SCHOOL EFFECTIVENESS & INEQUALITY INITIATIVE

Discussion Paper #2017.04

Do Parents Value School Effectiveness?

Atila Abdulkadiroglu
Parag A. Pathak
Jonathan Schellenberg
Christopher R. Walters

October 2017

MIT Department of Economics
77 Massachusetts Avenue, Bldg E52-300
Cambridge, MA 02139

National Bureau of Economic Research
1050 Massachusetts Avenue, 3rd Floor
Cambridge, MA 02138

Do Parents Value School Effectiveness?
Atila Abdulkadiroglu, Parag A. Pathak, Jonathan Schellenberg, and Christopher R. Walters
SEII Discussion Paper #2017.04
October 2017

ABSTRACT

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 and high school graduation. We also estimate impacts on 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. We find no relationship between preferences and school effectiveness after controlling for peer quality.

* We gratefully acknowledge funding from the National Science Foundation and the W. T. Grant Foundation. Thanks also go to 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, 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, and the University of Chicago Interactions Conference for suggestions and comments. We're grateful to Eryn Heying for invaluable administrative support. Abdulkadiroglu and Pathak are Scientific Advisory Board members of the Institute for Innovation in Public School Choice.

1 Introduction

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). This argument has motivated recent policy efforts to expand school choice (e.g., DeVos, 2017).

If choice is to improve educational effectiveness, however, 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 school attributes such as facilities, convenience, or peer composition in a manner that does not align with educational impacts (Hanushek, 1981). Moreover, 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 parent 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 et al., 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 et al., 2015). Recent evidence suggests that value-added models controlling only for observables provide quantitatively useful but biased estimates of causal effects due to selection on

[1] As we discuss in Section 2, 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 et al. (2015) propose empirical approaches to measuring student preferences without requiring that truth-telling is the unique equilibrium.

unobservables (Rothstein, 2010, 2017; Chetty et al., 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.

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, idiosyncratic match effects, and peer quality, defined as the component of a school's average outcome due to selection rather than effectiveness.

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. 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. Together, our findings imply that parents' choices tend to 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 than for improved school productivity.

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 et al., 2007; Hastings and Weinstein, 2008; Burgess et al., 2014; 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 et al., 2004; Hsieh and Urquiola, 2006; Barseghyan et al., 2014; Altonji et al., 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. Walters (forthcoming) documents that disadvantaged students in Boston are less likely to apply to charter schools than more advantaged students despite experiencing larger achievement benefits, while Angrist et al. (2013) and Abdulkadiroğlu et al. (2017) 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 achievement gains.

Finally, our analysis adds to a recent series of studies leveraging preference data from centralized school assignment mechanisms to investigate school demand (Hastings et al., 2009; Harris and Larsen, 2014; Fack et al., 2015; Abdulkadiroğlu et al., 2016; Glazerman and Dotter, 2016; Kapor et al., 2017; Agarwal and Somaini, forthcoming). Some of these studies analyze assignment mechanisms that provide incentives to strategically misreport preferences, while others measure school 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 success.

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 3 develops a conceptual framework for analyzing school effectiveness and peer quality, and Section 4 details our empirical approach. Section 5 summarizes estimated distributions of student preferences and school treatment effects. Section 6 links preferences to peer quality and school effectiveness. Section 7 concludes and discusses some directions for future research.

2 Setting and Data

2.1 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 et al., 2005, 2009). Abdulkadiroğlu et al. (2016) show that introducting coordinated assignment reduced the share of administratively assigned students and likely improved average household 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. Programs prioritize applicants 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 et al., 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 et al., 2016).

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, 2017). In the analysis to follow, we therefore interpret students' rank-ordered lists as truthful reports of their preferences.[2]

2.2 Data and Samples

The data used here are extracted from a New York City Department of Education (DOE) administrative information system covering all students enrolled in New York City public schools between the 2003-2004 and 2012-2013 school years. 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

[2] Abdulkadiroğlu et al. (2016) report that preference estimates using only the top ranked school, the top three schools, and all but the last ranked school are similar.

City public schools in eighth grade between 2003-2004 and 2006-2007. This set of students is used to construct several subsamples 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 additional samples corresponding to each outcome of interest. These outcome 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. Appendix 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 one if a student graduates within five years of her projected high school entry date given her eighth grade cohort. Likewise, college attendance equals one 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. (2017b), 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 an accurate 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, 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 shows 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 one 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.

2.3 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 half submit eight 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 et al. (2016) report that for 2003-04, 193 programs restricted eligibility to applicants who reside in the same borough. The preference analysis 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 zero and standard deviation one 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.

3 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:

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

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 $Y_i = \sum_j 1\{S_i = j\} Y_{ij}$, where $S_i \in \{1...J\}$ denotes school attendance.

We decompose potential outcomes into components explained by student ability, school effectiveness, and idiosyncratic factors. Let $A_i \equiv (1/J)\sum_j Y_{ij}$ 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 $X_i'(\beta_j - \bar{\beta})$) or because of unobserved factors that make student i more or less suitable for school j (captured by $\epsilon_{ij} - \bar{\epsilon}_i$).

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

$$E\left[Y_i | S_i = j\right] = Q_j + ATE_j + E\left[M_{ij} | S_i = j\right]. \tag{3}$$

Here $Q_j \equiv E\left[A_i | S_i = j\right]$ is the average ability of students enrolled at school j, a variable we label "peer quality." The quantity $E\left[M_{ij} | S_i = j\right]$ 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 this average match effect to be positive for all schools. 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.

4 Empirical Methods

The goal of our empirical analysis is to assess the roles of peer quality, school effectiveness, and idiosyncratic matching 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.

4.1 Estimating Preferences

Let U_{ij} denote student i's utility from enrolling in school j, and let $\mathcal{J} = \{1...J\}$ 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. U_{ij} may therefore 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

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

Student i's rank-order list is then $R_i = (R_{i1}...R_{i\ell(i)})'$, 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 et al., 2009 and Abdulkadiroğlu et al., 2016). We model unobserved tastes η_{ij} as following independent extreme value type I distributions conditional on X_i and $D_i = (D_{i1}...D_{iJ})'$. 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:

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

We allow flexible heterogeneity in tastes by estimating preference models separately for 360 covariate cells defined by the intersection of borough, sex, race (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 et al., forthcoming 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.

4.2 Estimating School Effectiveness

Our analysis of school effectiveness aims to recover the parameters of the potential outcome equations defined in Section 3. We take two approaches to estimating these parameters.

Approach 1: Selection on observables

The first set of estimates is based on the assumption:

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

This restriction, often labeled "selection on observables," requires potential outcomes to be mean-independent of high school enrollment 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 generalization of the Oaxaca-Blinder (1973) treatment effects estimator (Kline, 2011).[3]

OLS "value-added" models that control for lagged test scores are widely used to estimate the contributions of teachers and schools to student achievement (Koedel et al., 2015). The credibility of the selection on observables assumption underlying these models is a matter of continuing debate (Rothstein, 2010, 2017; Kane et al., 2013; Baicher-Hicks et al., 2014; Chetty et al., 2014a, 2016, 2017a; Guarino et al., 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 et al., 2014a). We therefore report OLS value-added 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:

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

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

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

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 that apply to the same schools. Similarly, (6) implies that 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:

$$E\left[Y_{ij}|X_i, D_i, \eta_{i1}...\eta_{iJ}, S_i\right] = \alpha_j + X_i'\beta_j + D_i'\gamma + \sum_{k=1}^{J} \psi_k \left(\eta_{ik} - \mu_\eta\right) + \varphi(\eta_{ij} - \mu_\eta),\ \ j = 1...J, \tag{7}$$

where $\mu_\eta \equiv E\left[\eta_{ij}\right]$ 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 et al., 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 match effects in student outcomes (Heckman et al., 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:

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

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

[5]As discussed in Section 6, we also estimate a model that replaces the control functions with fixed effects for first choice schools.

where $\lambda_k(X_i, D_i, R_i) \equiv E[\eta_{ik} - \mu_\eta | X_i, D_i, R_i]$ 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 mean outcomes for selection on unobservables (Heckman and Robb, 1985; Blundell and Matzkin, 2014; Wooldridge, 2015). As shown in Appendix 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 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] To account for estimation error in the control functions, we conduct inference via a two-step extension of the score bootstrap procedure of Kline and Santos (2012). As detailed in Appendix 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 $\hat{Q}_j = \sum_i 1\{S_i = j\}\hat{A}_i / \sum_i 1\{S_i = j\}$, the average predicted ability of enrolled students.

