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. [138,179]
- SACERDOTE, B. (2011): “Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far,” in *Handbook of Economics of Education*, ed. by E. A. Hanushek, S. Machin, and L. Woessmann. Amsterdam: Elsevier, 249–277. [139,175]
- SANBONMATSU, L., J. LUDWIG, L. KATZ, L. A. GENNETIAN, G. J. DUNCAN, R. C. KESSLER, E. A. T. W. MCDADE, AND S. T. LINDAU (2011): “Moving to Opportunity for Fair Housing Demonstration Program—Final Impacts Evaluation,” U.S. Department of Housing & Urban Development, PD&R. [138]
- TODD, P., AND K. WOLPIN (2003): “On the Specification and Estimation of the Production Function for Cognitive Achievement,” *Economic Journal*, 113, F3–F33. [173]
- VAZNIS, J. (2009): “Exam School Explore Adding Grade 6,” Boston Globe, Education Desk, June 9. [168]
- ZHANG, H. (2010): “Magnet Schools and Student Achievement: Evidence From a Randomized Natural Experiment in China,” Unpublished Paper, Chinese University of Hong Kong. [141]

*Dept. of Economics, Duke University, Durham, NC 27708, U.S.A.; atila.abdulkadiroglu@duke.edu,*

*Dept. of Economics, MIT, 77 Massachusetts Ave., Cambridge, MA 02142, U.S.A. and NBER; angrist@mit.edu,*

*and*

*Dept. of Economics, MIT, 77 Massachusetts Ave., Cambridge, MA 02142, U.S.A. and NBER; ppathak@mit.edu.*

*Manuscript received August, 2011; final revision received May, 2013.*