American Economic Review 2020, 110(5): 1502–1539
https://doi.org/10.1257/aer.20172040

Do Parents Value School Effectiveness?†

By Atila Abdulkadiroğlu, Parag A. Pathak, Jonathan Schellenberg, and Christopher R. Walters*

School choice may lead to improvements in school productivity if parents' choices reward effective schools and punish ineffective ones. This mechanism requires parents to choose schools based on causal effectiveness rather than peer characteristics. We study relationships among parent preferences, peer quality, and causal effects on outcomes for applicants to New York City's centralized high school assignment mechanism. We use applicants' rank-ordered choice lists to measure preferences and to construct selection-corrected estimates of treatment effects on test scores, high school graduation, college attendance, and college quality. Parents prefer schools that enroll high-achieving peers, and these schools generate larger improvements in short- and long-run student outcomes. Preferences are unrelated to school effectiveness and academic match quality after controlling for peer quality. (*JEL* D12, H75, I21, I26, I28)

Recent education reforms in the United States, including charter schools, school vouchers, and district-wide open enrollment plans, increase parents' power to choose schools for their children. School choice allows households to avoid undesirable schools and forces schools to satisfy parents' preferences or risk losing enrollment. Proponents of choice argue that this competitive pressure is likely to generate system-wide increases in school productivity and boost educational outcomes for students (Friedman 1962, Chubb and Moe 1990, Hoxby 2003). By decentralizing school quality assessment and allowing parents to act on local information, school choice is conjectured to provide better incentives for educational effectiveness than

*Abdulkadiroğlu: Duke University, and NBER (email: aa88@duke.edu); Pathak: Massachusetts Institute of Technology, and NBER (email: ppathak@mit.edu); Schellenberg: University of California, Berkeley (email: jschellenberg@econ.berkeley.edu); Walters: University of California, Berkeley, and NBER (email: crwalters@econ.berkeley.edu). Thomas Lemieux was the coeditor for this article. We gratefully acknowledge funding from the National Science Foundation and the W. T. Grant Foundation. We also thank Nikhil Agarwal, Josh Angrist, Stephane Bonhomme, David Card, David Chan, Michael Dinerstein, Will Dobbie, James Heckman, Peter Hull, Pat Kline, Thibaut Lamadon, Magne Mogstad, Jack Mountjoy, Derek Neal, Ariel Pakes, Stephen Raudenbush, Jesse Rothstein, Alex Torgovitsky, Miguel Urquiola, Jeffrey Wooldridge, Danny Yagan, Seth Zimmerman, and seminar participants at the University of Chicago Committee on Education Workshop, the University of Chicago Harris School, UC Berkeley, the University of Chicago Econometrics Lunch, the University of Chicago Interactions Conference, UW Madison, the 2017 NBER Labor Studies fall meetings, the University of Virginia, the University of Hawaii, the University of Rochester, Princeton University, the University of Maryland, and Arizona State University for suggestions and comments. We're grateful to Ray Han and Kate Bradley for outstanding research assistance, and to Eryn Heying for invaluable administrative support.
†Go to https://doi.org/10.1257/aer.20172040 to visit the article page for additional materials and author disclosure statements.

could be achieved by a centralized accountability system (Peterson and Campbell 2001). Choice may also improve outcomes by allowing students to sort into schools that suit their particular educational needs, resulting in improved match quality (Hoxby 2000). These arguments have motivated recent policy efforts to expand school choice (e.g., DeVos 2017).

If choice is to improve educational effectiveness, parents' choices must result in rewards for effective schools and sanctions for ineffective ones. Our use of the term "effective" follows Rothstein (2006): an effective school is one that generates causal improvements in student outcomes. Choice need not improve school effectiveness if it is not the basis for how parents choose between schools. For example, parents may value attributes such as facilities, convenience, student satisfaction, or peer composition in a manner that does not align with educational impacts (Hanushek 1981, Jacob and Lefgren 2007). Moreover, while models in which parents value schools according to their effectiveness are an important benchmark in the academic literature (e.g., Epple, Figlio, and Romano 2004), it may be difficult for parents to separate a school's effectiveness from the composition of its student body (Kane and Staiger 2002). If parent choices reward schools that recruit higher-achieving students rather than schools that improve outcomes, school choice may increase resources devoted to screening and selection rather than better instruction (Ladd 2002, MacLeod and Urquiola 2015). Consistent with these possibilities, Rothstein (2006) shows that cross-district relationships among school choice, sorting patterns, and student outcomes fail to match the predictions of a model in which school effectiveness is the primary determinant of parent preferences.

This paper offers new evidence on the links between preferences, school effectiveness, and peer quality based on choice and outcome data for more than 250,000 applicants in New York City's centralized high school assignment mechanism. Each year, thousands of New York City high school applicants rank-order schools, and the mechanism assigns students to schools using the deferred acceptance (DA) algorithm (Gale and Shapley 1962; Abdulkadiroğlu, Pathak, and Roth 2005). The DA mechanism is strategy-proof: truthfully ranking schools is a weakly dominant strategy for students (Dubins and Freedman 1981, Roth 1982). This fact motivates our assumption that applicants' rankings measure their true preferences for schools.[1] We summarize these preferences by fitting discrete choice models to applicants' rank-ordered preference lists.

We then combine the preference estimates with estimates of school treatment effects on test scores, high school graduation, college attendance, and college choice. Treatment effect estimates come from "value-added" regression models of the sort commonly used to measure causal effects of teachers and schools (Todd and Wolpin 2003; Koedel, Mihaly, and Rockoff 2015). We generalize the conventional value-added approach to allow for match effects in academic outcomes and to relax the selection-on-observables assumption underlying standard models. Recent evidence suggests that value-added models controlling only for observables provide

[1] As we discuss in Section I, DA is strategy-proof when students are allowed to rank every school, but the New York City mechanism only allows applicants to rank 12 choices. Most students do not fill their preference lists, however, and truthful ranking is a dominant strategy in this situation (Haeringer and Klijn 2009, Pathak and Sönmez 2013). Fack, Grenet, and He (2015) proposes empirical approaches to measuring student preferences without requiring that truth-telling is the unique equilibrium.

quantitatively useful but biased estimates of causal effects due to selection on unobservables (Rothstein 2010, 2017; Chetty, Friedman, and Rockoff 2014a; Angrist et al. 2017). We therefore use the rich information on preferences contained in students' rank-ordered choice lists to correct our estimates for selection on unobservables. This selection correction is implemented by extending the classic multinomial logit control function estimator of Dubin and McFadden (1984) to a setting where rankings of multiple alternatives are known. We show that predictions from our models match the effects of randomized lottery assignment for a subset of schools where lottery quasi-experiments are available, suggesting that our methods provide accurate measures of causal effects.

The final step of our analysis relates the choice model and treatment effect estimates to measure preferences for school effectiveness. The choice and outcome models we estimate allow preferences and causal effects to vary flexibly with student characteristics. Our specifications accommodate the possibility that schools are more effective for specific types of students and that applicants choose schools that are a good match for their student type. We compare the degree to which parent preferences are explained by overall school effectiveness, match quality, and peer quality, defined as the component of a school's average outcome due to selection rather than effectiveness. We explore these relationships for test scores as well as longer-run postsecondary outcomes, which is important in view of recent evidence that school quality is multidimensional and only imperfectly measured by effects on test scores (Beuermann et al. 2018).

We find preferences are positively correlated with both peer quality and causal effects on student outcomes. More effective schools enroll higher-ability students, however, and preferences are unrelated to school effectiveness after controlling for peer quality. We also find little evidence of selection on match effects: on balance, parents do not prefer schools that are especially effective for their own children, and students do not enroll in schools that are a better-than-average match. These patterns are similar for short-run achievement test scores and longer-run postsecondary outcomes. Looking across demographic and baseline achievement groups, we find no evidence that any subgroup places positive weight on school effectiveness once we adjust for peer quality. Our estimates are also similar across applicant cohorts, suggesting that the relationship between demand and effectiveness is stable over time.

The factors driving school popularity we uncover are noteworthy, but to translate them into implications about the incentives that schools face from demand-side forces requires isolating the causal impacts of school attributes on preferences. Since effectiveness and peer quality are not randomly assigned to schools, our estimates need not capture causal effects of these attributes on preferences if other school characteristics that influence demand are correlated with effectiveness or peer quality. We assess the potential for such omitted variables bias by conditioning on other school characteristics that predict demand, including measures of violent incidents, teacher education, and the school learning environment. This analysis reveals that parents prefer safer schools and schools with more educated teachers, but adding these covariates does not alter our main results characterizing the partial correlations between preferences, peer quality, and school effectiveness. This robustness exercise provides some reassurance that our estimates capture causal impacts of effectiveness on demand, though we cannot completely rule out the possibility that peer quality and school effectiveness are correlated with other unobservables.

It is worth cautioning that our findings do not mean that parents choose schools irrationally; they may use peer characteristics to proxy for school effectiveness if the latter is difficult to observe, or value peer quality independently of impacts on academic outcomes. Either way, our results imply that parents' choices penalize schools that enroll low achievers rather than schools that offer poor instruction. As a result, school choice programs may generate stronger incentives for screening and selection than for improved academic quality. We provide suggestive evidence that schools have responded to these incentives by increasing screening in the years following the introduction of centralized assignment in New York City.

Our analysis complements Rothstein's (2006) indirect test with a direct assessment of the relationships among parent preferences, peer quality, and school effectiveness based on unusually rich choice and outcome data. The results also contribute to a large literature studying preferences for school quality (Black 1999; Figlio and Lucas 2004; Bayer, Ferreira, and McMillan 2007; Hastings and Weinstein 2008; Burgess et al. 2015; Imberman and Lovenheim 2016). These studies show that housing prices and household choices respond to school performance levels, but they do not typically separate responses to causal school effectiveness and peer quality. Our findings are also relevant to theoretical and empirical research on the implications of school choice for sorting and stratification (Epple and Romano 1998; Epple, Figlio, and Romano 2004; Hsieh and Urquiola 2006; Barseghyan, Clark, and Coate 2014; Altonji, Huang, and Taber 2015; Avery and Pathak 2015; MacLeod and Urquiola 2015; MacLeod et al. 2017). In addition, our results help to reconcile some surprising findings from recent studies of school choice. Cullen, Jacob, and Levitt (2006) finds limited achievement effects of admission to preferred schools in Chicago, while Walters (2018) documents that disadvantaged students in Boston are less likely to apply to charter schools than more advantaged students despite experiencing larger achievement benefits. Angrist, Pathak, and Walters (2013) and Abdulkadiroğlu, Pathak, and Walters (2018) report on two settings where parents opt for schools that reduce student achievement. These patterns are consistent with our finding that school choices are not driven by school effectiveness.

Finally, our analysis adds to a recent series of studies leveraging preference data from centralized school assignment mechanisms to investigate school demand (Hastings, Kane, and Staiger 2009; Harris and Larsen 2014; Fack, Grenet, and He 2015; Abdulkadiroğlu, Agarwal, and Pathak 2017; Glazerman and Dotter 2016; Kapor, Neilson, and Zimmerman 2017; Agarwal and Somaini 2018). Some of these studies analyze assignment mechanisms that provide incentives to strategically misreport preferences, while others measure academic quality using average test scores rather than distinguishing between peer quality and school effectiveness or looking at longer-run outcomes. We build on this previous work by using data from a strategy-proof mechanism to separately estimate preferences for peer quality and causal effects on multiple measures of academic performance.

The rest of the paper is organized as follows. The next section describes school choice in New York City and the data used for our analysis. Section II develops a conceptual framework for analyzing school effectiveness and peer quality, and Section III details our empirical approach. Section IV summarizes estimated distributions of student preferences and school treatment effects. Section V links preferences to peer quality and school effectiveness, and Section VI discusses implications of these relationships. Section VII concludes and offers some directions for future research.

I. Setting and Data

A. *New York City High Schools*

The New York City public school district annually enrolls roughly 90,000 ninth graders at more than 400 high schools. Rising ninth graders planning to attend New York City's public high schools submit applications to the centralized assignment system. Before 2003 the district used an uncoordinated school assignment process in which students could receive offers from more than one school. Motivated in part by insights derived from the theory of market design, in 2003 the city adopted a coordinated single-offer assignment mechanism based on the student-proposing deferred acceptance (DA) algorithm (Gale and Shapley 1962; Abdulkadiroğlu, Pathak, and Roth 2005, 2009). Abdulkadiroğlu, Agarwal, and Pathak (2017) shows that introducing coordinated assignment reduced the share of administratively assigned students and likely improved average student welfare.

Applicants report their preferences for schooling options to the assignment mechanism by submitting rank-ordered lists of up to 12 academic programs. An individual school may operate more than one program. To aid families in their decision-making, the New York City Department of Education (DOE) distributes a directory that provides an overview of the high school admission process, key dates, and an information page for each high school. A school's information page includes a brief statement of its mission, a list of offered programs, courses and extracurricular activities, pass rates on New York Regents standardized tests, and the school's graduation rate (New York City Department of Education 2003). DOE also issues annual schools reports that list basic demographics, teacher characteristics, school expenditures, and Regents performance levels. During the time period of our study (2003–2007) these reports did not include measures of test score growth, though such measures have been added more recently (New York City Department of Education 2004, 2017).

Academic programs prioritize applicants in the centralized admission system using a mix of factors. Priorities depend on whether a program is classified as unscreened, screened, or an educational option program. Unscreened programs give priority to students based on residential zones and (in some cases) to those who attend an information session. Screened programs use these factors and may also assign priorities based on prior grades, standardized test scores, and attendance. Educational option programs use screened criteria for some of their seats and unscreened criteria for the rest. Random numbers are used to order applicants with equal priority. A small group of selective high schools, including New York City's exam schools, admit students in a parallel system outside the main round of the assignment process (Abdulkadiroğlu, Angrist, and Pathak 2014).

The DA algorithm combines student preferences with program priorities to generate a single program assignment for each student. In the initial step of the algorithm, each student proposes to her first-choice program. Programs provisionally accept students in order of priority up to capacity and reject the rest. In subsequent rounds, each student rejected in the previous step proposes to her most-preferred program among those that have not previously rejected her, and programs reject provisionally accepted applicants in favor of new applicants with higher priority. This process iterates until all students are assigned to a program or all unassigned students have been rejected by every program they have ranked. During our study time period, students left unassigned in the main round participate in a supplementary DA round in which they rank up to 12 additional programs with available seats. Any remaining students are administratively assigned by the district. About 82 percent, 8 percent, and 10 percent of applicants are assigned in the main, supplementary, and administrative rounds, respectively (Abdulkadiroğlu, Agarwal, and Pathak 2017).