The end result of our school quality estimation procedure is a vector of estimates for each school, $\hat{\theta}_j = (\hat{\alpha}_j, \hat{\beta}_j', \hat{Q}_j)'$. 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:

[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 zero for these students. We similarly code D_{ik} to zero 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.

$$\hat{\theta}_j|\theta_j \sim N(\theta_j, \Omega_j),$$
$$\theta_j \sim N(\mu_\theta, \Sigma_\theta). \quad (10)$$

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 empirical Bayes (EB) estimates characterize the joint distribution of peer quality and school treatment effect parameters, purged of the estimation error in $\hat{\theta}_j$.

This hierarchical model can also be used to improve estimates of parameters for individual schools. An 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 et al., 2014a; Hull, 2016; Angrist et al., 2017; Chetty and Hendren, 2017; Finkelstein et al., 2017). Appendix 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 $\hat{\theta}_j$ as a regressor (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.

4.3 Linking Preferences to School Effectiveness

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

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

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 et al. (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.

[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.

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, overall school effectiveness, and match effects.

5 Parameter Estimates

5.1 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 deviated 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 overdispersion in estimated preference parameters due to sampling error 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 fixed j, is 0.50, equivalent to about 1.5 (0.50/0.33) miles of distance. The larger within-cell standard deviation indicates that students in different subgroups tend to prefer the same schools.

5.2 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 zero 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 $0.15 - 0.2\sigma$; 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 et al., 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 (0.10σ), subsidized lunch status (0.06σ), 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.[8]

Our estimates imply that high-ability students tend to enroll in more effective schools. Table 5 reports correlations between Q_j and school treatment effect parameters based on control function estimates for Regents math scores. Corresponding value-added estimates appear in 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 et al., 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 elements of the slope coefficient vector 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

[8]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 one percent of a standard deviation ($0.005\sigma \times 2.0 = 0.01\sigma$).

with evidence from Deming et al. (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 more high-achieving students and produce larger gains on average 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. Section 6 presents a more systematic investigation of this pattern by documenting the net relationship between preferences and treatment effects combining all student characteristics.

Patterns of estimates for PSAT scores, high school graduation, college attendance, and college quality are generally similar to results for Regents math scores. Appendix Tables A3-A6 present estimated distributions of peer quality and school effectiveness for these additional 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 et al., 2013; Angrist et al., 2016a).

5.3 Decomposition of School Average Outcomes

We summarize the joint distribution of peer quality and school effectiveness by implementing the decomposition introduced in Section 3. Table 6 uses the control function estimates to decompose variation in school averages for each outcome into 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 45 percent (PSAT scores) to 83 percent (high school graduation), while the importance of average treatment effects ranges from 15 percent (PSAT scores) to 19 percent (log college quality).

Despite the significant variation in slope coefficients documented in Table 4, we find match effects have a limited role in explaining dispersion in school average outcomes. The variance of match effects accounts for only five 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.

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.

6 Preferences, Peer Quality, and School Effectiveness

6.1 Productivity vs. Peers

The last step of our analysis compares the relative strength of peer quality and school effectiveness as predictors of parent preferences. Table 7 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 one standard deviation increase in Q_j or ATE_j.

Our estimates show that while preferences are positively correlated with school effectiveness, this relationship is entirely explained by peer quality. Results based on the value-added model, reported in columns (1) and (2), imply that a one standard deviation increase in Q_j is associated with a 0.42 standard deviation increase in mean utility, while a one standard deviation increase in ATE_j is associated with a 0.25 standard deviation increase in mean utility. Column (3) 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 one 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; in fact, the control function results show a small, marginally statistically significant negative association between school effectiveness and popularity after controlling for peer quality.[9]

[9] Appendix Table A7 probes the robustness of these results to alternative controls for preference rankings by replacing the control functions in equation (8) with fixed effects for first choice schools. Appendix Table A8 investigates the robustness of our results to alternative measures of school popularity by 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 specifications produce similar results to those reported in Table 7.

Columns (4) and (8) of Table 7 explore the role of treatment effect heterogeneity by adding posterior mean predictions of match effects to equation (11), also scaled in standard deviation units of the distribution of match effects across schools and cells. 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 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, forthcoming).

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.

Together, the results in Table 7 and Figure 3 indicate that, among schools with similar student populations, parents do not rank more effective schools more favorably. This pattern may reflect a lack of interest in causal effects on academic outcomes or a lack of knowledge about these effects. Without direct information about school effectiveness, for example, parents may use peer characteristics as a proxy for school quality. In either case, however, these results have 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, a school wishing to boost its popularity must recruit better students. Our estimates imply that increasing average outcomes through improved causal effectiveness for a fixed set of students will have no effect on parent demand.

6.2 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 8 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 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 Kirkebøen et al. (2016), which show sorting into fields of study based on heterogeneous earnings gains. New York City's high school students do not respond to match quality in academic outcomes.

6.3 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 when choosing schools (Hastings et al., 2009). This pattern may reflect either greater responsiveness to peer quality or more sensitivity to causal school effectiveness. To distinguish between these possibilities, Table 9 estimates equation (11) separately by sex, race, subsidized lunch status, and baseline test score tercile.

We find that no subgroup of households responds to causal school effectiveness. Consistent with previous work, we find larger coefficients on peer quality among non-minority 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-level clustering of preferences will lead to a positive association between students' rankings and the enrollment share of others in the same group.[10] The key pattern in Table 9 is that, among schools with similar peer quality, no group prefers schools with greater causal impacts on academic achievement. This suggests that parent demand may yield limited incentives for improved causal effectiveness even among more-advantaged households.

7 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. Moreoever, no subgroup of parents systematically responds to causal school effectiveness.