An attractive theoretical property of the DA mechanism is that it is strategy-proof: since high-priority students can displace those with lower priority in later rounds of the process, listing schools in order of true preferences is a dominant strategy in the mechanism's canonical version. This property, however, requires students to have the option to rank all schools (Haeringer and Klijn 2009, Pathak and Sönmez 2013). As we show below, more than 70 percent of students rank fewer than 12 programs, meaning that truthful ranking of schools is a dominant strategy for the majority of applicants. The instructions provided with the New York City high school application also directly instruct students to rank schools in order of their true preferences (New York City Department of Education 2003). In the analysis to follow, we interpret students' rank-ordered lists as truthful reports of their preferences. We also probe the robustness of our findings to violations of this assumption by reporting results based on students who rank fewer than 12 choices.[2]

B. *Data and Samples*

The data used here are extracted from a DOE administrative information system covering all students enrolled in New York City public schools between the 2003–2004 and 2012–2013 school years (New York City Department of Education 2013). These data include school enrollment, student demographics, home addresses, scores on New York Regents standardized tests, Preliminary SAT (PSAT) scores, and high school graduation records, along with preferences submitted to the centralized high school assignment mechanism. A supplemental file from the National Student Clearinghouse (NSC) reports college enrollment for students graduating from New York City high schools between 2009 and 2012. A unique student identifier links records across these files.

We analyze high school applications and outcomes for four cohorts of students enrolled in New York City public schools in eighth grade between 2003–2004 and 2006–2007. This set of students is used to construct several samples for statistical analysis. The choice sample, used to investigate preferences for schools, consists of all high school applicants with baseline (eighth grade) demographic, test score, and address information. Our analysis of school effectiveness uses subsamples of the choice sample corresponding to each outcome of interest. These outcome

[2] Along similar lines, Abdulkadiroğlu, Agarwal, and Pathak (2017) shows that preference estimates using only the top-ranked school, the top three schools, and all but the last ranked school are similar.

samples include students with observed outcomes, baseline scores, demographics, and addresses, enrolled for ninth grade at one of 316 schools with at least 50 students for each outcome. The outcome samples also exclude students enrolled at the nine selective high schools that do not admit students via the main DA mechanism. Online Appendix Section A and Appendix Table A1 provide further details on data sources and sample construction.

Key outcomes in our analysis include Regents math standardized test scores, PSAT scores, high school graduation, college attendance, and college quality. The high school graduation outcome equals 1 if a student graduates within five years of her projected high school entry date given her eighth grade cohort. Likewise, college attendance equals 1 for students who enroll in any college (two or four year) within two years of projected on-time high school graduation. The college quality variable, derived from Internal Revenue Service tax record statistics reported by Chetty et al. (2017), equals the mean 2014 income for children born between 1980 and 1982 who attended a student's college. The mean income for the non-college population is assigned to students who do not enroll in a college. While this metric does not distinguish between student quality and causal college effectiveness, it provides a measure of the selectivity of a student's college. It has also been used elsewhere to assess effects of education programs on the intensive margin of college attendance (Chetty et al. 2011; Chetty, Friedman, and Rockoff 2014b). College attendance and quality are unavailable for the 2003–2004 cohort because the NSC data window does not allow us to determine whether students in this cohort were enrolled in college within two years of projected high school graduation.

Descriptive statistics for the choice and outcome samples appear in Table 1. These statistics show that New York City schools serve a disadvantaged urban population. Seventy-three percent of students are black or Hispanic, and 65 percent are eligible for a subsidized lunch. Data from the 2011–2015 American Community Surveys show that the average student in the choice sample lives in a census tract with a median household income of $50,136 in 2015 dollars. Observed characteristics are generally similar for students in the choice and outcome samples. The average PSAT score in New York City is 116, about 1 standard deviation below the US average (the PSAT is measured on a 240-point scale, normed to have a mean of 150 and a standard deviation of 30). The five-year high school graduation rate is 61 percent, and 48 percent of students attend some college within two years of graduation.

C. *Choice Lists*

New York City high school applicants tend to prefer schools near their homes, and most do not fill their choice lists. These facts are shown in Table 2, which summarizes rank-ordered preference lists in the choice sample. As shown in column 1, 93 percent of applicants submit a second choice, about one-half submit 8 or more choices, and 28 percent submit the maximum 12 allowed choices. Column 2 shows that students prefer schools located in their home boroughs: 85 percent of first-choice schools are in the same borough as the student's home address, and the fraction of other choices in the home borough are also high. Abdulkadiroğlu, Agarwal, and Pathak (2017) reports that for 2003–2004, 193 programs restricted eligibility to applicants who reside in the same borough. The preference analysis

Table 1—Descriptive Statistics for New York City Eighth Graders

		Outcome samples			
	Choice sample	Regents math	PSAT	HS graduation	College
	(1)	(2)	(3)	(4)	(5)
Female	0.497	0.518	0.532	0.500	0.500
Black	0.353	0.377	0.359	0.376	0.372
Hispanic	0.381	0.388	0.384	0.399	0.403
Subsidized lunch	0.654	0.674	0.667	0.680	0.700
Census tract median income	$50,136	$50,004	$49,993	$49,318	$49,243
Bronx	0.231	0.221	0.226	0.236	0.239
Brooklyn	0.327	0.317	0.335	0.339	0.333
Manhattan	0.118	0.118	0.119	0.116	0.116
Queens	0.259	0.281	0.255	0.250	0.253
Staten Island	0.065	0.063	0.064	0.059	0.059
Regents math score	0.000	−0.068	0.044	−0.068	−0.044
PSAT score	120	116	116	116	115
High school graduation	0.587	0.763	0.789	0.610	0.624
Attended college	0.463	0.588	0.616	0.478	0.478
College quality	$31,974	$33,934	$35,010	$31,454	$31,454
Observations	270,157	155,850	149,365	230,087	173,254

Notes: This table shows descriptive statistics for applicants to New York City public high schools between the 2003–2004 and 2006–2007 school years. Column 1 reports average characteristics and outcomes for all applicants with complete information on preferences, demographics, and eighth-grade test scores. Columns 2–5 display characteristics for the Regents math, PSAT, high school graduation, and college outcome samples. Outcome samples are restricted to students with data on the relevant outcome, enrolled in ninth grade at schools with at least 50 students for each outcome. Regents math scores are normalized to mean 0 and standard deviation 1 in the choice sample. High school graduation equals 1 for students who graduate from a New York City high school within five years of the end of their eighth grade year. College attendance equals 1 for students enrolled in any college within two years of projected high school graduation. College quality is the mean 2014 income for individuals in the 1980–1982 birth cohorts who attended a student's college. This variable equals the mean income in the non-college population for students who did not attend college. The college outcome sample excludes students in the 2003–2004 cohort. Census tract median income is median household income measured in 2015 dollars using data from the 2011–2015 American Community Surveys. Regents math, PSAT, graduation, and college outcome statistics exclude students with missing values.

Table 2—Correlates of Preference Rankings for New York City High Schools

	Fraction reporting	Same borough	Distance	Regents math score
	(1)	(2)	(3)	(4)
Choice 1	1.000	0.849	2.71	0.200
Choice 2	0.929	0.844	2.94	0.149
Choice 3	0.885	0.839	3.04	0.116
Choice 4	0.825	0.828	3.12	0.085
Choice 5	0.754	0.816	3.18	0.057
Choice 6	0.676	0.803	3.23	0.030
Choice 7	0.594	0.791	3.28	0.009
Choice 8	0.523	0.780	3.29	−0.013
Choice 9	0.458	0.775	3.31	−0.031
Choice 10	0.402	0.773	3.32	−0.051
Choice 11	0.345	0.774	3.26	−0.071
Choice 12	0.278	0.787	3.04	−0.107

Notes: This table reports average characteristics of New York City high schools by student preference rank. Column 1 displays fractions of student applications listing each choice. Column 2 reports the fraction of listed schools located in the same borough as a student's home address. Column 3 reports the mean distance between a student's home address and each ranked school, measured in miles. This column excludes schools outside the home borough. Column 4 shows average Regents math scores in standard deviation units relative to the New York City average.

to follow, therefore, treats schools in a student's home borough as her choice set and aggregates schools in other boroughs into a single outside option. Column 3, which reports average distances (measured as great-circle distance in miles) for each choice restricted to schools in the home borough, shows that students rank nearby schools higher within boroughs as well.

Applicants also prefer schools with strong academic performance. The last column of Table 2 reports the average Regents high school math score for schools at each position on the rank list. Regents scores are normalized to have mean 0 and standard deviation 1 in the New York City population. To earn a high school diploma in New York state, students must pass a Regents math exam. These results reveal that higher-ranked schools enroll students with better math scores. The average score at a first-choice school is 0.2 standard deviations (σ) above the city average, and average scores monotonically decline with rank. PSAT, graduation, college enrollment, and college quality indicators also decline with rank. Students and parents clearly prefer schools with high achievement levels. Our objective in the remainder of this paper is to decompose this pattern into components due to preferences for school effectiveness and peer quality.

II. Conceptual Framework

Consider a population of students indexed by i, each of whom attends one of J schools. Let Y_{ij} denote the potential value of some outcome of interest for student i if she attends school j. The projection of Y_{ij} on a vector of observed characteristics, X_i, is written as

$$Y_{ij} = \alpha_j + X_i'\beta_j + \epsilon_{ij}, \tag{1}$$

where $E[\epsilon_{ij}] = E[X_i\epsilon_{ij}] = 0$ by definition of α_j and β_j. The coefficient vector β_j measures the returns to observed student characteristics at school j, while ϵ_{ij} reflects variation in potential outcomes unexplained by these characteristics. We further normalize $E[X_i] = 0$, so $\alpha_j = E[Y_{ij}]$ is the population mean potential outcome at school j. The realized outcome for student i is $Y_i = \sum_j \mathbf{1}\{S_i = j\} Y_{ij}$, where $S_i \in \{1, \ldots, J\}$ denotes school attendance.

We decompose potential outcomes into components explained by student ability, school effectiveness, and idiosyncratic factors. Let $A_i = (1/J)\sum_j Y_{ij}$ denote student i's general ability, defined as the average of her potential outcomes across all schools. This variable describes how the student would perform at the average school. Adding and subtracting A_i on the right-hand side of (1) yields

$$Y_{ij} = \underbrace{\bar{\alpha} + X_i'\bar{\beta} + \bar{\epsilon}_i}_{A_i} + \underbrace{(\alpha_j - \bar{\alpha})}_{ATE_j} + \underbrace{X_i'(\beta_j - \bar{\beta}) + (\epsilon_{ij} - \bar{\epsilon}_i)}_{M_{ij}}, \tag{2}$$

where $\bar{\alpha} = (1/J)\sum_j \alpha_j$, $\bar{\beta} = (1/J)\sum_j \beta_j$, and $\bar{\epsilon}_i = (1/J)\sum_j \epsilon_{ij}$. Equation (2) shows that student i's potential outcome at school j is the sum of three terms: the student's general ability, A_i; the school's average treatment effect, ATE_j, defined as the causal effect of school j relative to an average school for an average student; and a match effect, M_{ij}, which reflects student i's idiosyncratic suitability for school j. Match effects may arise either because of an interaction between student i's observed characteristics and the extra returns to characteristics at school j (captured by $X_i'(\beta_j - \overline{\beta})$) or because of unobserved factors that make student i more or less suitable for school j (captured by $\epsilon_{ij} - \overline{\epsilon}_i$).

This decomposition allows us to interpret variation in observed outcomes across schools using three terms. The average outcome at school j is given by

$$(3) \qquad E[Y_i | S_i = j] = Q_j + ATE_j + E[M_{ij} | S_i = j].$$

Here $Q_j = E[A_i | S_i = j]$ is the average ability of students enrolled at school j, a variable we label "peer quality." The quantity $E[M_{ij} | S_i = j]$ is the average suitability of j's students for this particular school. In a Roy (1951)-style model in which students sort into schools on the basis of comparative advantage in the production of Y_i, we would expect these average match effects to be positive. Parents and students may also choose schools on the basis of peer quality Q_j, overall school effectiveness ATE_j, or the idiosyncratic match M_{ij} for various outcomes.

III. Empirical Methods

The goal of our empirical analysis is to assess the roles of peer quality, school effectiveness, and academic match quality in applicant preferences. Our analysis proceeds in three steps. We first use rank-ordered choice lists to estimate preferences, thereby generating measures of each school's popularity. Next, we estimate schools' causal effects on test scores, high school graduation, college attendance, and college choice. Finally, we combine these two sets of estimates to characterize the relationships among school popularity, peer quality, and causal effectiveness.

A. *Estimating Preferences*

Let U_{ij} denote student i's utility from enrolling in school j, and let $\mathcal{J} = \{1, \ldots, J\}$ represent the set of available schools. We abstract from the fact that students rank programs rather than schools by ignoring repeat occurrences of any individual school on a student's choice list. Therefore, U_{ij} may be interpreted as the indirect utility associated with student i's favorite program at school j. The school ranked first on a student's choice list is

$$R_{i1} = \arg\max_{j \in \mathcal{J}} U_{ij},$$

while subsequent ranks satisfy

$$R_{ik} = \arg\max_{j \in \mathcal{J} \setminus \{R_{im}: m < k\}} U_{ij}, \quad k > 1.$$

Student i's rank-order list is then $R_i = (R_{i1}, \ldots, R_{i\ell(i)})'$, where $\ell(i)$ is the length of the list submitted by this student.

We summarize these preference lists by fitting random utility models with parameters that vary according to observed student characteristics. Student i's utility from enrolling in school j is modeled as

$$U_{ij} = \delta_{c(X_i)j} - \tau_{c(X_i)} D_{ij} + \eta_{ij}, \tag{4}$$

where the function $c(X_i)$ assigns students to covariate cells based on the variables in the vector X_i, and D_{ij} records distance from student i's home address to school j. The parameter δ_{cj} is the mean utility of school j for students in covariate cell c, and τ_c is a cell-specific distance parameter or "cost." We include distance in the model because a large body of evidence suggests it plays a central role in school choices (e.g., Hastings, Kane, and Staiger 2009 and Abdulkadiroğlu et al. 2017). We model unobserved tastes η_{ij} as following independent extreme value type I distributions conditional on X_i and $D_i = (D_{i1}, \ldots, D_{iJ})'$. Equation (4) is therefore a rank-ordered multinomial logit model (Hausman and Ruud 1987).