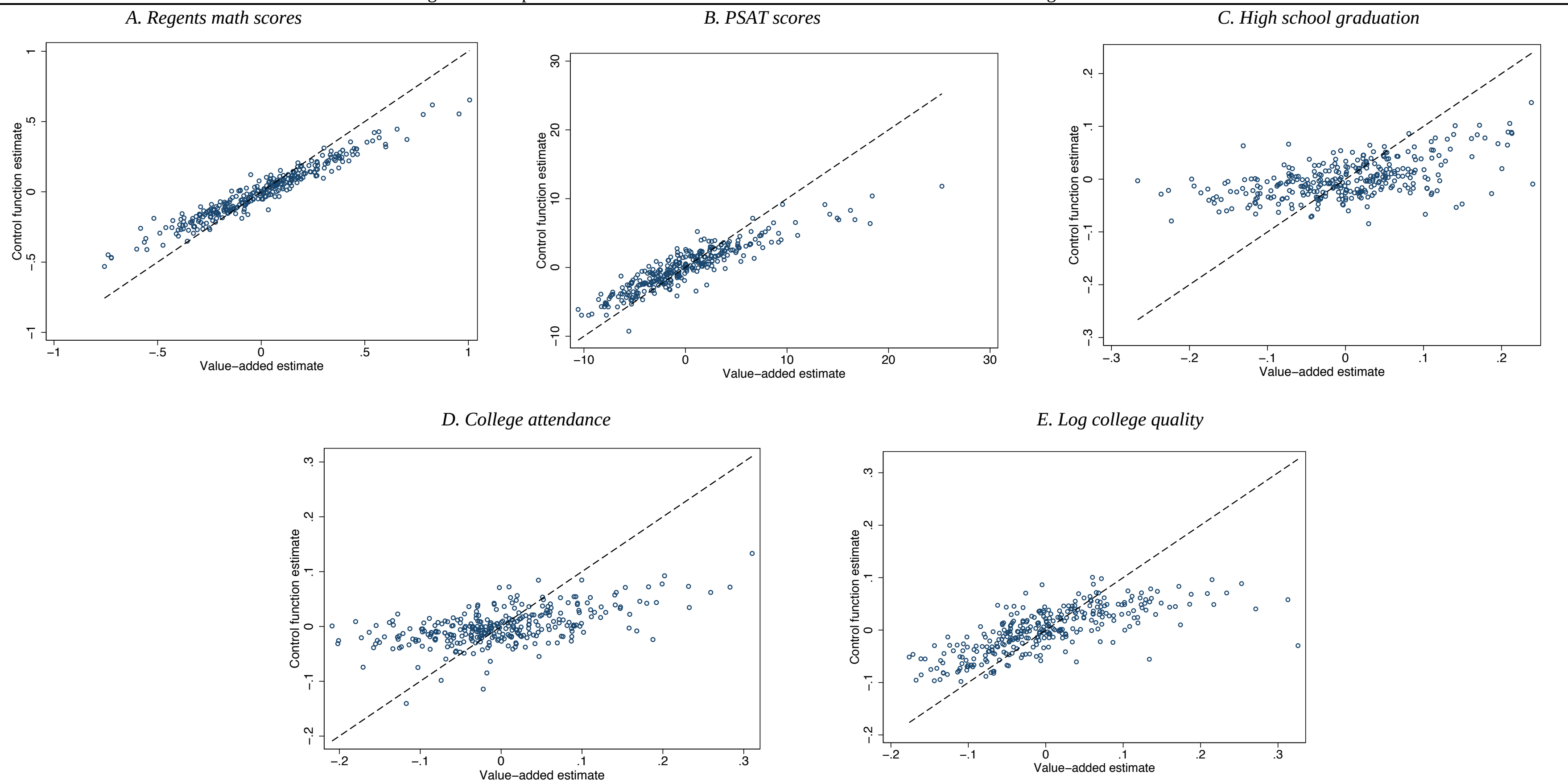
This pattern of findings has important implications for the effects of school choice programs. If parents respond to peer quality but not causal effectiveness, a school's easiest path to boosting its popularity is to improve the average ability of its student population. Since peer quality is a fixed resource, this creates the potential for socially costly zero-sum competition as schools invest in mechanisms to attract

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

the best students. MacLeod and Urquiola (2015) argue 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.

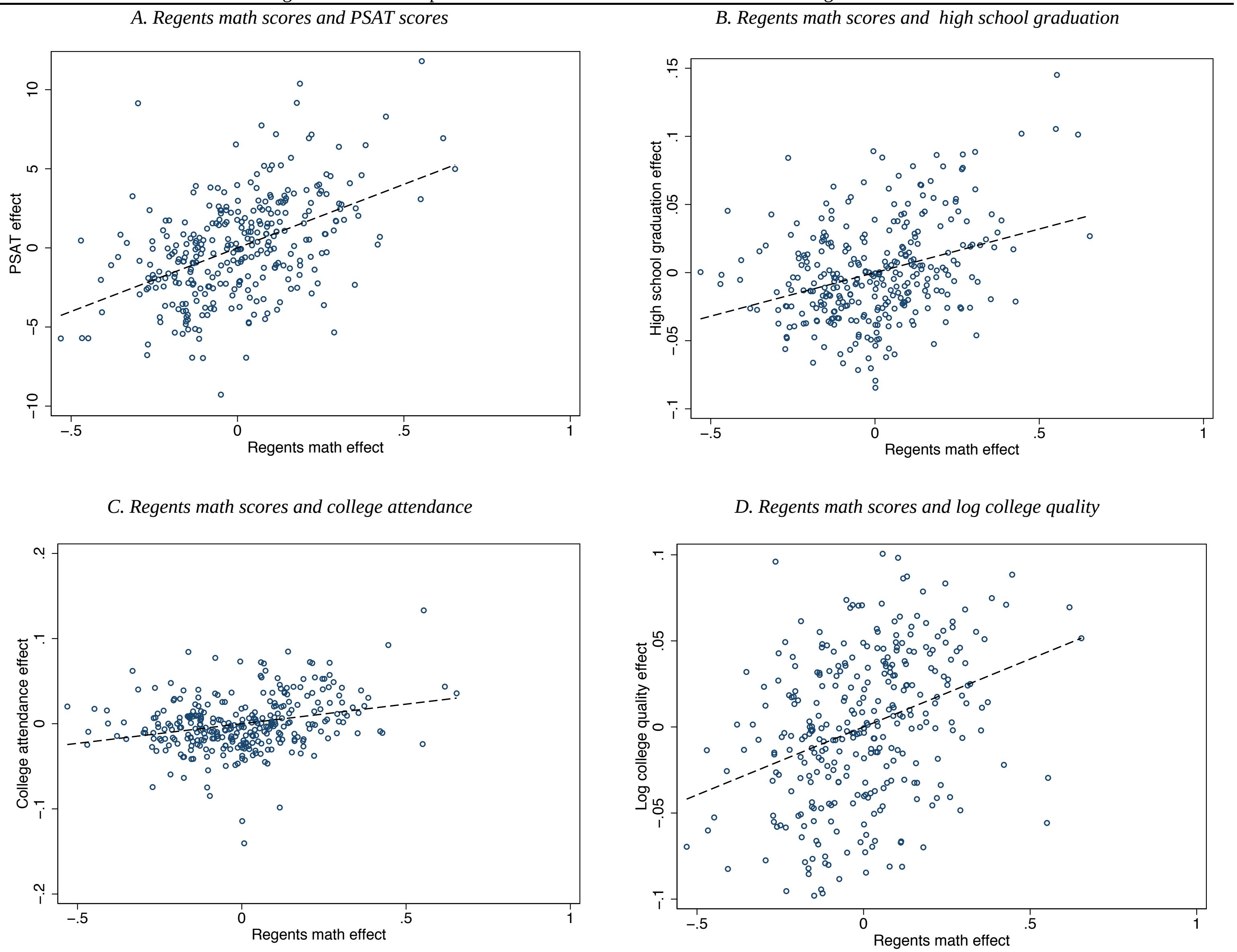
While we have shown that parents do not choose schools based on causal effects for a variety of educational outcomes, we cannot rule out the possibility that preferences are determined by effects on unmeasured outcomes. Parents may be sensitive to school safety or other non-academic amenities, for example. Our analysis also does not address why parents put more weight on peer quality than on treatment effects. Parents may prefer schools with high-achieving peers because of an intrinsic preference for peer achievement or because student composition serves as a signal of causal effectiveness, as in MacLeod and Urquiola (2015). Our results are consistent with either of these possibilities. 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.

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.

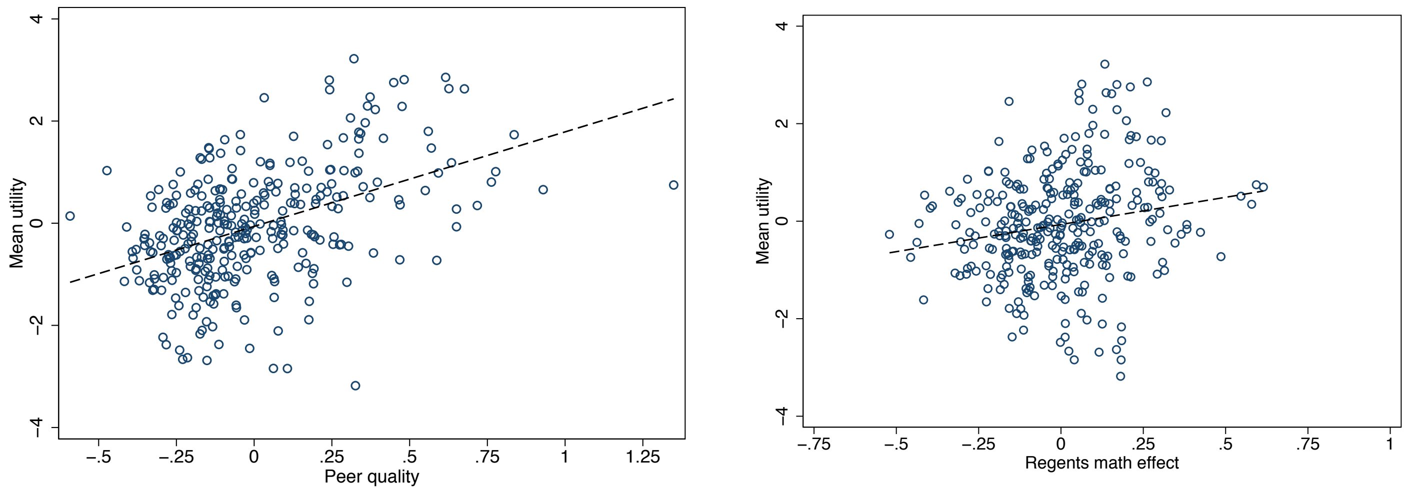
Figure 2: Relationships between effects on test scores and effects on long run outcomes



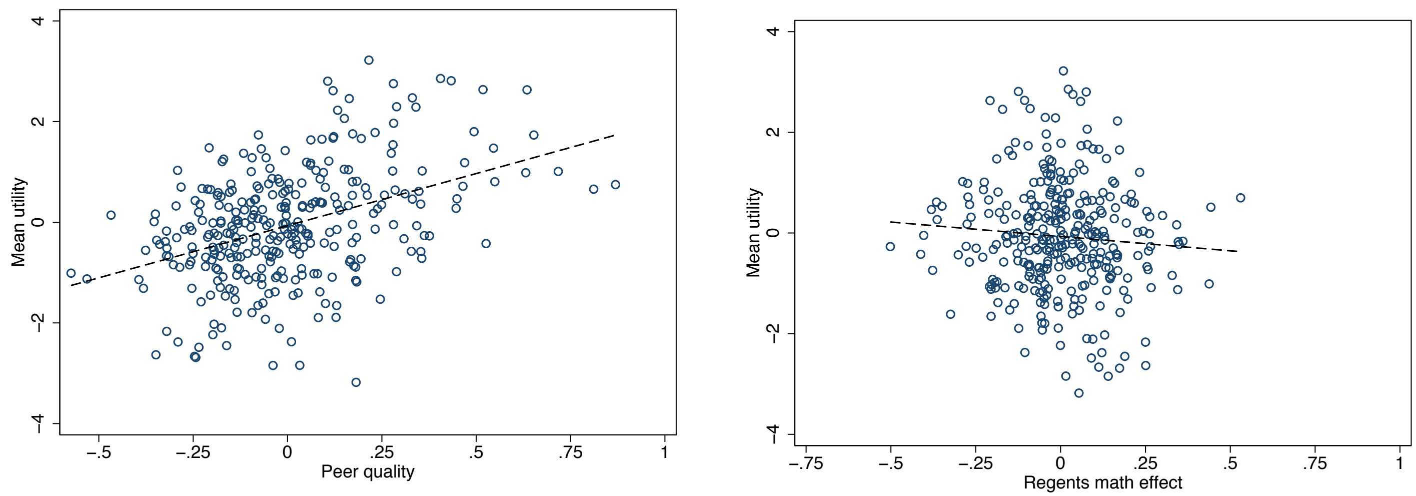
Notes: This figure plots estimates of causal effects on Regents math scores against estimates of effects on longer-run outcomes. Treatment effects are empirical Bayes posterior mean estimates of school average treatment effects from control function models. Panel A plots the relationship between Regents 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.

Figure 3: Relationships among preferences, peer quality, and Regents math effects

A. Bivariate relationships



B. Multivariate relationships



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.

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
N	270157	155850	149365	230087	173254

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 for ninth grade at schools with at least 50 students for each outcome. Regents math scores are normalized to mean zero and standard deviation one in the choice sample. High school graduation equals one 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 one 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 (1)	Same borough (2)	Distance (3)	Regents math score (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.

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		270157		
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.

Table 4. Distributions of peer quality and treatment effect parameters for Regents math scores

	Value-added model		Control function model	
	Mean	Std. dev.	Mean	Std. dev.
	(1)	(2)	(3)	(4)
Peer quality	0	0.285	0	0.305
	-	(0.012)	-	(0.012)
ATE	0	0.289	0	0.233
	-	(0.012)	-	(0.014)
Female	-0.049	0.067	-0.030	0.062
	(0.006)	(0.006)	(0.005)	(0.006)
Black	-0.124	0.175	-0.110	0.121
	(0.013)	(0.013)	(0.011)	(0.012)
Hispanic	-0.107	0.155	-0.091	0.096
	(0.012)	(0.013)	(0.009)	(0.012)
Subsidized lunch	0.003	0.054	0.026	0.055
	(0.006)	(0.006)	(0.005)	(0.006)
Log census tract median income	0.020	0.042	0.013	0.045
	(0.005)	(0.007)	(0.005)	(0.006)
Eighth grade math score	0.626	0.110	0.599	0.106
	(0.008)	(0.006)	(0.007)	(0.006)
Eighth grade reading score	0.163	0.050	0.143	0.053
	(0.005)	(0.004)	(0.004)	(0.004)
Preference coefficient (ψ_j)	-	-	-0.001	0.007
			(0.001)	(0.001)
Match coefficient (φ)	-	-	0.005	-
			(0.002)	

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 zero for out-of-borough schools and indicators for missing values are included.

Table 5. Correlations of peer quality and treatment effect parameters for Regents math scores

	Peer quality	Control function parameters							
		ATE	Female	Black	Hispanic	Sub. lunch	Log tract inc.	Math score	Reading score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ATE	0.587								
	(0.052)								
Female	0.079	0.290							
	(0.078)	(0.101)							
Black	0.013	0.106	-0.202						
	(0.076)	(0.104)	(0.139)						
Hispanic	0.005	0.073	-0.301	0.995					
	(0.075)	(0.106)	(0.157)	(0.007)					
Subsidized lunch	0.046	-0.155	0.059	-0.068	-0.073				
	(0.085)	(0.115)	(0.138)	(0.150)	(0.148)				
Log census tract income	0.040	0.049	-0.061	-0.222	-0.210	-0.219			
	(0.099)	(0.134)	(0.161)	(0.174)	(0.173)	(0.182)			
Eighth grade math score	-0.076	0.039	-0.078	-0.029	-0.020	0.051	0.017		
	(0.064)	(0.083)	(0.098)	(0.101)	(0.100)	(0.111)	(0.130)		
Eighth grade reading score	-0.414	-0.455	-0.190	-0.054	-0.032	-0.006	0.094	0.242	
	(0.068)	(0.093)	(0.116)	(0.127)	(0.128)	(0.132)	(0.153)	(0.098)	
Preference coefficient (ψ_j)	0.427	0.245	0.200	-0.093	-0.112	-0.108	0.311	-0.236	-0.281
	(0.063)	(0.092)	(0.104)	(0.104)	(0.105)	(0.116)	(0.131)	(0.084)	(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.

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	1.600	0.012	0.016	0.021
Variance of peer quality	0.093	0.715	0.010	0.010	0.009
Variance of ATE	0.054	0.240	0.002	0.003	0.004
Variance of match	0.008	0.026	0.002	0.002	0.001
2Cov(peer quality, ATE)	0.081	0.746	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	-0.066	-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.