The logit model implies the conditional likelihood of the rank list R_i is

$$\mathcal{L}(R_i | X_i, D_i) = \prod_{k=1}^{\ell(i)} \frac{\exp\left(\delta_{c(X_i)R_{ik}} - \tau_{c(X_i)} D_{iR_{ik}}\right)}{\sum_{j \in \mathcal{J} \setminus \{R_{im}: m<k\}} \exp\left(\delta_{c(X_i)j} - \tau_{c(X_i)} D_{ij}\right)}.$$

We allow flexible heterogeneity in tastes by estimating preference models separately for 360 covariate cells defined by the intersection of borough, sex, race/ethnicity (black, Hispanic, or other), subsidized lunch status, above-median census tract income, and terciles of the mean of eighth grade math and reading scores. This specification follows several recent studies that flexibly parametrize preference heterogeneity in terms of observable characteristics (e.g., Hastings, Hortaçsu, and Syverson 2017 and Langer 2016). Students rarely rank schools outside their home boroughs, so covariate cells often include zero students ranking any given out-of-borough school. We therefore restrict the choice set $\mathcal{J}$ to schools located in the home borough and aggregate all other schools into an outside option with utility normalized to zero. Maximum likelihood estimation of the preference parameters produces a list of school mean utilities along with a distance coefficient for each covariate cell.

B. *Estimating School Effectiveness*

Our analysis of school effectiveness aims to recover the parameters of the potential outcome equations defined in Section II. We take two approaches to estimating these parameters.

Approach 1: Selection on observables.

The first set of estimates is based on the assumption

$$E\left[Y_{ij} | X_i, S_i\right] = \alpha_j + X_i' \beta_j, \quad j = 1, \ldots, J. \tag{5}$$

This restriction, often labeled "selection on observables," requires school enrollment to be as good as random conditional on the covariate vector X_i, which includes sex, race, subsidized lunch status, the log of median census tract income, and eighth grade math and reading scores. Assumption (5) implies that an ordinary least squares (OLS) regression of Y_i on school indicators interacted with X_i recovers unbiased estimates of α_j and β_j for each school. This fully interacted specification is a multiple-treatment extension of the Oaxaca-Blinder (1973) treatment effects estimator (Kline 2011).[3] By allowing school effectiveness to vary with student characteristics, we generalize the constant effects "value-added" approach commonly used to estimate the contributions of teachers and schools to student achievement (Koedel, Mihaly, and Rockoff 2015).

The credibility of the selection on observables assumption underlying value-added estimators is a matter of continuing debate (Rothstein 2010, 2017; Kane et al. 2013; Bacher-Hicks, Kane, and Staiger 2014; Chetty, Friedman, and Rockoff 2014a, 2016, 2017; Guarino, Reckase, and Wooldridge 2015). Comparisons to results from admission lotteries indicate that school value-added models accurately predict the impacts of random assignment but are not perfectly unbiased (Deming 2014; Angrist et al. 2016b, 2017). Selection on observables may also be more plausible for test scores than for longer-run outcomes, for which lagged measures of the dependent variable are not available (Chetty, Friedman, and Rockoff 2014a). We therefore report OLS estimates as a benchmark and compare these to estimates from a more general strategy that relaxes assumption (5).

Approach 2: Rank-ordered control functions.

Our second approach is motivated by the restriction

$$(6) \quad E\left[Y_{ij} \mid X_i, D_i, \eta_{i1}, \ldots, \eta_{iJ}, S_i\right] = \alpha_j + X_i'\beta_j + g_j(D_i, \eta_{i1}, \ldots, \eta_{iJ}), \quad j = 1, \ldots, J.$$

This restriction implies that any omitted variable bias afflicting OLS value-added estimates is due either to spatial heterogeneity captured by distances to each school (D_i) or to the preferences underlying the rank-ordered lists submitted to the assignment mechanism (η_{ij}). The function $g_j(\cdot)$ allows potential outcomes to vary arbitrarily across students with different preferences over schools. Factors that lead students with the same observed characteristics, spatial locations, and preferences to ultimately enroll in different schools, such as school priorities, random rationing due to oversubscription, or noncompliance with the assignment mechanism, are presumed to be unrelated to potential outcomes.

Under assumption (6), comparisons of matched sets of students with the same covariates, values of distance, and rank-ordered choice lists recover causal effects of school attendance. This model is therefore similar to the "self-revelation" model proposed by Dale and Krueger (2002, 2014) in the context of postsecondary enrollment. Dale and Krueger assume that students reveal their unobserved "types" via the selectivity of their college application portfolios, so college enrollment is as good as random among students who apply to the same schools. Similarly, (6) implies that

[3] We also include main effects of borough so that the model includes the same variables used to define covariate cells in the preference model.

high school applicants reveal their types through the content of their rank-ordered preference lists.

Though intuitively appealing, full nonparametric matching on rank-ordered lists is not feasible in practice because few students share the exact same rankings. We therefore use the structure of the logit choice model in equation (4) to derive a parametric approximation to this matching procedure. Specifically, we replace equation (6) with the assumption

$$
\begin{aligned}
(7)\quad & E\left[Y_{ij} \mid X_i, D_i, \eta_{i1}, \ldots, \eta_{iJ}, S_i\right] \\
& = \alpha_j + X_i'\beta_j + D_i'\gamma + \sum_{k=1}^{J} \psi_k \times \left(\eta_{ik} - \mu_\eta\right) + \varphi \times \left(\eta_{ij} - \mu_\eta\right), \quad j = 1, \ldots, J,
\end{aligned}
$$

where $\mu_\eta = E\left[\eta_{ij}\right]$ is Euler's constant.[4] As in the multinomial logit selection model of Dubin and McFadden (1984), equation (7) imposes a linear relationship between potential outcomes and the unobserved logit errors. Functional form assumptions of this sort are common in multinomial selection models with many alternatives, where requirements for nonparametric identification are very stringent (Lee 1983; Dahl 2002; Heckman, Urzua, and Vytlacil 2008).[5]

Equation (7) accommodates a variety of forms of selection on unobservables. The coefficient ψ_k represents an effect of the preference for school k common to all potential outcomes. This permits students with strong preferences for particular schools to have higher or lower general ability A_i. The parameter φ captures an additional match effect of the preference for school j on the potential outcome at this specific school. The model therefore allows for "essential" heterogeneity linking preferences to unobserved match effects in student outcomes (Heckman, Urzua, and Vytlacil 2006). A Roy (1951)-style model of selection on gains would imply $\varphi > 0$, but we do not impose this restriction.

By iterated expectations, equation (7) implies that mean observed outcomes at school j are

$$
\begin{aligned}
(8)\quad & E\left[Y_i \mid X_i, D_i, R_i, S_i = j\right] \\
& = \alpha_j + X_i'\beta_j + D_i'\gamma + \sum_{k=1}^{J} \psi_k \lambda_k\left(X_i, D_i, R_i\right) + \varphi \lambda_j\left(X_i, D_i, R_i\right),
\end{aligned}
$$

where $\lambda_k\left(X_i, D_i, R_i\right) = E\left[\eta_{ik} - \mu_\eta \mid X_i, D_i, R_i\right]$ gives the mean preference for school k conditional on a student's characteristics, spatial location, and preference list. The $\lambda_k(\cdot)$s serve as "control functions" correcting for selection on unobservables (Heckman and Robb 1985, Blundell and Matzkin 2014, Wooldridge 2015). As shown in online Appendix Section B.1, these functions are generalizations of the formulas derived by Dubin and McFadden (1984), extended to account for the fact that we observe a list of several ranked alternatives rather than just the most preferred choice.

Note that equation (8) includes main effects of distance to each school; we do not impose an exclusion restriction for distance. Identification of the selection parameters ψ_k and φ comes from variation in preference rankings for students who enroll at

[4] The means of both X_i and D_i are normalized to zero to maintain the interpretation that $\alpha_j = E\left[Y_{ij}\right]$.

[5] As discussed in Section V, we also estimate an alternative model that includes fixed effects for first choice schools.

the same school conditional on covariates and distance. Intuitively, if students who rank school j highly do better than expected given their observed characteristics at all schools, we will infer that $\psi_j > 0$. If these students do better than expected at school j but not elsewhere, we will infer that $\varphi > 0$.

We use the choice model parameters to build first-step estimates of the control functions, then estimate equation (8) in a second-step OLS regression of Y_i on school indicators and their interactions with X_i, controlling for D_i and the estimated $\lambda_k(\cdot)$ functions.[6] We adjust inference for estimation error in the control functions via a two-step extension of the score bootstrap procedure of Kline and Santos (2012). As detailed in online Appendix Section B.2, the score bootstrap avoids the need to recalculate the first-step logit estimates or the inverse variance matrix of the second-step regressors in the bootstrap iterations.

The Joint Distribution of Peer Quality and School Effectiveness.—Estimates of equations (5) and (7) may be used to calculate each school's peer quality. A student's predicted ability in the value-added model is

$$\hat{A}_i = \frac{1}{J}\sum_{j=1}^{J}\left[\hat{\alpha}_j + X_i'\hat{\beta}_j\right], \tag{9}$$

where $\hat{\alpha}_j$ and $\hat{\beta}_j$ are OLS value-added coefficients. Predicted ability in the control function model adds estimates of the distance and control function terms in equation (8). Estimated peer quality at school j is then $\hat{Q}_j = \sum_i \mathbf{1}\{S_i = j\}\hat{A}_i / \sum_i \mathbf{1}\{S_i = j\}$, the average predicted ability of enrolled students.

The end result of our school quality estimation procedure is a vector of estimates for each school, $\hat{\theta}_j = \left(\hat{\alpha}_j, \hat{\beta}_j', \hat{Q}_j\right)'$. The vector of parameters for the control function model also includes an estimate of the selection coefficient for school j, $\hat{\psi}_j$. These estimates are unbiased but noisy measures of the underlying school-specific parameters θ_j. We investigate the distribution of θ_j using the following hierarchical model:

$$\begin{aligned} \hat{\theta}_j \,|\, \theta_j &\sim N\left(\theta_j, \Omega_j\right), \\ \theta_j &\sim N\left(\mu_\theta, \Sigma_\theta\right). \end{aligned} \tag{10}$$

Here Ω_j is the sampling variance of the estimator $\hat{\theta}_j$, while μ_θ and Σ_θ govern the distribution of latent parameters across schools. In a hierarchical Bayesian framework μ_θ and Σ_θ are hyperparameters describing a prior distribution for θ_j. We estimate these hyperparameters by maximum likelihood applied to model (10), approximating Ω_j with an estimate of the asymptotic variance of $\hat{\theta}_j$.[7] The resulting estimates of μ_θ and Σ_θ characterize the joint distribution of peer quality and school treatment effect parameters, purged of the estimation error in $\hat{\theta}_j$.

[6] The choice model uses only preferences over schools in students' home boroughs, so $\lambda_k(\cdot)$ is undefined for students outside school k's borough. We therefore include dummies for missing values and code the control functions to 0 for these students. We similarly code D_{ik} to 0 for students outside of school k's borough and include borough indicators so that the distance coefficients are estimated using only within-borough variation. Our key results are not sensitive to dropping students attending out-of-borough schools from the sample.

[7] The peer quality estimates $\hat{Q}_j$ are typically very precise, so we treat peer quality as known rather than estimated when fitting the hierarchical model.

This hierarchical model can also be used to improve estimates of parameters for individual schools. An empirical Bayes (EB) posterior mean for θ_j is given by

$$\theta_j^* = \left(\hat{\Omega}_j^{-1} + \hat{\Sigma}_\theta^{-1}\right)^{-1}\left(\hat{\Omega}_j^{-1}\hat{\theta}_j + \hat{\Sigma}_\theta^{-1}\hat{\mu}_\theta\right),$$

where $\hat{\Omega}_j$, $\hat{\mu}_\theta$, and $\hat{\Sigma}_\theta$ are estimates of Ω_j, μ_θ, and Σ_θ. Relative to the unbiased but noisy estimate $\hat{\theta}_j$, this EB shrinkage estimator uses the prior distribution to reduce sampling variance at the cost of increased bias, yielding a minimum mean squared error (MSE) prediction of θ_j (Robbins 1956, Morris 1983). This approach parallels recent work applying shrinkage methods to estimate causal effects of teachers, schools, neighborhoods, and hospitals (Chetty, Friedman, and Rockoff 2014a; Hull 2016; Angrist et al. 2017; Chetty and Hendren 2017; Finkelstein et al. 2017). Online Appendix Section B.3 further describes our EB estimation strategy. In addition to reducing MSE, empirical Bayes shrinkage eliminates attenuation bias that would arise in models using elements of $\hat{\theta}_j$ as regressors (Jacob and Lefgren 2008). We exploit this property by regressing estimates of school popularity on EB posterior means in the final step of our empirical analysis.

C. *Linking Preferences to School Effectiveness*

We relate preferences to peer quality and causal effects with regressions of the form

$$\hat{\delta}_{cj} = \kappa_c + \rho_1 Q_j^* + \rho_2 ATE_j^* + \rho_3 M_{cj}^* + \xi_{cj}, \tag{11}$$

where $\hat{\delta}_{cj}$ is an estimate of the mean utility of school j for students in covariate cell c, κ_c is a cell fixed effect, and Q_j^* and ATE_j^* are EB posterior mean predictions of peer quality and average treatment effects. The variable M_{cj}^* is an EB prediction of the mean match effect of school j for students in cell c. Observations in equation (11) are weighted by the inverse sampling variance of $\hat{\delta}_{cj}$. We use the variance estimator proposed by Cameron, Gelbach, and Miller (2011) to double-cluster inference by cell and school. Two-way clustering accounts for correlated estimation errors in $\hat{\delta}_{cj}$ across schools within a cell as well as unobserved determinants of popularity common to a given school across cells. We estimate equation (11) separately for Regents test scores, PSAT scores, high school graduation, college attendance, and college quality. The parameters ρ_1, ρ_2, and ρ_3 measure how preferences relate to peer quality, school effectiveness, and match quality.[8]

[8] The control function version of our estimation procedure is closely related to classic selection-correction methods from studies of labor supply decisions. In their review of identification of labor supply models, French and Taber (2011) details a procedure that estimates labor market participation probabilities in a first step, uses these probabilities to selection-correct a wage equation in a second step, then relates participation to the unselected wage equation parameters in a third "structural probit" step. Similarly, we use preference estimates to selection-correct equations for student outcomes, then link the selection-corrected outcome estimates back to preferences to understand relationships between choices and treatment effects.

Table 3—Variation in Student Preference Parameters

		Standard deviations		
	Mean	Within cells	Between cells	Total
	(1)	(2)	(3)	(4)
School mean utility	—	1.117	0.500	1.223
		(0.045)	(0.003)	(0.018)
Distance cost	0.330	—	0.120	0.120
	(0.006)		(0.005)	(0.005)
Number of students		270,157		
Number of schools		316		
Number of covariate cells		360		

Notes: This table summarizes variation in school value-added and utility parameters across schools and covariate cells. Utility estimates come from rank-ordered logit models fit to student preference rankings. These models include school indicators and distance to school and are estimated separately in covariate cells defined by borough, gender, race, subsidized lunch status, an indicator for above or below the median of census tract median income, and tercile of the average of eighth grade math and reading scores. Column 1 shows the mean of the distance coefficient across cells weighted by cell size. Column 2 shows the standard deviation of school mean utilities across schools within a cell, and column 3 shows the standard deviation of a given school's mean utility across cells. School mean utilities are deviated from cell averages to account for differences in the reference category across cells. Estimated standard deviations are adjusted for sampling error by subtracting the average squared standard error of the parameter estimates from the total variance.

IV. Parameter Estimates

A. *Preference Parameters*

Table 3 summarizes the distribution of household preference parameters across the 316 high schools and 360 covariate cells in the choice sample. The first row reports estimated standard deviations of the mean utility δ_{cj} across schools and cells, while the second row displays the mean and standard deviation of the cell-specific distance cost τ_c. School mean utilities are deviations from cell averages to account for differences in the reference category across boroughs, and calculations are weighted by cell size. We adjust these standard deviations for sampling error in the estimated preference parameters by subtracting the average squared standard error from the sample variance of mean utilities.

Consistent with the descriptive statistics in Table 1, the preference estimates indicate that households dislike more distant schools. The mean distance cost is 0.33. This implies that increasing the distance to a particular school by one mile reduces the odds that a household prefers this school to another in the same borough by 33 percent. The standard deviation of the distance cost across covariate cells is 0.12. While there is significant heterogeneity in distastes for distance, all of the estimated distance costs are positive, suggesting that all subgroups prefer schools closer to home.

The estimates in Table 3 reveal significant heterogeneity in tastes for schools both within and between subgroups. The within-cell standard deviation of school mean utilities, which measures the variation in δ_{cj} across schools j for a fixed cell c, equals 1.12. This is equivalent to roughly 3.4 (1.12/0.33) miles of distance, implying that households are willing to travel substantial distances to attend more popular schools. The between-cell standard deviation, which measures variation in δ_{cj} across c for a

Table 4—Distributions of Peer Quality and Treatment Effect Parameters for Regents Math Scores

	Value-added model		Control function model	
	Mean (1)	SD (2)	Mean (3)	SD (4)
Peer quality	0	0.288	0	0.305
	—	(0.012)	—	(0.012)
ATE	0	0.290	0	0.233
	—	(0.012)	—	(0.014)
Female	−0.048	0.062	−0.029	0.062
	(0.005)	(0.006)	(0.005)	(0.006)
Black	−0.112	0.130	−0.108	0.120
	(0.011)	(0.011)	(0.010)	(0.011)
Hispanic	−0.097	0.114	−0.085	0.105
	(0.010)	(0.011)	(0.010)	(0.012)
Subsidized lunch	0.001	0.052	0.026	0.054
	(0.005)	(0.006)	(0.005)	(0.006)
log census tract median income	0.020	0.037	0.013	0.045
	(0.005)	(0.007)	(0.005)	(0.006)
Eighth grade math score	0.622	0.105	0.599	0.105
	(0.007)	(0.006)	(0.007)	(0.006)
Eighth grade reading score	0.159	0.048	0.143	0.052
	(0.004)	(0.004)	(0.004)	(0.004)
Preference coefficient (ψ_j)	—	—	−0.001	0.007
			(0.001)	(0.000)
Match coefficient (φ)	—	—	0.006	—
			(0.001)	

Notes: This table reports estimated means and standard deviations of peer quality and school treatment effect parameters for Regents math scores. Peer quality is a school's average predicted test score given the characteristics of its students. The ATE is a school's average treatment effect, and other treatment effect parameters are school-specific interactions with student characteristics. Estimates come from maximum likelihood models fit to school-specific regression coefficients. Columns 1 and 2 report estimates from an OLS regression that includes interactions of school indicators with sex, race, subsidized lunch, the log of the median income in a student's census tract, and eighth grade reading and math scores. This model also includes main effects of borough. Columns 3 and 4 show estimates from a control function model that adds distance to each school and predicted unobserved preferences from the choice model. Control functions and distance variables are set to 0 for out-of-borough schools and indicators for missing values are included.

fixed j, is 0.50, equivalent to about 1.5 (0.5/0.33) miles of distance. The larger within-cell standard deviation indicates that students in different subgroups tend to prefer the same schools.

B. *School Effectiveness and Peer Quality*

Our estimates of school treatment effects imply substantial variation in both causal effects and sorting across schools. Table 4 reports estimated means and standard deviations of peer quality Q_j, average treatment effects ATE_j, and slope coefficients β_j. We normalize the means of Q_j and ATE_j to 0 and quantify the variation in these parameters relative to the average school. As shown in column 2, the value-added model produces standard deviations of Q_j and ATE_j for Regents math scores equal to 0.29σ. This is somewhat larger than corresponding estimates of variation in school value-added from previous studies (usually around 0.15–0.2σ; see, e.g., Angrist et al. 2017). One possible reason for this difference is that most students in our sample attend high school for two years before taking Regents math exams, while previous studies look at impacts after one year.

As shown in columns 3 and 4 of Table 4, the control function model attributes some of the variation in Regents math value-added parameters to selection bias. Adding controls for unobserved preferences and distance increases the estimated standard deviation of Q_j to 0.31σ and reduces the estimated standard deviation of ATE_j to 0.23σ. Figure 1, which compares value-added and control function estimates for all five outcomes, demonstrates that this pattern holds for other outcomes as well: adjusting for selection on unobservables compresses the estimated distributions of treatment effects. This compression is more severe for high school graduation, college attendance, and college quality than for Regents math and PSAT scores. Our findings are therefore consistent with previous evidence that bias in OLS value-added models is more important for longer-run and non-test score outcomes (see, e.g., Chetty, Friedman, and Rockoff 2014b).

The bottom rows of Table 4 show evidence of substantial treatment effect heterogeneity across students. For example, the standard deviation of the slope coefficient on a black indicator equals 0.12σ in the control function model. This implies that holding the average treatment effect ATE_j fixed, a one standard deviation improvement in a school's match quality for black students boosts scores for these students by about a tenth of a standard deviation relative to whites. We also find significant variation in slope coefficients for gender (0.06σ), Hispanic (0.11σ), subsidized lunch status (0.05σ), the log of median census tract income (0.05σ), and eighth grade math and reading scores (0.11σ and 0.05σ). The final row of column 3 reports a control function estimate of φ, the parameter capturing matching between unobserved preferences and Regents scores. This estimate indicates a positive relationship between preferences and the unobserved component of student-specific test score gains, but the magnitude of the coefficient is very small.[9]

Our estimates imply that high-ability students tend to enroll in more effective schools. Table 5 reports correlations between Q_j and key school treatment effect parameters based on control function estimates for Regents math scores. Corresponding value-added estimates appear in online Appendix Table A2. The estimated correlation between peer quality and average treatment effects is 0.59. This may reflect either positive peer effects or higher-achieving students' tendency to enroll in schools with better inputs. Our finding that schools with high-ability peers are more effective contrasts with recent studies of exam schools in New York City and Boston, which show limited treatment effects for highly selective public schools (Abdulkadiroğlu, Angrist, and Pathak 2014; Dobbie and Fryer 2014). Within the broader New York public high school system, we find a strong positive association between school effectiveness and average student ability.

Table 5 also reports estimated correlations of Q_j and ATE_j with the slope coefficients β_j. Schools with larger average treatment effects tend to be especially good for girls: the correlation between ATE_j and the female slope coefficient is positive and statistically significant. This is consistent with evidence from Deming et al.

[9]The average predicted value of $(\eta_{ij} - \mu_\eta)$ for a student's enrolled school in our sample is 2.0. Our estimate of φ therefore implies that unobserved match effects increase average test scores by about 1 percent of a standard deviation ($0.006\sigma \times 2.0 = 0.012\sigma$).

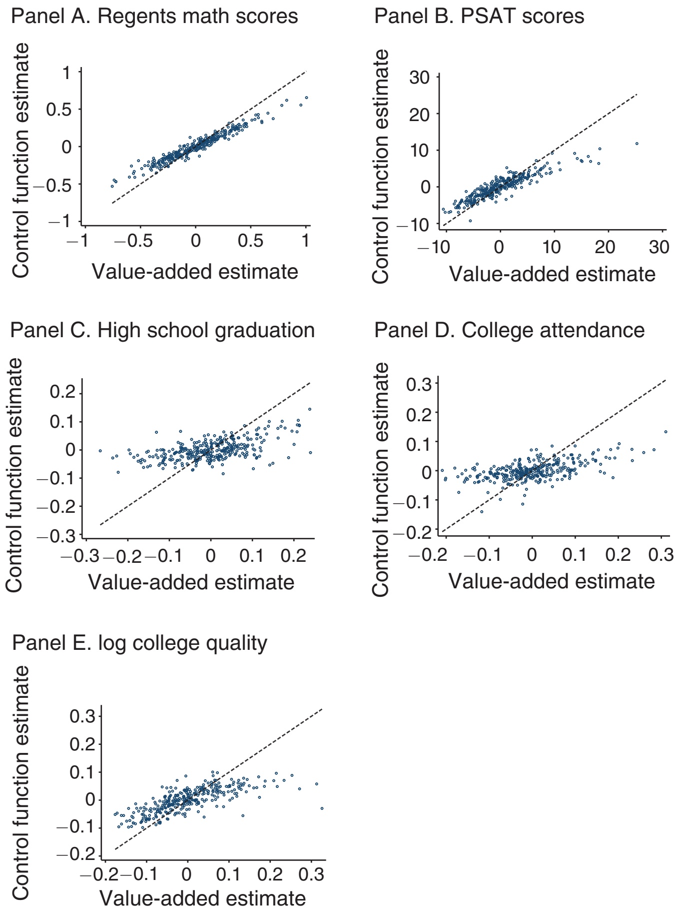

Figure 1. Comparison of Value-Added and Control Function Estimates of School Average Treatment Effects

Notes: This figure plots school average treatment effect (ATE) estimates from value-added models against corresponding estimates from models including control functions that adjust for selection on unobservables. Value-added estimates come from regressions of outcomes on school indicators interacted with gender, race, subsidized lunch status, the log of census tract median income, and eighth grade math and reading scores. Control function models add distance to school and predicted unobserved tastes from the choice model. Points in the figure are empirical Bayes posterior means from models fit to the distribution of school-specific estimates. Dashed lines show the 45-degree line.

(2014) showing that girls' outcomes are more responsive to school value-added. We estimate a very high positive correlation between black and Hispanic coefficients, suggesting that match effects tend to be similar for these two groups.

The slope coefficient on eighth grade reading scores is negatively correlated with peer quality and the average treatment effect. Both of these estimated correlations are below -0.4 and statistically significant. In other words, schools that enroll higher-ability students and produce larger achievement gains are especially effective at teaching low-achievers. In contrast to our estimate of the parameter φ, this suggests negative selection on the observed component of match effects in student achievement. A similar selection pattern is documented by Walters (2018), which

Table 5—Correlations of Peer Quality and Treatment Effect Parameters for Regents Math Scores

		Control function parameters						
	Peer quality (1)	ATE (2)	Female (3)	Black (4)	Hispanic (5)	Sub. lunch (6)	Math score (7)	Reading score (8)
ATE	0.588							
	(0.052)							
Female	0.078	0.299						
	(0.078)	(0.101)						
Black	0.006	0.107	−0.177					
	(0.077)	(0.106)	(0.142)					
Hispanic	−0.013	0.115	−0.235	0.922				
	(0.080)	(0.112)	(0.150)	(0.028)				
Subsidized lunch	0.045	−0.168	0.066	−0.038	0.004			
	(0.086)	(0.117)	(0.140)	(0.153)	(0.159)			
Eighth grade math score	−0.075	0.037	−0.074	−0.005	−0.007	0.060		
	(0.064)	(0.083)	(0.099)	(0.102)	(0.109)	(0.113)		
Eighth grade reading score	−0.418	−0.452	−0.193	−0.090	−0.078	0.004	0.256	
	(0.068)	(0.094)	(0.117)	(0.130)	(0.138)	(0.135)	(0.099)	
Preference coefficient (ψ_j)	0.429	0.247	0.212	−0.083	−0.058	−0.127	−0.241	−0.281
	(0.063)	(0.092)	(0.104)	(0.106)	(0.111)	(0.116)	(0.083)	(0.099)

Notes: This table reports estimated correlations between peer quality and school treatment effect parameters for Regents math scores. The ATE is a school's average treatment effect, and other treatment effect parameters are school-specific interactions with student characteristics. Estimates come from maximum likelihood models fit to school-specific regression coefficients from a control function model controlling for observed characteristics, distance to school, and unobserved tastes from the choice model.

shows that lower-scoring students in Boston are less likely to apply to charter schools despite receiving larger achievement benefits. Section V presents a more systematic investigation of relationships between preferences and match effects.

Patterns of estimates for PSAT scores, high school graduation, college attendance, and college quality are generally similar to results for Regents math scores. Online Appendix Tables A3–A6 present estimated distributions of peer quality and school effectiveness for these longer-run outcomes. For all five outcomes, we find substantial variation in peer quality and average treatment effects, a strong positive correlation between these variables, and significant effect heterogeneity with respect to student characteristics. Overall, causal effects for the longer-run outcomes are highly correlated with effects on Regents math scores. This is evident in Figure 2, which plots EB posterior mean predictions of average treatment effects on Regents scores against corresponding predictions for the other four outcomes. These results are consistent with recent evidence that short-run test score impacts reliably predict effects on longer-run outcomes (Chetty et al. 2011; Dynarski, Hyman, and Schanzenbach 2013; Angrist et al. 2016a).