Table 7. Preferences for peer quality and Regents math effects

	Value-added models				Control function models			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peer quality	0.420		0.442	0.409	0.410		0.467	0.426
	(0.062)		(0.064)	(0.067)	(0.057)		(0.061)	(0.061)
ATE		0.246	-0.033	-0.022		0.216	-0.084	-0.081
		(0.047)	(0.046)	(0.047)		(0.047)	(0.046)	(0.045)
Match effect				-0.072				-0.157
				(0.047)				(0.050)
N	21684							

Notes: This table reports estimates from regressions of school-by-covariate cell mean utility estimates on peer quality and Regents math treatment effect parameter estimates. Mean utilities, peer quality, and treatment effects are scaled in standard deviation units. 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. Columns (1)-(4) report results from value-added models, while columns (5)-(8) report results from control function models. Treatment effect parameters are empirical Bayes posterior means. 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.

Table 8. 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		0.455		0.329		0.267		0.361
		(0.082)		(0.069)		(0.057)		(0.066)
ATE	0.335	-0.057	0.116	-0.065	0.252	0.098	0.194	0.082
	(0.056)	(0.087)	(0.043)	(0.053)	(0.049)	(0.052)	(0.056)	(0.075)
Match effect		0.026		-0.058		-0.044		0.132
		(0.052)		(0.039)		(0.038)		(0.061)
N	21684							

Notes: This table reports estimates from regressions of school-by-covariate cell mean utility estimates on student quality and treatment effect parameter estimates. Mean utilities, peer quality, and treatment effects are scaled in standard deviation units. 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 means from control function models. 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.

Table 9. Heterogeneity in preferences for peer quality and Regents math effects

	By sex		By race			By subsidized lunch		By eighth grade test score tercile		
	Male	Female	Black	Hispanic	Other	Eligible	Ineligible	Lowest	Middle	Highest
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Peer quality	0.432	0.419	0.390	0.361	0.684	0.398	0.490	0.247	0.385	0.680
	(0.062)	(0.067)	(0.063)	(0.066)	(0.131)	(0.059)	(0.079)	(0.058)	(0.064)	(0.090)
ATE	-0.117	-0.046	-0.089	-0.041	-0.208	-0.072	-0.105	-0.054	-0.060	-0.140
	(0.050)	(0.046)	(0.050)	(0.047)	(0.097)	(0.045)	(0.052)	(0.046)	(0.046)	(0.061)
Match effect	-0.149	-0.166	-0.167	-0.133	-0.146	-0.160	-0.149	-0.150	-0.145	-0.110
	(0.049)	(0.053)	(0.056)	(0.062)	(0.056)	(0.051)	(0.050)	(0.061)	(0.056)	(0.049)
N	10795	10889	7467	7433	6784	11043	10641	7264	7286	7134

Notes: This table reports estimates from regressions of school-by-covariate cell mean utility estimates on student quality and Regents math effects separately by student subgroup. 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. Treatment effect estimates are empirical Bayes posterior means from control function models. 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.

References

Abdulkadiroğlu, A., N. Agarwal, and P. A. Pathak (2016): "The welfare effects of coordinated assignment: evidence from the New York City high school match," Forthcoming, *American Economic Review*.

Abdulkadiroğlu, A., J. D. Angrist, and P. A. Pathak (2014): "The elite illusion: achievement effects at Boston and New York exam schools," *Econometrica*, 82, 137–196.

Abdulkadiroğlu, A., P. A. Pathak, and A. E. Roth (2005): "The New York City high school match," *American Economic Review: Papers & Proceedings*, 95, 364–367.

——— (2009): "Strategy-proofness versus efficiency in matching with indifferences: redesigning the NYC high school match," *American Economic Review*, 99, 1954–78.

Abdulkadiroğlu, A., P. A. Pathak, and C. R. Walters (2017): "Free to choose: can school choice reduce student achievement?" Forthcoming, *American Economic Journal: Applied Economics*.

Agarwal, N. and P. Somaini (forthcoming): "Demand analysis using strategic reports: an application to a school choice mechanism," *Econometrica*.

Altonji, J. G., C.-I. Huang, and C. R. Taber (2015): "Estimating the cream skimming effect of school choice," *Journal of Political Economy*, 123, 266–324.

Angrist, J. D., S. R. Cohodes, S. M. Dynarski, P. A. Pathak, and C. R. Walters (2016a): "Stand and deliver: effects of Boston's charter high schools on college preparation, entry and choice," *Journal of Labor Economics*, 34, 275–318.

Angrist, J. D., P. D. Hull, P. A. Pathak, and C. R. Walters (2016b): "Interpreting tests of school VAM validity," *American Economic Review: Papers & Proceedings*, 106, 388–392.

——— (2017): "Leveraging lotteries for school value-added: testing and estimation," *Quarterly Journal of Economics*, 132, 871–919.

Angrist, J. D., P. A. Pathak., and C. R. Walters (2013): "Explaining charter school effectiveness," *American Economic Journal: Applied Economics*, 5, 1–27.

Avery, C. N. and P. A. Pathak (2015): "The distributional consequences of public school choice," NBER Working Paper 21525.

Baicher-Hicks, A., T. J. Kane, and D. O. Staiger (2014): "Validating teacher effect estimates using changes in teacher assignments in Los Angeles," Mimeo, Harvard University.

Barseghyan, L., D. Clark, and S. Coate (2014): "Public school choice: an economic analysis," NBER working paper no. 20701.

Bayer, P., F. Ferreira, and R. McMillan (2007): "A unified framework for measuring preferences for schools and neighborhoods," *Journal of Political Economy*, 115, 588–638.

Black, S. E. (1999): "Do better schools matter? Parental valuation of elementary education," *Quarterly Journal of Economics*, 14, 577–599.

Blinder, A. S. (1973): "Wage discrimination: reduced form and structural estimates," *Journal of Human Resources*, 8, 436–455.

Blundell, R. and R. L. Matzkin (2014): "Control functions in nonseparable simultaneous equations models," *Quantitative Economics*, 5, 271–295.

Burgess, S., E. Greaves, A. Vignoles, and D. Wilson (2014): "What parents want: school preferences and school choice," *The Economic Journal*, 125, 1262–1289.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011): "Robust inference with multiway clustering," *Journal of Business and Economic Statistics*, 29, 238–249.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011): "How does your kindergarten classroom affect your earnings? Evidence from Project STAR," *Quarterly Journal of Economics*, 126, 1593–1660.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014a): "Measuring the impact of teachers I: evaluating bias in teacher value-added estimates," *American Economic Review*, 104, 2593–2563.

——— (2014b): "Measuring the impact of teachers II: teacher value-added and student outcomes in adulthood," *American Economic Review*, 104, 2633–2679.

——— (2016): "Using lagged outcomes to evaluate bias in value-added models," *American Economic Review: Papers & Proceedings*, 106, 393–399.

——— (2017a): "Measuring the impacts of teachers: reply," *American Economic Review*, 107, 1685–1717.

Chetty, R., J. N. Friedman, E. Saez, N. Turner, and D. Yagan (2017b): "Mobility report cards: the role of colleges in intergenerational mobility," The Equality of Opportunity Project, January.

Chetty, R. and N. Hendren (2017): "The impacts of neighborhoods on intergenerational mobility II: county-level estimates," NBER working paper no. 23002.

Chubb, J. E. and T. M. Moe (1990): *Politics, Markets, and America's Schools*, Washington, DC: Brookings Institutution Press.

Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg (2016): "Who benefits from universal childcare? Estimating marginal returns to early childcare attendance," Working paper.

Dahl, G. B. (2002): "Mobility and the return to education: testing a Roy model with multiple markets," *Econometrica*, 70, 2367–2420.

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

——— (2014): "Estimating the effects of college characteristics over the career using administrative earnings data," *Journal of Human Resources*, 49, 323–358.

Deming, D. (2014): "Using school choice lotteries to test measures of school effectiveness," *American Economic Review: Papers & Proceedings*, 104, 406–411.

Deming, D. J., J. S. Hastings, T. J. Kane, and D. O. Staiger (2014): "School choice, school quality, and postsecondary attainment," *American Economic Review*, 104, 991–1013.

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

Dobbie, W. and R. 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, 58–75.