C. *Decomposition of School Average Outcomes*

We summarize the joint distribution of peer quality and school effectiveness by implementing the decomposition introduced in Section II. Table 6 uses the control function estimates to decompose variation in school averages for each outcome into

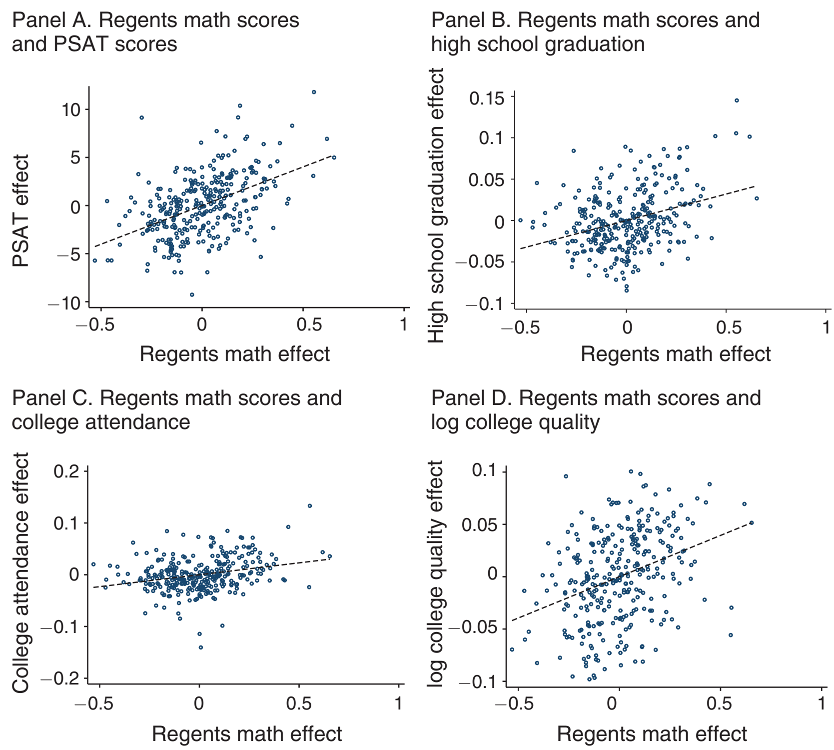

FIGURE 2. RELATIONSHIPS BETWEEN EFFECTS ON TEST SCORES AND EFFECTS ON LONG-RUN OUTCOMES

Notes: This figure plots estimates of causal effects on Regents math scores against estimates of effects on longer-run outcomes. Treatment effects are empirical Bayes posterior mean estimates of school average treatment effects from control function models. Panel A plots the relationship between Regents math effects and effects on PSAT scores. Panels B, C, and D show corresponding results for high school graduation, college attendance, and log college quality.

components explained by peer quality, school effectiveness, average match effects, and covariances of these components.

Consistent with the estimates in Table 4, both peer quality and school effectiveness play roles in generating variation in school average outcomes, but peer quality is generally more important. Peer quality explains 47 percent of the variance in average Regents scores (0.093/0.191), while average treatment effects explain 28 percent (0.054/0.191). The explanatory power of peer quality for other outcomes ranges from 49 percent (PSAT scores) to 83 percent (high school graduation), while the importance of average treatment effects ranges from 10 percent (PSAT scores) to 19 percent (log college quality).

Despite the significant variation in slope coefficients documented in Table 4, match effects are unimportant in explaining dispersion in school average outcomes. The variance of match effects accounts for only 5 percent of the variation in average Regents scores, and corresponding estimates for the other outcomes are also small. Although school treatment effects vary substantially across subgroups, there is not much sorting of students to schools on this basis, so the existence of potential match effects is of little consequence for realized variation in outcomes across schools.

TABLE 6—DECOMPOSITION OF SCHOOL AVERAGE OUTCOMES

	Regents math (1)	PSAT score/10 (2)	High school graduation (3)	College attendance (4)	log college quality (5)
Total variance of average outcome	0.191	1.586	0.012	0.016	0.021
Variance of peer quality	0.093	0.781	0.010	0.010	0.009
Variance of ATE	0.054	0.160	0.002	0.003	0.004
Variance of match	0.008	0.027	0.002	0.002	0.001
2cov(peer quality, ATE)	0.081	0.745	0.005	0.008	0.011
2cov(peer quality, match)	−0.023	−0.061	−0.003	−0.003	−0.002
2cov(ATE, match)	−0.022	−0.068	−0.004	−0.005	−0.003

Notes: This table decomposes variation in average outcomes across schools into components explained by student characteristics, school average treatment effects (ATE), and the match between student characteristics and school effects. Estimates come from control function models adjusting for selection on unobservables. Column 1 shows results for Regents math scores in standard deviation units, column 2 reports estimates for PSAT scores, column 3 displays estimates for high school graduation, column 4 reports results for college attendance, and column 5 shows results for log college quality. The first row reports the total variance of average outcomes across schools. The second row reports the variance of peer quality, defined as the average predicted outcome as a function of student characteristics and unobserved tastes. The third row reports the variance of ATE, and the fourth row displays the variance of the match effect. The remaining rows show covariances of these components.

The final three rows of Table 6 quantify the contributions of covariances among peer quality, treatment effects, and match effects. As a result of the positive relationship between peer quality and school effectiveness, the covariance between Q_j and ATE_j substantially increases cross-school dispersion in mean outcomes. The covariances between match effects and the other variance components are negative. This indicates that students at highly effective schools and schools with higher-ability students are less appropriately matched on the heterogeneous component of treatment effects, slightly reducing variation in school average outcomes.

D. *Testing for Bias in Estimates of School Effectiveness*

As described in Section I, New York City's centralized assignment mechanism breaks ties at random for students with the same preferences and priorities. Abdulkadiroğlu et al. (2017, 2019) derive methods for using the random assignment implicit in such systems for impact evaluation. The core of this approach uses student preferences and priorities along with the structure of the assignment mechanism to derive a probability of assignment (propensity score) for each student to each school where random tie-breaking occurs. Conditioning on the propensity score is sufficient to isolate the random component of school assignment, generating valid instruments for school enrollment (Rosenbaum and Rubin 1983). The priority information necessary to produce lottery propensity scores is only available for the 2003 New York City high school applicant cohort, and random assignment occurs for only a minority of schools. While these data constraints prevent us from producing a lottery-based estimate of effectiveness for every school, we can use the shifts in school attendance resulting from these lottery assignments to test the accuracy of our value-added and control function estimates.

To implement these tests we first construct DA mechanism-based propensity scores and then apply the lottery-based test for bias in non-experimental estimators of school effectiveness proposed by Angrist et al. (2016b, 2017). For a set of L lotteries, this test is implemented by estimating the following two-stage least squares (2SLS) system:

$$Y_i = \kappa_0 + \phi \hat{Y}_i + \sum_{\ell=1}^{L} \kappa_\ell p_{i\ell} + e_i, \tag{12}$$

$$\hat{Y}_i = \pi_0 + \sum_{\ell=1}^{L} \left[\pi_\ell Z_{i\ell} + \omega_\ell p_{i\ell}\right] + \upsilon_i, \tag{13}$$

where $\hat{Y}_i$ is the fitted value generated by a non-experimental estimation procedure (either value-added or control function), $Z_{i\ell}$ is an indicator equal to 1 if student i is assigned to school ℓ, and $p_{i\ell}$ is the propensity score measuring student i's probability of assignment to school ℓ. The first-stage coefficients π_ℓ describe the non-experimental estimator's predicted effects of assignment in each lottery, which are nonzero because lottery offers shift students across schools. Angrist et al. (2017) shows that the "forecast coefficient" ϕ should equal 1 if the estimator used to generate $\hat{Y}_i$ correctly predicts the effects of random lottery assignments on average, while the overidentification test for the system defined by (12) and (13) measures whether the estimator has the same predictive validity in every lottery. As discussed in Angrist et al. (2016b), the combination of these restrictions can be viewed as a Hausman (1983)-style test comparing lottery-based instrumental variables (IV) and OLS value-added estimates. If the restrictions underlying the value-added model are true, the value-added estimates should match lottery-based estimates of the same causal parameters.

As shown in Table 7, this lottery-based test suggests that our value-added and control function estimates accurately capture the causal effects of schools on student outcomes.[10] Column 1 reports tests of an "uncontrolled" model that measures school effectiveness as the unadjusted mean outcome at the school. This model generates forecast coefficients far from one and decisive rejections of the overidentification test for all three outcomes available for the 2003 cohort (Regents math, PSAT scores, and high school graduation), indicating that the available lotteries have power to detect bias in the most naïve nonexperimental estimators. Columns 2 and 3 show that the addition of controls for observed student characteristics generates forecast coefficients much closer to 1 and overidentification tests that generally do not reject at conventional levels. Unfortunately, we cannot use lotteries to validate our estimates for postsecondary outcomes since college attendance data are not available for the 2003 applicant cohort, so estimates for these outcomes should be viewed more cautiously.

While the estimates in Table 7 are encouraging, it's worth noting that the lotteries available in New York may have weak power to detect bias in our value-added and control function models. Specifically, the first stage F-statistics for equation (13) are below the rule-of-thumb value of 10 commonly used to diagnose weak instruments (Staiger and Stock 1997), implying that the lotteries tend to shift students

[10] We validate our approach to reconstructing school lotteries by reporting relationships between lottery offers and students characteristics in Appendix Table A7. Without controls for propensity scores, school offers are strongly correlated with baseline test scores and other observables, but we cannot reject that offers at all schools are independent of observed characteristics after controlling for the propensity scores. This indicates that our strategy successfully isolates randomized lottery assignments.

TABLE 7—LOTTERY-BASED TESTS FOR BIAS IN ESTIMATES OF SCHOOL EFFECTIVENESS

	2SLS			UJIVE		
	Uncontrolled	Value-added	Control function	Uncontrolled	Value-added	Control function
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Math						
Forecast coefficient	0.599	0.965	0.967	0.598	0.961	0.963
	(0.040)	(0.038)	(0.038)	(0.040)	(0.046)	(0.045)
p-value	0.000	0.354	0.388	0.000	0.343	0.376
Overid. $\chi^2(123)$ stat.	174.1	84.88	86.50	174.1	84.88	86.50
p-value	0.000	0.996	0.995	0.002	0.996	0.995
First-stage F-stat.	91.1	6.1	5.9	91.1	6.1	5.9
Number of lotteries		124			124	
Number of students		22,515			22,515	
Panel B. PSAT						
Forecast coefficient	0.306	0.879	0.912	0.296	0.815	0.862
	(0.049)	(0.048)	(0.048)	(0.050)	(0.083)	(0.084)
p-value	0.000	0.012	0.066	0.000	0.025	0.101
Overid. $\chi^2(123)$ stat.	145.8	112.9	106.8	145.7	111.7	106.1
p-value	0.079	0.732	0.851	0.0792	0.759	0.861
First-stage F-stat.	62.9	2.3	2.1	62.9	2.3	2.1
Number of lotteries		124			124	
Number of students		16,554			16,554	
Panel C. High school graduation						
Forecast coefficient	0.333	0.905	0.914	0.329	0.893	0.901
	(0.063)	(0.076)	(0.076)	(0.064)	(0.093)	(0.094)
p-value	0.000	0.214	0.262	0.000	0.245	0.295
Overid. $\chi^2(123)$ stat.	205.1	145.6	147.7	205.1	145.6	147.7
p-value	0.000	0.080	0.064	0.000	0.080	0.064
First-stage F-stat.	92.1	5.4	5.0	92.1	5.4	5.0
Number of lotteries		124			124	
Number of students		32,131			32,131	

Notes: This table reports the results of lottery-based tests for bias in estimates of school effectiveness. The sample is restricted to students who have nondegenerate risk for at least one school and lotteries with 100 or more students at risk. Students are considered to have risk at a given school if their propensity score is strictly between 0 and 1 and they are in a score cell with variation in school offers. Columns 1 and 4 measure school effectiveness as the school mean outcome, columns 2 and 5 use value-added estimates, and columns 3 and 6 use control function estimates. Forecast coefficients and overidentification tests in columns 1–3 come from two-stage least squares regressions of test scores on the fitted values from the non-lottery estimation procedure, instrumenting with school-specific lottery offer indicators and controlling for school-specific propensity scores. Columns 4–6 use the Unbiased Jackknife Instrumental Variables (UJIVE) estimator of Kolesár (2013) instead of 2SLS.

across schools with similar estimated effectiveness. Columns 4–6 demonstrate that we obtain similar results based on the Unbiased Jackknife Instrumental Variables (UJIVE) estimator proposed by Kolesár (2013). The UJIVE estimator performs well with weak instruments and (unlike other common approaches such as limited information maximum likelihood) it is robust to the presence of treatment effect heterogeneity. The UJIVE estimates suggest that our non-experimental estimators are accurate, particularly the control function estimator, for which we cannot reject forecast unbiasedness or the overidentifying restrictions for any outcome (we reject forecast unbiasedness for the value-added estimator for PSAT scores). The similarity of 2SLS and UJIVE also eases concerns about weak instrument bias in 2SLS, though the weak first stage also indicates that tests based on both estimators are likely to have low power. Taken together, the results in Table 7 suggest that our estimation strategies generate reliable measures of causal effects, though the available lottery variation may be insufficient to detect modest violations.

V. Preferences, Peer Quality, and School Effectiveness

A. *Productivity versus Peers*

The last step of our analysis compares the relative strength of peer quality and school effectiveness as predictors of parent preferences. Table 8 reports estimates of equation (11) for Regents math scores, first including Q_j^* and ATE_j^* one at a time and then including both variables simultaneously. Mean utilities, peer quality, and treatment effects are scaled in standard deviations of their respective school-level distributions, so the estimates can be interpreted as the standard deviation change in mean utility associated with a 1 standard deviation increase in Q_j or ATE_j.

Bivariate regressions show that school popularity is positively correlated with both peer quality and school effectiveness. Results based on the OLS value-added model, reported in columns 1 and 2 of panel A, imply that a 1 standard deviation increase in Q_j is associated with a 0.42 standard deviation increase in mean utility, while a 1 standard deviation increase in ATE_j is associated with a 0.24 standard deviation increase in mean utility. The latter result contrasts with studies reporting no average test score impact of attending preferred schools (Cullen, Jacob, and Levitt 2006; Hastings, Kane, and Staiger 2009). These studies rely on admission lotteries that shift relatively small numbers of students across a limited range of schools. Our results show that looking across all high schools in New York City, more popular schools tend to be more effective on average.

While preferences are positively correlated with school effectiveness, however, this relationship is entirely explained by peer quality. Column 3 of panel A shows that when both variables are included together, the coefficient on peer quality is essentially unchanged, while the coefficient on the average treatment effect is rendered small and statistically insignificant. The ATE_j coefficient also remains precise: we can rule out increases in mean utility on the order of 0.06 standard deviations associated with a 1 standard deviation change in school value-added at conventional significance levels. The control function estimates in columns 5–7 are similar to the value-added estimates, showing no association between school effectiveness and popularity after controlling for peer quality.

Columns 4 and 8 of Table 8 explore the role of treatment effect heterogeneity by adding posterior mean predictions of match quality to equation (11), also scaled in standard deviation units of the distribution of match effects across schools and cells.[11] The match coefficient is negative for both the value-added and control function models, and the control function estimate is statistically significant. This reflects the negative correlation between baseline test score slope coefficients and peer

[11] Equation (11) captures the projection of cell mean utility on peer quality and cell mean treatment effects. In the control function model the projection of student-specific utility on student-specific treatment effects also includes the idiosyncratic match component $\varphi \eta_{ij}$. Our small estimate of the matching coefficient φ implies this term is negligible, so we focus on relationships at the cell mean level for both the value-added and control function models.

TABLE 8—PREFERENCES FOR PEER QUALITY AND REGENTS MATH EFFECTS

	Value-added models				Control function models			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. No controls for school characteristics								
Peer quality	0.416		0.438	0.406	0.407		0.439	0.437
	(0.061)		(0.063)	(0.067)	(0.057)		(0.059)	(0.059)
ATE		0.244	−0.033	−0.022		0.219	−0.051	−0.047
		(0.047)	(0.046)	(0.047)		(0.046)	(0.043)	(0.043)
Match effect				−0.072				−0.172
				(0.047)				(0.054)
Observations	21,684							
Panel B. With controls for school characteristics								
Peer quality	0.310		0.314	0.286	0.299		0.303	0.308
	(0.060)		(0.059)	(0.060)	(0.056)		(0.056)	(0.056)
ATE		0.157	−0.005	0.005		0.144	−0.008	−0.003
		(0.042)	(0.039)	(0.040)		(0.040)	(0.035)	(0.035)
Match effect				−0.068				−0.142
				(0.039)				(0.044)
Observations	20,200							

Notes: This table reports estimates from regressions of school popularity on peer quality and school effectiveness. School popularity is measured as the estimated mean utility for each school and covariate cell in the choice model from Table 4. Covariate cells are defined by borough, gender, race, subsidized lunch status, an indicator for students above the median of census tract median income, and tercile of the average of eighth grade math and reading scores. Peer quality is constructed as the average predicted Regents math score for enrolled students. Treatment effect estimates are empirical Bayes posterior mean predictions of Regents math effects. Mean utilities, peer quality, and treatment effects are scaled in standard deviation units. Columns 1–4 report results from value-added models, while columns 5–8 report results from control function models. All regressions include cell indicators and weight by the inverse of the squared standard error of the mean utility estimates. Panel A includes no additional controls, while panel B controls for the school environment score, violent and disruptive incidents per student, and percent of teachers with master's degrees. Standard errors are double-clustered by school and covariate cell.

quality reported in Table 5: schools that are especially effective for low-achieving students tend to be more popular among high-achievers and therefore enroll more of these students despite their lower match quality. This is consistent with recent studies of selection into early-childhood programs and charter schools, which also find negative selection on test score match effects (Cornelissen et al. 2016, Kline and Walters 2016, Walters 2018).

Correlations between popularity and effectiveness may fail to capture the causal effects of school effectiveness on parent demand if other determinants of parent preferences are correlated with school effectiveness.[12] It's worth noting that for omitted variables bias to explain our finding that parents do not place positive weight on effectiveness conditional on peer quality, an omitted amenity that parents value would need to be negatively correlated with effectiveness after controlling for Q_j. Since we might expect any omitted variables to be positively correlated with both effectiveness and demand (as is the case with peer quality itself), this sort of selection bias seems implausible. Nevertheless, Panel B of Table 8 investigates

[12] An analogous problem arises in studies of teacher and school value-added, which often analyze relationships between teacher or school effects and observed characteristics without quasi-experimental variation in the characteristics (see, e.g., Kane, Rockoff, and Staiger 2008 and Angrist, Pathak, and Walters 2013).

the potential for such omitted variable bias by adding controls for other important school attributes to equation (11). The sensitivity of regression coefficients to controls for observables is a common diagnostic for assessing the scope for selection on unobservables (Altonji, Elder, and Taber 2005; Oster 2019).

We explore the impact of controlling for three school covariates. The first is a measure of the quality of the school environment derived from New York City's Learning Environments Survey (New York City Department of Education 2008). This survey is taken each year by New York City students in grades 6 through 12 as well as parents and teachers. We construct an overall school environment score by taking the first principle component of the Safety and Respect, Communication, Engagement, and Academic Expectations summary measures from the school survey. Second, we add a measure of violent and disruptive incidents (VADI) per student reported by the New York State Department of Education (New York State Department of Education 2007b). Finally, we control for the fraction of teachers with master's degrees as reported on New York school report cards distributed between 2005 and 2007 (New York State Department of Education 2007a). The results in panel B of Table 8 show that our main conclusions are unaffected by the addition of these control variables: the coefficient on Q_j remains large and statistically insignificant, while the coefficient on ATE_j remains close to zero.[13] While we cannot control for all unobserved factors that influence preferences for schools, the robustness of our results to controls for observed characteristics suggests that our key findings are unlikely to be driven by omitted variable bias.

Figure 3 presents a graphical summary of the links among preferences, peer quality, and treatment effects by plotting bivariate and multivariate relationships between mean utility (averaged across covariate cells) and posterior predictions of Q_j and ATE_j from the control function model. Panel A shows strong positive bivariate correlations for both variables. Panel B plots mean utilities against residuals from a regression of Q_j^* on ATE_j^* (left-hand panel) and residuals from a regression of ATE_j^* on Q_j^* (right-hand panel). Adjusting for school effectiveness has little effect on the relationship between preferences and peer quality. In contrast, partialing out peer quality eliminates the positive association between popularity and effectiveness.

B. *Preferences and Effects on Longer-Run Outcomes*

Parents may care about treatment effects on outcomes other than short-run standardized test scores. We explore this by estimating equation (11) for PSAT scores, high school graduation, college attendance, and log college quality.

Results for these outcomes are similar to the findings for Regents math scores: preferences are positively correlated with average treatment effects in a bivariate sense but are uncorrelated with treatment effects conditional on peer quality. Table 9 reports results based on control function estimates of treatment effects. The magnitudes of all treatment effect coefficients are small, and the overall pattern of results suggests no systematic relationship between preferences and school effectiveness

[13] Online Appendix Table A8 reports the coefficients on the control variables. These estimates show that parents prefer schools with fewer violent incidents and those with more teachers with master's degrees, while the school environment score is uncorrelated with demand.

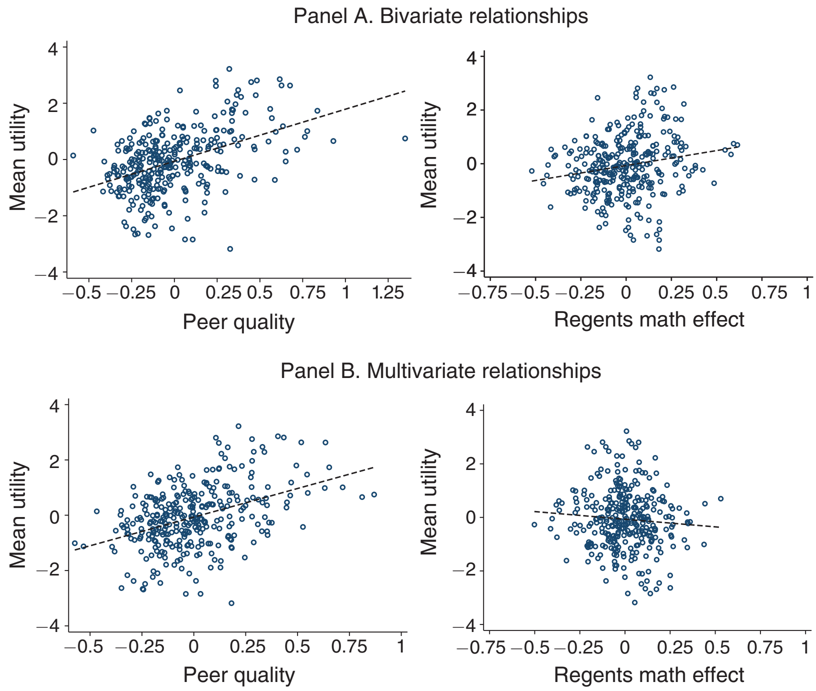

Figure 3. Relationships among Preferences, Peer Quality, and Regents Math Effects

Notes: This figure plots school mean utility estimates against estimates of peer quality and Regents math average treatment effects. Mean utilities are school average residuals from a regression of school-by-covariate cell mean utility estimates on cell indicators. Peer quality is defined as the average predicted Regents math score for enrolled students. Regents math effects are empirical Bayes posterior mean estimates of school average treatment effects from control function models. The left plot in panel A displays the bivariate relationship between mean utility and per quality, while the right plot shows the bivariate relationship between mean utility and Regents math effects. The left plot in panel B displays the relationship between mean utility and residuals from a regression of peer quality on Regents math effects, while the right plot shows the relationship between mean utility and residuals from a regression of Regents math effects on peer quality. Dashed lines are ordinary least squares regression lines.

conditional on peer composition. We find a modest positive relationship between preferences and match effects for log college quality, but corresponding estimates for PSAT scores, high school graduation, and college attendance are small and statistically insignificant. This pattern contrasts with results for the Norwegian higher education system, reported by Kirkeboen, Leuven, and Mogstad (2016), which show sorting into fields of study based on heterogeneous earnings gains. Unlike Norwegian college students, New York City's high school students do not prefer schools with higher academic match quality.

C. *Heterogeneity in Preferences for Peer and School Quality*

Previous evidence suggests that parents of higher-income, higher-achieving students place more weight on academic performance levels when choosing schools (Hastings, Kane, and Staiger 2009). This pattern may reflect either greater responsiveness to peer quality or more sensitivity to causal school effectiveness. If parents of high-achievers value school effectiveness, choice may indirectly create incentives

Table 9—Preferences for Peer Quality and School Effectiveness by Outcome

	PSAT score		High school graduation		College attendance		log college quality	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peer quality		0.467		0.430		0.235		0.322
		(0.070)		(0.070)		(0.054)		(0.065)
ATE	0.325	−0.092	0.103	−0.174	0.273	0.132	0.199	0.029
	(0.056)	(0.074)	(0.045)	(0.054)	(0.048)	(0.054)	(0.059)	(0.080)
Match effect		−0.049		−0.065		−0.017		0.053
		(0.047)		(0.044)		(0.050)		(0.061)
Observations	21,684							

Notes: This table reports estimates from regressions of school popularity on peer quality and school effectiveness separately by outcome. School popularity is measured as the estimated mean utility for each school and covariate cell in the choice model from Table 4. Covariate cells are defined by borough, gender, race, subsidized lunch status, an indicator for students above the median of census tract median income, and tercile of the average of eighth grade math and reading scores. Peer quality is constructed as the average predicted outcome for enrolled students. Treatment effect estimates are empirical Bayes posterior mean predictions from control function models. Mean utilities, peer quality, and treatment effects are scaled in standard deviation units. All regressions include cell indicators and weight by the inverse of the squared standard error of the mean utility estimates. Standard errors are double-clustered by school and covariate cell.

for schools to improve because better instruction will attract high-ability students, raising peer quality and therefore demand from other households. In Table 10 we investigate this issue by estimating equation (11) separately by race, subsidized lunch status, and baseline test score category.

We find that no subgroup of households responds to causal school effectiveness. Consistent with previous work, we find larger coefficients on peer quality among nonminority students, richer students (those ineligible for subsidized lunches), and students with high baseline achievement. We do not interpret this as direct evidence of stronger preferences for peer ability among higher-ability students; since students are more likely to enroll at schools they rank highly, any group component to preferences will lead to a positive association between students' rankings and the enrollment share of others in the same group.[14] The key pattern in Table 10 is that, among schools with similar peer quality, no group prefers schools with greater causal impacts on academic achievement.

D. *Changes in Demand over Time*

If parents do not have perfect information about school quality we might expect changes in demand over time as parents learn more about which schools are effective. The evolution of choice behavior is of particular interest in our sample since New York City changed from an uncoordinated assignment process to a coordinated single-offer system in 2003, the first year of our data (Abdulkadiroğlu, Agarwal, and Pathak 2017). Online Appendix Table A9 assesses whether parents systematically select more effective schools over time by reporting estimates of equation (11) based on preference models fit separately for each of the four applicant cohorts in

[14] This is a version of the "reflection problem" that plagues econometric investigations of peer effects (Manski 1993).

Table 10—Heterogeneity in Preferences for Peer Quality and Regents Math Effects

	By race			By subsidized lunch		By eighth grade test score tercile		
	Black (1)	Hispanic (2)	Other (3)	Eligible (4)	Ineligible (5)	Lowest (6)	Middle (7)	Highest (8)
Peer quality	0.396	0.370	0.705	0.410	0.501	0.251	0.395	0.686
	(0.060)	(0.063)	(0.128)	(0.057)	(0.077)	(0.055)	(0.062)	(0.092)
ATE	−0.047	−0.011	−0.192	−0.036	−0.076	−0.015	−0.029	−0.117
	(0.045)	(0.044)	(0.094)	(0.042)	(0.050)	(0.042)	(0.042)	(0.059)
Match effect	−0.200	−0.144	−0.149	−0.180	−0.155	−0.166	−0.169	−0.125
	(0.056)	(0.066)	(0.061)	(0.054)	(0.054)	(0.061)	(0.058)	(0.055)
Observations	7,467	7,433	6,784	11,043	10,641	7,264	7,286	7,134

Notes: This table reports estimates from regressions of school popularity on peer quality and school effectiveness separately by student subgroup. School popularity is measured as the estimated mean utility for each school and covariate cell in the choice model from Table 4. Peer quality is constructed as the average predicted Regents math score for enrolled students. Treatment effect estimates are empirical Bayes posterior mean predictions of Regents math effects from control function models. Mean utilities, peer quality, and treatment effects are scaled in standard deviation units. Peer quality is constructed as the average predicted Regents math score for enrolled students. All regressions include cell indicators and weight by the inverse of the squared standard error of the mean utility estimates. Standard errors are double-clustered by school and covariate cell.

our sample. The results reveal remarkably stable patterns of choice: in each year the coefficient on peer quality is large and positive, the coefficient on the average treatment effect is a precise zero, and the match effect coefficient is zero or negative. Evidently, more experience with the centralized matching process did not lead to a stronger relationship between preferences and effectiveness for parents in New York City. This suggests that either parents do not learn much about school effectiveness over time, or the patterns we identify reflect preferences for other school attributes rather than a lack of information about effectiveness.