Dubin, J. A. and D. L. McFadden (1984): "An econometric analysis of residential electric appliance holdings and consumption," *Econometrica*, 52, 345–362.

Dubins, L. E. and D. A. Freedman (1981): "Machiavelli and the Gale-Shapley algorithm," 88, 485–494.

Dynarski, S., J. Hyman, and D. W. Schanzenbach (2013): "Experimental evidence on the effect of childhood investments on postsecondary attainment and degree completion," *Journal of Policy Analysis and Management*, 32, 692–717.

Epple, D., D. N. Figlio, and R. Romano (2004): "Competition between private and public schools: testing stratification and pricing predictions," *Journal of Public Economics*, 88, 1215–1245.

Epple, D. and R. Romano (1998): "Competition between private and public schools, vouchers, and peer-group effects," *American Economic Review*, 62, 33–62.

Fack, G., J. Grenet, and Y. He (2015): "Beyond truth-telling: preference estimation with centralized school choice," Working Paper, Paris School of Economics.

Figlio, D. N. and M. E. Lucas (2004): "What's in a grade? School report cards and the housing market," *American Economic Review*, 94, 591–604.

Finkelstein, A., M. Gentzkow, P. Hull, and H. Williams (2017): "Adjusting risk adjustment - accounting for variation in diagnostic intensity," *New England Journal of Medicine*, 376, 608–610.

Friedman, M. (1962): *Capitalism and Freedom*, Cambridge University Press.

Gale, D. and L. S. Shapley (1962): "College admissions and the stability of marriage," *The American Mathematical Monthly*, 69, 9–15.

Glazerman, S. and D. Dotter (2016): "Market signals: evidence on the determinants and consequences of school choice from a citywide lottery," Mathematica Policy Research working paper.

Guarino, C. M., M. D. Reckase, and J. M. Wooldridge (2015): "Can value-added measures of teacher performance be trusted?" *Education Finance and Policy*, 10, 117–156.

Haeringer, G. and F. Klijn (2009): "Constrained school choice," *Journal of Economic Theory*, 144, 1921–1947.

Hanushek, E. A. (1981): "Throwing money at schools," *Journal of Policy Analysis and Management*, 1, 19–41.

Harris, D. N. and M. Larsen (2014): "What schools do families want (and why)?" Technical report, Education Research Alliance for New Orleans.

Hastings, J., A. Hortačsu, and C. Syverson (forthcoming): "Advertising and competition in privatized social security: the case of Mexico," *Econometrica*.

Hastings, J. S., T. J. Kane, and D. O. Staiger (2009): "Heterogeneous preferences and the efficacy of public school choice," Working paper.

Hastings, J. S. and J. M. Weinstein (2008): "Information, school choice, and academic achievement: evidence from two experiments," *Quarterly Journal of Economics*, 123, 1373–1414.

Hausman, J. A. and P. A. Ruud (1987): "Specifying and testing econometric models for rank-ordered data," *Journal of Econometrics*, 34(1-2), 83–104.

Heckman, J. J. and R. Robb (1985): "Alternative methods for evaluating the impact of interventions: an overview," *Journal of Applied Econometrics*, 30, 239–267.

Heckman, J. J., S. Urzua, and E. Vytlacil (2006): "Understanding instrumental variables estimates in models with essential heterogeneity," *Review of Economics and Statistics*, 88, 389–432.

——— (2008): "Instrumental variables in models with multiple outcomes: the general unordered case," *Annales d'Economie et de Statistique*, 91-92, 151–174.

Hoxby, C. M. (2003): "School choice and school productivity: could school choice be a tide that lifts all boats?" in *The Economics of School Choice*, ed. by C. M. Hoxby, Chicago, IL: University of Chicago Press.

Hsieh, C. and M. Urquiola (2006): "The effects of generalized school choice on achievement and stratification: evidence from Chile's voucher program," *Journal of Public Economics*, 90, 1477–1503.

Hull, P. D. (2016): "Estimating hospital quality with quasi-experimental data," Working paper.

Imberman, S. A. and M. 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, 104–121.

Jacob, B. A. and L. Lefgren (2008): "Principals as agents: subjective performance assessment in education," *Journal of Labor Economics*, 26, 101–136.

Kane, T. J., D. F. McCaffrey, and D. O. Staiger (2013): "Have we identified effective teachers? Validating measures of effective teaching using random assignment," *Gates Foundation Report*.

Kane, T. J. and D. O. Staiger (2002): "The promise and pitfalls of using imprecise school accountability measures," *Journal of Economic Perspectives*, 16, 91–114.

Kapor, A., C. A. Neilson, and S. D. Zimmerman (2017): "Heterogeneous beliefs and school choice mechanisms," Working paper.

Kirkebøen, L. J., E. Leuven, and M. Mogstad (2016): "Field of study, earnings, and self-selection," *Quarterly Journal of Economics*, 131, 1057–1111.

Kline, P. (2011): "Oaxaca-Blinder as a reweighting estimator," *American Economic Review: Papers & Proceedings*, 101, 532–537.

Kline, P. and A. Santos (2012): "A score based approach to wild bootstrap inference," *Journal of Econometric Methods*, 1, 23–41.

Kline, P. and C. R. Walters (2016): "Evaluating public programs with close substitutes: the case of Head Start," *Quarterly Journal of Economics*, 131, 1795–1848.

Koedel, C., K. Mihaly, and J. E. Rockoff (2015): "Value-added modeling: a review," *Economics of Education Review*, 47, 180–195.

Ladd, H. F. (2002): *Market-based Reforms in Urban Education*, Washington, DC: Economic Policy Institute.

Langer, A. (2016): "(Dis)Incentives for demographic price discrimination in the new vehicle market," Working paper.

Lee, L.-F. (1983): "Generalized econometric models with selectivity," *Econometrica*, 51, 507–512.

MacLeod, W. B., E. Riehl, J. E. Saavedra, and M. Urquiola (2017): "The big sort: college reputation and labor market outcomes," *American Economic Journal: Applied Economics*, 9, 223–261.

MacLeod, W. B. and M. Urquiola (2015): "Anti-lemons: school reputation and educational quality," *American Economic Review*, 3471–3488.

Manski, C. F. (1993): "Identification of endogenous social effects: the reflection problem," *Review of Economic Studies*, 60, 531–542.

Morris, C. N. (1983): "Parametric empirical Bayes inference: theory and applications," *Journal of the American Statistical Association*, 78, 47–55.

New York City Department of Education (2017): "2017 New York High School Directory," New York, New York.

Oaxaca, R. (1973): "Male-female wage differentials in urban labor markets," *International Economic Review*, 14, 693–709.

Pathak, P. A. and T. Sönmez (2013): "School admissions reform in Chicago and England: comparing mechanisms by their vulnerability to manipulation," *American Economic Review*, 103, 80–106.

Robbins, H. (1956): "An empirical Bayes approach to statistics," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1, 157–163.

Roth, A. E. (1982): "The economics of matching: stability and incentives," *Mathematics of Operations Research*, 7, 617–628.

Rothstein, J. (2006): "Good principals or good peers? Parental valuation of school characteristics, Tiebout equilibrium, and the incentive effects of competition among jurisdictions," *American Economic Review*, 96, 1333–1350.

——— (2010): "Teacher quality in educational production: tracking, decay, and student achievement," *Quarterly Journal of Economics*, 125, 175–214.

——— (2017): "Measuring the impacts of teachers: comment," *American Economic Review*, 107, 1656–1684.

Roy, A. (1951): "Some thoughts on the distribution of earnings," *Oxford Economic Papers*, 3, 135–146.

Todd, P. E. and K. I. Wolpin (2003): "On the specification and estimation of the production function for cognitive achievement," *The Economic Journal*, 113, F3–F33.

Walters, C. R. (forthcoming): "The demand for effective charter schools," *Journal of Political Economy*.

Wooldridge, J. M. (2015): "Control function methods in applied econometrics," *Journal of Human Resources*, 50, 420–445.