E. *Alternative Specifications*

We investigate the robustness of our key results by estimating a variety of alternative specifications, reported in online Appendix Tables A10 and A11. To assess the sensitivity of our estimates to reasonable changes in our measure of school popularity, columns 1–4 of online Appendix Table A10 display results from models replacing $\hat{\delta}_{cj}$ in equation (11) with the log share of students in a cell ranking a school first or minus the log sum of ranks in the cell (treating unranked schools as tied). These alternative measures of demand produce very similar results to the rank-ordered logit results in Table 8.

Estimates based on students' submitted rankings may not accurately describe demand if students strategically misreport their preferences in response to the 12-choice constraint on list length. As noted in Section I, truthful reporting is a dominant strategy for the 72 percent of students who list fewer than 12 choices. Columns 5 and 6 of online Appendix Table A10 report results based on rank-ordered logit models estimated in the subsample of unconstrained students. Results here are again similar to the full sample estimates, suggesting that strategic misreporting is not an important concern in our setting.

Our preference estimation approach models students' choices among all schools in their home boroughs. Students may be unaware of some schools and therefore consider only a subset of the available alternatives. A conservative approach to defining consideration sets is to assume students are only aware of the schools ranked on their preference lists. Columns 7 and 8 of online Appendix Table A10 show results based on preference estimates that omit all unranked alternatives from the choice set. This approach produces similar estimates as well.

Equation (8) parameterizes the relationship between potential outcomes and preference rankings through the control functions $\lambda_k(\cdot)$. Columns 1–4 of online Appendix Table A11 present an alternative parameterization that replaces the control functions with fixed effects for first choice schools. This approach ignores information on lower-ranked schools but more closely parallels the application portfolio matching approach in Dale and Krueger (2002, 2014). As a second alternative specification, columns 5–8 report estimates from a control function model that drops the distance control variables from equation (8). This model relies on an exclusion restriction for distance, a common identification strategy in the literature on educational choice (Card 1995, Neal 1997, Booker et al. 2011, Mountjoy 2017, Walters 2018). These alternative approaches to estimating school effectiveness produce no meaningful changes in the results.

VI. Discussion

The findings reported here inform models of school choice commonly considered in the literature. Theoretical analyses often assume parents know students' potential achievement outcomes and choose between schools on this basis. For example, Epple, Figlio, and Romano (2004) and Epple and Romano (2008) study models in which parents value academic achievement and consumption of other goods, and care about peer quality only insofar as it produces higher achievement through peer effects. Hoxby (2000) argues that school choice may increase achievement by allowing students to sort on match quality. Such models imply that demand should be positively correlated with both average treatment effects and match effects conditional on peer quality, a prediction that is inconsistent with the pattern in Table 8.

Parents may choose between schools based on test score levels rather than treatment effects. Cullen, Jacob, and Levitt (2006) suggests confusion between levels and gains may explain limited effects of admission to preferred schools in Chicago. Since our setting has substantial variation in both levels and value-added, we can more thoroughly investigate this model of parent decision-making. If parents choose between schools based on average outcomes, increases in these outcomes due to selection and causal effectiveness should produce equal effects on popularity. In contrast, we find that demand only responds to the component of average outcomes that is due to enrollment of higher-ability students. That is, we can reject the view that parental demand is driven by performance levels: demand places no weight on the part of performance levels explained by value-added but significant weight on the part explained by peer quality.

It is important to note that our findings do not imply parents are uninterested in school effectiveness. Without direct information about treatment effects, for example, parents may use peer characteristics as a proxy for school quality, as in MacLeod and Urquiola (2015). In view of the positive correlation between peer quality and school effectiveness, this is a reasonable strategy for parents that cannot observe treatment effects and wish to choose effective schools. Effectiveness varies widely conditional on peer quality, however, so parents make substantial sacrifices in academic quality by not ranking schools based on effectiveness. Table 11 compares Regents math effects for observed preference rankings versus hypothetical rankings in which parents order schools according to their effectiveness. The average treatment effect of first-choice schools would improve from 0.07σ to 0.43σ if parents ranked schools based on effectiveness, and the average match effect would increase from -0.04σ to 0.16σ. This implies that the average student loses more than one-half of a standard deviation in math achievement by enrolling in her first-choice school rather than the most effective option.

The statistics in Table 11 suggest that if information frictions prevent parents from ranking schools based on effectiveness, providing information about school effectiveness could alter school choices considerably. These changes may be particularly valuable for disadvantaged students. As shown in online Appendix Table A12, gaps in effectiveness between observed first-choice schools and achievement-maximizing choices are larger for students with lower baseline achievement. This is driven by the stronger relationship between peer quality and preferences for more-advantaged parents documented in Table 10. These results suggest reducing information barriers could lead to differential increases in school quality for disadvantaged students and reduce inequality in student achievement. On the other hand, the patterns documented here may also reflect parents' valuation of school amenities other than academic effectiveness rather than a lack of information about treatment effects. It is also important to note that our estimates capture average impacts of changing an individual student's rankings, holding fixed the behavior of other students; we might expect schools' treatment effect parameters to change if all students changed behavior simultaneously due to changes in peer effects or other inputs.

Regardless of why parents respond to peer quality rather than school effectiveness, our results have important implications for the incentive effects of school choice programs. Since parents only respond to the component of school average outcomes that can be predicted by the ability of enrolled students, our estimates imply a school wishing to boost its popularity must recruit better students; improving outcomes by increasing causal effectiveness for a fixed set of students will have no impact on parent demand. Our results therefore suggest that choice may create incentives for schools to invest in screening and selection.

The evolution of admissions criteria used at New York City's high schools is consistent with the implication that schools have an increased incentive to screen applicants due to parents' demand for high-ability peers. After the first year of the new assignment mechanism, several school programs eliminated all lottery-based admissions procedures and became entirely screened. In the 2003–2004 high school brochure, 36.8 percent of programs are screened, and this fraction jumps to 40.3 percent two years later. The Beacon High School in Manhattan, for example, switched from a school where one-half of the seats were assigned via random lottery in 2003–2004 to a screened school the following year, where admissions is based on test performance, an interview, and a portfolio of essays. Leon Goldstein High School for Sciences in Brooklyn underwent a similar transition. Both high

Table 11—Potential Achievement Gains from Ranking Schools by Effectiveness

	Observed rankings			Rankings based on effectiveness		
	Peer quality (1)	ATE (2)	Match (3)	Peer quality (4)	ATE (5)	Match (6)
Choice 1	0.112	0.071	−0.037	0.286	0.427	0.162
Choice 2	0.057	0.055	−0.020	0.182	0.352	0.108
Choice 3	0.021	0.045	−0.012	0.087	0.275	0.113
Choice 4	−0.013	0.036	−0.006	0.105	0.247	0.103
Choice 5	−0.046	0.027	−0.002	0.124	0.228	0.092
Choice 6	−0.074	0.019	−0.001	0.103	0.209	0.085
Choice 7	−0.097	0.014	0.001	0.118	0.197	0.075
Choice 8	−0.114	0.012	0.001	0.099	0.169	0.066
Choice 9	−0.127	0.007	0.001	0.064	0.333	0.111
Choice 10	−0.139	0.004	0.003	0.046	0.165	0.063
Choice 11	−0.146	0.003	0.003	0.028	0.157	0.056
Choice 12	−0.156	−0.002	0.002	0.013	0.146	0.053

Notes: This table summarizes Regents math score gains that parents could achieve by ranking schools based on effectiveness. Columns 1–3 report average peer quality, average treatment effects, and average match quality for students' observed preference rankings. Columns 4–6 display corresponding statistics for hypothetical rankings that list schools in order of their treatment effects. Treatment effect estimates come from control function models. All calculations are restricted to ranked schools within the home borough.

schools frequent lists of New York City's best public high schools.[15] Compared to the first years of the new system, there has also been growth in the number of limited unscreened programs, which use a lottery but also give priority to students who attend an open house or high school fair. Compared to unscreened programs, prioritizing applicants who attend an information session provides an ordeal that favors applicants with time and resources thus resulting in positive selection.[16] The number of limited unscreened programs nearly doubled from 106 to 210 from 2005 to 2012 (Nathanson, Corcoran, and Baker-Smith 2013).

VII. Conclusion

A central motivation for school choice programs is that parents' choices generate demand-side pressure for improved school productivity. We investigate this possibility by comparing estimates of school popularity and treatment effects based on rank-ordered preference data for applicants to public high schools in New York City. Parents prefer schools that enroll higher-achieving peers. Conditional on peer quality, however, parents' choices are unrelated to causal school effectiveness. Moreover, no subgroup of parents systematically responds to causal school effectiveness. We also find no relationship between preferences for schools and estimated match quality. This indicates that choice does not lead students to sort into schools on the basis of comparative advantage in academic achievement.

[15] Mary Kay Linge and Joshua Tanzer, "The Top 40 Public High Schools in NYC," *New York Post*, September 17, 2016, https://nypost.com/2016/09/17/the-top-40-public-high-schools-in-nyc/.

[16] Monica Disare, "City to Eliminate High School Admissions Method That Favored Families with Time and Resources," *Chalkbeat*, June 6, 2017, https://www.chalkbeat.org/posts/ny/2017/06/06/city-to-eliminate-high-school-admissions-method-that-favored-families-with-time-and-resources/ (accessed December 2017).

This pattern of findings has important implications for the expected effects of school choice programs. Our results on match quality suggest choice is unlikely to increase allocative efficiency. Our findings regarding peer quality and average treatment effects suggest choice may create incentives for increased screening rather than academic effectiveness. If parents respond to peer quality but not causal effects, a school's easiest path to boosting its popularity is to improve the ability of its student population. Since peer quality is a fixed resource, this creates the potential for socially costly zero-sum competition as schools invest in mechanisms to attract the best students. MacLeod and Urquiola (2015) argues that restricting a school's ability to select pupils may promote efficiency when student choices are based on school reputation. The impact of school choice on effort devoted to screening is an important empirical question for future research.

While we have shown that parents do not choose schools based on causal effects for a variety of educational outcomes, we cannot rule out the possibility that preferences are determined by effects on unmeasured outcomes. Our analysis also does not address why parents put more weight on peer quality than on treatment effects. If parents rely on student composition as a proxy for effectiveness, coupling school choice with credible information on causal effects may strengthen incentives for improved productivity and weaken the association between preferences and peer ability. Distinguishing between true tastes for peer quality and information frictions is another challenge for future work.

REFERENCES

Abdulkadiroğlu, Atila, Nikhil Agarwal, and Parag A. Pathak. 2017. "The Welfare Effects of Coordinated Assignment: Evidence from the New York City High School Match." *American Economic Review* 107 (12): 3635–89.

Abdulkadiroğlu, Atila, Joshua D. Angrist, Yusuke Narita, and Parag A. Pathak. 2017. "Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation." *Econometrica* 85 (5): 1373–1432.

Abdulkadiroğlu, Atila, Joshua D. Angrist, Yusuke Narita, and Parag A. Pathak. 2019. "Breaking Ties: Regression Discontinuity Design Meets Market Design." Cowles Foundation Discussion Paper 2170.

Abdulkadiroğlu, Atila, Joshua D. Angrist, and Parag A. Pathak. 2014. "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools." *Econometrica* 82 (1): 137–96.

Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth. 2005. "The New York City High School Match." *American Economic Review* 95 (2): 364–67.

Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth. 2009. "Strategy-Proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match." *American Economic Review* 99 (5): 1954–78.

Abdulkadiroğlu, Atila, Parag A. Pathak, Jonathan Schellenberg, and Christopher R. Walters. 2020. "Replication Data for: Do Parents Value School Effectiveness?" American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E112266V1.

Abdulkadiroğlu, Atila, Parag A. Pathak, and Christopher R. Walters. 2018. "Free to Choose: Can School Choice Reduce Student Achievement?" *American Economic Journal: Applied Economics* 10 (1): 175–206.

Agarwal, Nikhil, and Paulo Somaini. 2018. "Demand Analysis Using Strategic Reports: An Application to a School Choice Mechanism." *Econometrica* 86 (2): 391–444.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113 (1): 151–84.

Altonji, Joseph G., Ching-I Huang, and Christopher R. Taber. 2015. "Estimating the Cream Skimming Effect of School Choice." *Journal of Political Economy* 123 (2): 266–324.

Angrist, Joshua D., Sarah R. Cohodes, Susan M. Dynarski, Parag A. Pathak, and Christopher R. Walters. 2016a. "Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry, and Choice." *Journal of Labor Economics* 34 (2): 275–318.

Angrist, Joshua D., Peter D. Hull, Parag A. Pathak, and Christopher R. Walters. 2016b. "Interpreting Tests of School VAM Validity." *American Economic Review* 106 (5): 388–92.

Angrist, Joshua D., Peter D. Hull, Parag A. Pathak, and Christopher R. Walters. 2017. "Leveraging Lotteries for School Value-Added: Testing and Estimation." *Quarterly Journal of Economics* 132 (2): 871–919.

Angrist, Joshua D., Parag A. Pathak., and Christopher R. Walters. 2013. "Explaining Charter School Effectiveness." *American Economic Journal: Applied Economics* 5 (4): 1–27.

Avery, Christopher, and Parag A. Pathak. 2015. "The Distributional Consequences of Public School Choice." NBER Working Paper 21525.

Bacher-Hicks, Andrew, Thomas J. Kane, and Douglas O. Staiger. 2014. "Validating Teacher Effect Estimates Using Changes in Teacher Assignments in Los Angeles." Unpublished.

Barseghyan, Levon, Damon Clark, and Stephen Coate. 2014. "Public School Choice: An Economic Analysis." NBER Working Paper 20701.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy* 115 (4): 588–638.

Beuermann, Diether, C. Kirabo Jackson, Laia Navarro-Sola, and Francisco Pardo. 2018. "What Is a Good School, and Can Parents Tell? Evidence on the Multidimensionality of School Output." NBER Working Paper 25342.

Black, Sandra E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *Quarterly Journal of Economics* 114 (2): 577–99.

Blinder, Alan S. 1973. "Wage Discrimination: Reduced Form and Structural Estimates." *Journal of Human Resources* 8 (4): 436–55.