Appendix A: Data

The data used for this project were provided by the NYC Department of Education (DOE). This Appendix describes the DOE data files and explains the process used to construct our working extract from these files.

A.1 Application Data

Data on NYC high school applications are controlled by the Student Enrollment Office. We received all applications for the 2003-2004 through 2006-2007 school years. Application records include students' rank-ordered lists of academic programs submitted in each round of the application process, along with school priorities and student attributes such as special education status, race, gender, and address. The raw application files contained all applications, including private school students and first-time ninth graders who wished to change schools as well as new high school applicants. From these records we selected the set of eighth graders who were enrolled as NYC public school students in the previous school year.

A.2 Enrollment Data

We received registration and enrollment files from the Office of School Performance and Accountability (OSPA). These data include every student's grade and building code, or school ID, as of October of each school year. A separate OSPA file contains biographical information, including many of the same demographic variables from the application data. We measure demographics from the application records for variables that appeared in both files and use the OSPA file to gather additional background information such as subsidized lunch status.

OSPA also provided an attendance file with days attended and absent for each student at every school he or she attended in a given year. We use these attendance records to assign students to ninth-grade schools. If a student was enrolled in multiple schools, we use the school with the greatest number of days attended in the year following their final application to high school. A final OSPA file included scores on New York State Education Department eighth grade achievement tests. We use these test scores to assign baseline math and English Language Arts (reading) scores. Baseline scores are normalized to have mean zero and standard deviation one in our applicant sample.

A.3 Outcome Data

Our analysis studies five outcomes: Regents math scores, PSAT scores, high school graduation, college attendance, and college quality. We next describe the construction of each of these outcomes.

The Regents math test is one of five tests NYC students must pass to receive a Regents high school diploma from the state of New York. We received records of scores on all Regents tests taken between

2004 and 2008. We measured Regents math scores based on the lowest level math test offered in each year, which changed over the course of our sample. For the first three cohorts the lowest level math test offered was the Math A (Elementary Algebra and Planar Geometry) test. In 2007, the Board of Regents began administering the Math E (Integrated Algebra I) exam in addition to the Math A exam; the latter was phased out completely by 2009. We assign the earliest high school score on either of these two exams as the Regents math outcome for students in our sample. The majority of students took Math A in tenth grade, while most students taking Math E did so in ninth grade.

PSAT scores were provided to the NYC DOE by the College Board for 2003-2012. We retain PSAT scores that include all three test sections: math, reading, and writing (some subtests are missing for some observations, particularly in earlier years of our sample). If students took the PSAT multiple times, we use the score from the first attempt.

High school graduation is measured from graduation files reporting discharge status for all public school students between 2005 and 2012. These files indicate the last school attended by each student and the reason for discharge, including graduation, equivalent achievement (e.g. receiving a general equivalency diploma), or dropout. Discharge status is reported in years 4, 5, and 6 from expected graduation based on a student's year of ninth grade enrollment; our data window ends in 2012, so we only observe 4-year and 5-year high school discharge outcomes for students enrolled in eighth grade for the 2006-2007 year. We therefore focus on 5-year graduation for all four cohorts. Our graduation outcome equals one if a student received either a local diploma, a Regents diploma, or an Advanced Regents diploma within 5 years of her expected graduation date. Students not present in the graduation files are coded as not graduating.

College outcomes are measured from National Student Clearinghouse (NSC) files. The NSC records enrollment for the vast majority of post-secondary institutions, though a few important New York City-area institutions, including Rutgers and Columbia University, were not included in the NSC during our sample period.[11] The NYC DOE submitted identifying information for all NYC students graduating between 2009 and 2012 for matching to the NSC. Since many students in the 2003-04 eighth grade cohort graduated in 2008, NSC data are missing for a large fraction of this cohort. Our college outcomes are therefore defined only for the last three cohorts in the sample. For these years we code a student as attending college if she enrolled in a post-secondary institution within five years of applying to high school. This captures students who graduated from high school on time and enrolled in college the following fall, as well as students that delayed high school graduation or college enrollment by one year.

We measure college quality based on the mean 2014 incomes of students enrolled in each institution among those born between 1980 and 1982. These average incomes are reported by Chetty et al. (2017b). Fewer than 100 observations in the NSC sample failed to match to institutions in the Chetty et al. (2017b) sample. For students who enrolled in multiple postsecondary institutions, we assign the quality of the first institution attended. If a student enrolled in multiple schools simultaneously, we use the institution with

[11]In addition, about 100 parents opted out of the NSC in 2011 and 2012.

the highest mean earnings.

A.4 Matching Data Files

To construct our final analysis sample, we begin with the set of high school applications submitted by students enrolled in eighth grade between the 2003-2004 and 2006-2007 school years. We match these applications to the student enrollment file using a unique student identifier known as the OSISID and retain individuals that appear as eighth graders in both data sets. If a student submits multiple high school applications as an eighth grader, we select the final application for which data is available. We then use the OSISID to match applicant records to the OSPA attendance and test scores files (used to assign ninth grade enrollment and baseline test scores), and the Regents, PSAT, graduation, and NSC outcome files.

This merged sample is used to construct the set of 316 high schools that enrolled at least 50 students with observations for each of the five outcomes, excluding selective schools that do not participate in the main DA round. The final choice sample includes the set of high school applicants reporting at least one of these 316 schools on their preference lists. The five outcome samples are subsets of the choice sample with observed data on the relevant outcome and enrolled in one of our sample high schools for ninth grade. Table A1 displays the impact of each restriction on sample size for the four cohorts in our analysis sample.

Appendix B: Econometric Methods

B.1 Rank-Ordered Control Functions

This section provides formulas for the rank-ordered control functions in equation (8). The choice model is

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

where $V_{ij} \equiv \delta_{c(X_i)j} - \tau_{c(X_i)} D_{ij}$ represents the observed component of student i's utility for school j and η_{ij} is the unobserved component. The control functions are given by $\lambda_{ij} = E[\eta_{ij} - \mu_\eta | X_i, D_i, R_i] = E[\eta_{ij}|R_i, V_i] - \mu_\eta$, where $V_i = (V_{i1}, ..., V_{iJ})'$. To compute the conditional mean of η_{ij}, it will be useful to define the following functions for any set of mean utilities S and subset $S' \subseteq S$:

$$P(S'|S) = \frac{\sum_{v \in S'} \exp(v)}{\sum_{v \in S} \exp(v)},$$

$$\mathcal{I}(S) = \mu_\eta + \log\left(\sum_{v \in S} \exp(v)\right).$$

$P(S'|S)$ gives the probability that an individual chooses an option in S' from the set S when the value of each option is the sum of its mean utility and an extreme value type I error term, while $\mathcal{I}(S)$ gives the expected maximum utility of choosing an option in S, also known as the inclusive value. We provide expressions for the control functions for two cases: (1) when a student ranks all available alternatives, and (2) when the student leaves some alternatives unranked.

B.1.1 All alternatives ranked

Control function for the highest-ranked alternative

Without loss of generality label alternatives in decreasing order of student i's preferences, so that $R_{ij} = j$ for $j = 1...J$. The control function associated with the highest ranked alternative is

$$\lambda_{i1} = -(V_{i1} + \mu_\eta) + E\left[U_{i1}|R_i, V_i\right]$$

$$= -(V_{i1} + \mu_\eta) + \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{u_1} \int_{-\infty}^{u_2} .... \int_{-\infty}^{u_{J-1}} \left[u_1 \prod_{j=1}^{J} f(u_j|V_{ij})\right] du_J ... du_2 du_1}{\prod_{j=1}^{J-1} P\left(V_{ij}|V_{ij}....V_{iJ}\right)},$$

where $f(u|V) = \exp\left(V - u - \exp(V - u)\right)$ is the density function of a Gumbel random variable with location parameter V. This simplifies to

$$\lambda_{i1} = -(V_{i1} + \mu_\eta) + \frac{\prod_{j=1}^{J} P\left(V_{ij}|V_{ij}....V_{iJ}\right) \times \mathcal{I}(V_{i1}...V_{iJ})}{\prod_{j=1}^{J-1} P\left(V_{ij}|V_{ij}....V_{iJ}\right)}.$$

$$= -V_{i1} + \left(\mathcal{I}(V_{i1}...V_{iJ}) - \mu_\eta\right)$$

$$= -\log P(V_{i1}|V_{i1}...V_{iJ}),$$

which coincides with the control function for the best alternative in the multinomial logit model of Dubin and McFadden (1984). This shows that knowledge of the rankings of less-preferred alternatives does not affect the expected utility associated with the best choice.