Blundell, Richard, and Rosa L. Matzkin. 2014. "Control Functions in Nonseparable Simultaneous Equations Models." *Quantitative Economics* 5 (2): 271–95.

Booker, Kevin, Tim R. Sass, Brian Gill, and Ron Zimmer. 2011. "The Effects of Charter High Schools on Educational Attainment." *Journal of Labor Economics* 29 (2): 377–415.

Burgess, Simon, Ellen Greaves, Anna Vignoles, and Deborah Wilson. 2015. "What Parents Want: School Preferences and School Choice." *Economic Journal* 125 (587): 1262–89.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2011. "Robust Inference with Multiway Clustering." *Journal of Business and Economic Statistics* 29 (2): 238–49.

Card, David. 1995. "Using Geographic Variation in College Proximity to Estimate the Return to Schooling." In *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, edited by Louis N. Christofides, E. Kenneth Grant, and Robert Swidinsky. Toronto: University of Toronto Press.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *Quarterly Journal of Economics* 126 (4): 1593–1660.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014a. "Measuring the Impact of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104 (9): 2593–2632.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014b. "Measuring the Impact of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review* 104 (9): 2633–79.

Chetty, Raj, John N. Friedman, and Jonah Rockoff. 2016. "Using Lagged Outcomes to Evaluate Bias in Value-Added Models." *American Economic Review* 106 (5): 393–99.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2017. "Measuring the Impacts of Teachers: Reply." *American Economic Review* 107 (6): 1685–1717.

Chetty, Raj, John N. Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. 2017. "Mobility Report Cards: The Role of Colleges in Intergenerational Mobility." The Equality of Opportunity Project, January.

Chetty, Raj, and Nathaniel Hendren. 2017. "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates." NBER Working Paper 23002.

Chubb, John E., and Terry M. Moe. 1990. *Politics, Markets, and America's Schools.* Washington, DC: Brookings Institution Press.

Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg. 2016. "Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance." Unpublished.

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

Dahl, Gordon B. 2002. "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets." *Econometrica* 70 (6): 2367–2420.

Dale, Stacy Berg, and Alan 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.

Dale, Stacy Berg, and Alan B. Krueger. 2014. "Estimating the Effects of College Characteristics over the Career Using Administrative Earnings Data." *Journal of Human Resources* 49 (2): 323–58.

Deming, David J. 2014. "Using School Choice Lotteries to Test Measures of School Effectiveness." *American Economic Review* 104 (5): 406–11.

Deming, David J., Justine S. Hastings, Thomas J. Kane, and Douglas O. Staiger. 2014. "School Choice, School Quality, and Postsecondary Attainment." *American Economic Review* 104 (3): 991–1013.

DeVos, Betsy. 2017. Comments at Brookings Institution event on the 2016 Education Choice and Competition Index, March 29. https://www.brookings.edu/events/the-2016-education-choice-and-competition-index.

Dobbie, Will, and Roland G. Fryer. 2014. "The Impact of Attending a School with High-Achieving Peers: Evidence from the New York City Exam Schools." *American Economic Journal: Applied Economics* 6 (3): 58–75.

Dubin, Jeffrey A., and Daniel L. McFadden. 1984. "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption." *Econometrica* 52 (2): 345–62.

Dubins, L. E., and D. A. Freedman. 1981. "Machiavelli and the Gale-Shapley Algorithm." *American Mathematical Monthly* 88 (7): 485–94.

Dynarski, Susan, Joshua Hyman, and Diane Whitmore Schanzenbach. 2013. "Experimental Evidence on the Effect of Childhood Investments on Postsecondary Attainment and Degree Completion." *Journal of Policy Analysis and Management* 32 (4): 692–717.

Epple, Dennis, David N. Figlio, and Richard Romano. 2004. "Competition between Private and Public Schools: Testing Stratification and Pricing Predictions." *Journal of Public Economics* 88 (7–8): 1215–45.

Epple, Dennis, and Richard Romano. 1998. "Competition between Private and Public Schools, Vouchers, and Peer-Group Effects." *American Economic Review* 88 (1): 33–62.

Epple, Dennis, and Richard Romano. 2008. "Educational Vouchers and Cream Skimming." *International Economic Review* 49 (4): 1395–1435.

Fack, Gabrielle, Julien Grenet, and Yinghua He. 2015. "Beyond Truth-Telling: Preference Estimation with Centralized School Choice." Paris School of Economics Working Paper.

Figlio, David N., and Maurice E. Lucas. 2004. "What's in a Grade? School Report Cards and the Housing Market." *American Economic Review* 94 (3): 591–604.

Finkelstein, Amy, Matthew Gentzkow, Peter Hull, and Heidi Williams. 2017. "Adjusting Risk Adjustment: Accounting for Variation in Diagnostic Intensity." *New England Journal of Medicine* 376: 608–10.

French, Eric, and Christopher Taber. 2011. "Identification of Models of the Labor Market." In *Handbook of Labor Economics*, Vol. 4A, edited by Orley Ashenfelter and David Card, 537–617. Amsterdam: Elsevier.

Friedman, Milton. 1962. *Capitalism and Freedom.* Chicago: University of Chicago Press.

Gale, D., and L. S. Shapley. 1962. "College Admissions and the Stability of Marriage." *American Mathematical Monthly* 69 (1): 9–15.

Glazerman, Steven, and Dallas Dotter. 2016. "Market Signals: Evidence on the Determinants and Consequences of School Choice from a Citywide Lottery." Mathematica Policy Research Working Paper 45.

Guarino, Cassandra M., Mark D. Reckase, and Jeffrey M. Wooldridge. 2015. "Can Value-Added Measures of Teacher Performance Be Trusted?" *Education Finance and Policy* 10 (1): 117–56.

Haeringer, Guillaume, and Flip Klijn. 2009. "Constrained School Choice." *Journal of Economic Theory* 144 (5): 1921–47.

Hanushek, Eric A. 1981. "Throwing Money at Schools." *Journal of Policy Analysis and Management* 1 (1): 19–41.

Harris, Douglas N., and Matthew Larsen. 2014. "What Schools Do Families Want (And Why)?" Technical report, Education Research Alliance for New Orleans.

Hastings, Justine, Ali Hortaçsu, and Chad Syverson. 2017. "Sales Force and Competition in Financial Product Markets: The Case of Mexico's Social Security Privatization." *Econometrica* 85 (6): 1723–61.

Hastings, Justine S., Thomas J. Kane, and Douglas O. Staiger. 2009. "Heterogeneous Preferences and the Efficacy of Public School Choice." Unpublished.

Hastings, Justine S., and Jeffrey M. Weinstein. 2008. "Information, School Choice, and Academic Achievement: Evidence from Two Experiments." *Quarterly Journal of Economics* 123 (4): 1373–1414.

Hausman, Jerry A. 1983. "Specification and Estimation of Simultaneous Equation Models." In *Handbook of Econometrics*, Vol. 1, edited by Zvi Griliches and Michael D. Intriligator, 391–448. New York: Elsevier.

Hausman, Jerry A., and Paul A. Ruud. 1987. "Specifying and Testing Econometric Models for Rank-Ordered Data." *Journal of Econometrics* 34 (1–2): 83–104.

Heckman, James J., and Richard Robb. 1985. "Alternative Methods for Evaluating the Impact of Interventions: An Overview." *Journal of Econometrics* 30 (1–2): 239–67.

Heckman, James J., Sergio Urzua, and Edward Vytlacil. 2006. "Understanding Instrumental Variables in Models with Essential Heterogeneity." *Review of Economics and Statistics* 88 (3): 389–432.

Heckman, James J., Sergio Urzua, and Edward Vytlacil. 2008. "Instrumental Variables in Models with Multiple Outcomes: The General Unordered Case." *Annales d'Économie et de Statistique* (91–92): 151–74.

Hoxby, Caroline M. 2000. "Does Competition among Public Schools Benefit Students and Taxpayers?" *American Economic Review* 90 (5): 1209–38.

Hoxby, Caroline M. 2003. "School Choice and School Productivity: Could School Choice Be a Tide That Lifts All Boats?" In *The Economics of School Choice*, edited by Caroline M. Hoxby. Chicago: University of Chicago Press.

Hsieh, Chang-Tai, and Miguel Urquiola. 2006. "The Effects of Generalized School Choice on Achievement and Stratification: Evidence from Chile's Voucher Program." *Journal of Public Economics* 90 (8–9): 1477–1503.

Hull, Peter. 2016. "Estimating Hospital Quality with Quasi-Experimental Data." Unpublished.

Imberman, Scott A., and Michael F. Lovenheim. 2016. "Does the Market Value Value-Added? Evidence from Housing Prices after a Public Release of School and Teacher Value-Added." *Journal of Urban Economics* 91: 104–21.

Jacob, Brian A., and Lars Lefgren. 2007. "What Do Parents Value in Education? An Empirical Investigation of Parents' Revealed Preferences for Teachers." *Quarterly Journal of Economics* 122 (4): 1603–37.

Jacob, Brian A., and Lars Lefgren. 2008. "Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education." *Journal of Labor Economics* 26 (1): 101–36.

Kane, Thomas J., Daniel F. McCaffrey, Trey Miller, and Douglas O. Staiger. 2013. "Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment." Gates Foundation Report.

Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. 2008. "What Does Certification Tell Us about Teacher Effectiveness? Evidence from New York City." *Economics of Education Review* 27 (6): 615–31.

Kane, Thomas J., and Douglas O. Staiger. 2002. "The Promise and Pitfalls of Using Imprecise School Accountability Measures." *Journal of Economic Perspectives* 16 (4): 91–114.

Kapor, Adam, Christopher A. Neilson, and Seth D. Zimmerman. 2017. "Heterogeneous Beliefs and School Choice Mechanisms." Unpublished.

Kirkeboen, Lars J., Edwin Leuven, and Magne Mogstad. 2016. "Field of Study, Earnings, and Self-Selection." *Quarterly Journal of Economics* 131 (3): 1057–1111.

Kline, Patrick. 2011. "Oaxaca-Blinder as a Reweighting Estimator." *American Economic Review* 101 (3): 532–37.

Kline, Patrick, and Andres Santos. 2012. "A Score Based Approach to Wild Bootstrap Inference." *Journal of Econometric Methods* 1 (1): 23–41.

Kline, Patrick, and Christopher R. Walters. 2016. "Evaluating Public Programs with Close Substitutes: The Case of Head Start." *Quarterly Journal of Economics* 131 (4): 1795–1848.

Koedel, Cory, Kata Mihaly, and Jonah E. Rockoff. 2015. "Value-Added Modeling: A Review." *Economics of Education Review* 47: 180–95.

Kolesár, Michal. 2013. "Estimation in an Instrumental Variables Model with Treatment Effect Heterogeneity." Unpublished.

Ladd, Helen F. 2002. "Market-Based Reforms in Urban Education." Unpublished.

Langer, Ashley. 2016. "(Dis)Incentives for Demographic Price Discrimination in the New Vehicle Market." Unpublished.

Lee, Lung-Fei. 1983. "Generalized Econometric Models with Selectivity." *Econometrica* 51 (2): 507–12.

MacLeod, W. Bentley, Evan Riehl, Juan E. Saavedra, and Miguel Urquiola. 2017. "The Big Sort: College Reputation and Labor Market Outcomes." *American Economic Journal: Applied Economics* 9 (3): 223–61.

MacLeod, W. Bentley, and Miguel Urquiola. 2015. "Reputation and School Competition." *American Economic Review* 105 (11): 3471–88.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies* 60 (3): 531–42.

Morris, Carl N. 1983. "Parametric Empirical Bayes Inference: Theory and Applications." *Journal of the American Statistical Association* 78 (381): 47–55.

Mountjoy, Jack. 2017. "Community Colleges and Upward Mobility." Unpublished.

Nathanson, Lori, Sean P. Corcoran, and Christine Baker-Smith. 2013. "High School Choice in New York City: A Report on the School Choices and Placements of Low-Achieving Students." Research Alliance for New York City Schools, April.

Neal, Derek. 1997. "The Effects of Catholic Secondary Schooling on Educational Achievement." *Journal of Labor Economics* 15 (1): 98–123.

New York City Department of Education. 2003. "Directory of the New York City Public High Schools, 2003–2004." New York.

New York City Department of Education. 2004. "2003–2004 Annual School Reports." New York.

New York City Department of Education. 2008. "NYC School Survey." https://opendata.cityofnewyork.us.

New York City Department of Education. 2013. "Student Administrative Data System." New York.

New York City Department of Education. 2017. "2016–2017 School Quality Reports." http://schools.nyc.gov/accountability/tools/report/default.htm.

New York State Department of Education. 2007a. "School Report Card Database." https://data.nysed.gov/downloads.php.

New York State Department of Education. 2007b. "School Safety and Educational Climate Reports." http://www.p12.nysed.gov/irs/school_safety/vadir_archive.html.

Oaxaca, Ronald. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review* 14 (3): 693–709.

Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business and Economic Statistics* 37 (2): 187–204.

Pathak, Parag A., and Tayfun Sönmez. 2013. "School Admissions Reform in Chicago and England: Comparing Mechanisms by Their Vulnerability to Manipulation." *American Economic Review* 103 (1): 80–106.

Peterson, Paul E., and David E. Campbell. 2001. *Charters, Vouchers, and Public Education.* Washington, DC: Brookings Institution Press.

Robbins, Herbert. 1956. "An Empirical Bayes Approach to Statistics." In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability,* Vol. 1, edited by Jerzy Neyman, 157–63. Berkeley: University of California Press.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1): 41–55.

Roth, Alvin E. 1982. "The Economics of Matching: Stability and Incentives." *Mathematics of Operations Research* 7 (4): 617–28.

Rothstein, Jesse M. 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): 1333–50.

Rothstein, Jesse M. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *Quarterly Journal of Economics* 125 (1): 175–214.

Rothstein, Jesse M. 2017. "Measuring the Impacts of Teachers: Comment." *American Economic Review* 107 (6): 1656–84.

Roy, A. D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3 (2): 135–46.

Staiger, Douglas, and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (3): 557–86.

Todd, Petra E., and Kenneth I. Wolpin. 2003. "On the Specification and Estimation of the Production Function for Cognitive Achievement." *Economic Journal* 113 (485): F3–F33.

Walters, Christopher R. 2018. "The Demand for Effective Charter Schools." *Journal of Political Economy* 126 (6): 2179–2223.

Wooldridge, Jeffrey M. 2015. "Control Function Methods in Applied Econometrics." *Journal of Human Resources* 50 (2): 420–45.