Control functions for lower-ranked alternatives

To work out λ_{ij} for $j > 1$, define the following functions:

$$G_{i0}(u) = 1,$$

$$G_{ik}(u) = \int_u^\infty f(x|V_{ik})G_{i(k-1)}(x)dx, \ k = 1...J.$$

It can be shown that

$$G_{ik}(u) = \sum_{j=1}^{k} B_{ik}^j \left[1 - F(u|\mathcal{I}(V_j...V_k) - \mu_\eta)\right],$$

where $F(u|V) = \exp(-\exp(V-u))$ is the Gumbel CDF with location V, and the coefficients B_{ik}^j are:

$$B_{i1}^1 = 1,$$

$$B_{ik}^j = -B_{i(k-1)}^j \times P\left(V_{ik}|V_{ij}....V_{ik}\right), \ k > 1, \ j \neq k,$$

$$B_{ik}^k = \sum_{j=1}^{k-1} B_{i(k-1)}^j, \ k > 1.$$

Then for $j > 1$, we have

$$\lambda_{ij} = -(V_{ij} + \mu_\eta) + \frac{\int_{-\infty}^\infty \int_{u_j}^\infty \int_{u_{j-1}}^\infty ... \int_{u_2}^\infty \int_{-\infty}^{u_j} \int_{-\infty}^{u_{j+1}} ... \int_{-\infty}^{u_{J-1}} \left[u_j \prod_{k=1}^J f(u_k|V_{ik})\right] du_J...du_{j+1}du_1...du_j}{\prod_{k=1}^{J-1} P\left(V_{ik}|V_{ik}....V_{iJ}\right)}$$

$$= -(V_{ij} + \mu_\eta) + \frac{\int_{-\infty}^\infty u_j f(u_j|\mathcal{I}(V_{ij}...V_{iJ}) - \mu_\eta)G_{i(j-1)}(u_j)du_j}{\prod_{k=1}^{j-1} P\left(V_{ik}|V_{ik}....V_{iJ}\right)}$$

$$= -(V_{ij} + \mu_\eta) + \frac{\sum_{m=1}^{j-1} B_{i(j-1)}^m \left[\mathcal{I}(V_{ij}...V_{iJ}) - P(V_{ij}...V_{iJ}|V_{im}...V_{iJ})\mathcal{I}(V_{im}...V_{iJ})\right]}{\prod_{k=1}^{j-1} P\left(V_{ik}|V_{ik}....V_{iJ}\right)}.$$

B.1.2 Unranked alternatives

To derive the control functions for a case in which some alternatives are unranked, assign arbitrary labels $\ell(i) + 1....J$ to unranked schools. The control functions for all ranked alternatives can be obtained by defining a composite unranked alternative with observed utility $V_{iu} = \mathcal{I}\left(V_{ik} : k > \ell(i)\right) - \mu_\eta$ and treating this as the lowest-ranked option in the calculations in section B.1.1. The control function for an unranked alternative $j > \ell(i)$ is defined by the expression

$$\lambda_{ij} + (V_{ij} + \mu_\eta) = E\left[U_{ij}|U_{i1} > ... > U_{i\ell(i)}, \ U_{i\ell(i)} > U_{ik} \ \forall k > \ell(i), V_i\right]$$

$$= \frac{\int_{-\infty}^{\infty} \int_{u_j}^{\infty} \int_{u_{\ell(i)}}^{\infty} \int_{u_{\ell(i)-1}}^{\infty} \cdots \int_{u_2}^{\infty} \int_{-\infty}^{u_{\ell(i)}} .. \int_{-\infty}^{u_{\ell(i)}} u_j \prod_{k=1}^{J} f(u_k|V_{ik}) du_{\ell(i)+1} du_{j-1} du_{j+1} ... du_J du_1 ... du_{\ell(i)} du_j}{\prod_{k=1}^{\ell(i)} P(V_{ik}|V_{ik}...V_{iJ})}$$

$$= \frac{\int_{-\infty}^{\infty} u_j f(u_j|V_{ij}) \left[\int_{u_j}^{\infty} f\left(u_{\ell(i)}|\mathcal{I}(S_i^{-j}(\ell(i))) - \mu_\eta\right) G_{i(\ell(i)-1)}(u_{\ell(i)}) du_{\ell(i)}\right] du_j}{P(V_{i\ell(i)}|S_i^{-j}(\ell(i)))^{-1} \times \prod_{k=1}^{\ell(i)} P(V_{ik}|V_{ik}...V_{iJ})},$$

where $S_i^{-j}(m) = \{V_{ik} : k \geq m\} \backslash \{V_{ij}\}$ is the set of i's mean utilities for alternatives m and higher excluding alternative j. When $\ell(i) = 1$, we have $G_{i(\ell(i)-1)}(u_\ell) = 1$ and this expression collapses to

$$\lambda_{ij} = \frac{P(V_{ij}|V_{i1}...V_{iJ})}{1 - P(V_{ij}|V_{i1}...V_{iJ})} \log P(V_{ij}|V_{i1}...V_{iJ}),$$

which is the expression derived by Dubin and McFadden (1984) for the expected errors of alternatives that are not selected in the multinomial logit model. For $\ell(i) > 1$, we have

$$\lambda_{ij} = -(V_{ij} + \mu_\eta)$$

$$+ \frac{\sum_{m=1}^{\ell(i)-1} B_{i(\ell(i)-1)}^{m} \left[(1 - P(S_i^{-j}(\ell(i))|S_i^{-j}(m)))\mathcal{I}(V_{ij}) - P(V_{ij}|V_{i\ell(i)}..V_{iJ})\mathcal{I}(V_{i\ell(i)}..V_{iJ}) + P(S_i^{-j}(\ell(i))|S_i^{-j}(m))P(V_{ij}|V_{im}..V_{iJ})\mathcal{I}(V_{im}..V_{iJ})\right]}{P(V_{i\ell(i)}|S_i^{-j}(\ell(i)))^{-1} \times \prod_{k=1}^{\ell(i)} P(V_{ik}|V_{ik}...V_{iJ})}.$$

B.2 Two-Step Score Bootstrap

We use a two-step modification of the score bootstrap of Kline and Santos (2012) to conduct inference for the control function models. Let $\Delta = (\delta_{11}...\delta_{1J}, \tau_1...\delta_{C1}...\delta_{CJ}, \tau_C)'$ denote the vector of choice model parameters for all covariate cells. Maximum likelihood estimates of these parameters are given by:

$$\hat{\Delta} = \arg\max_{\Delta} \sum_i \log \mathcal{L}(R_i|X_i, D_i; \Delta),$$

where $\mathcal{L}(R_i|X_i, D_i; \Delta)$ is the likelihood function defined in Section 4.1, now explicitly written as a function of the choice model parameters.

Let $\Gamma = (\alpha_1, \beta_1', \psi_1...\alpha_J, \beta_J', \psi_J, \gamma', \varphi)'$ denote the vector of outcome equation parameters. Second-step estimates of these parameters are

$$\hat{\Gamma} = \left[\sum_i W_i(\hat{\Delta}) W_i(\hat{\Delta})'\right]^{-1} \times \sum_i W_i(\hat{\Delta}) Y_i,$$

where $W_i(\Delta)$ is the vector of regressors in equation (8). This vector depends on Δ through the control functions $\lambda_j(X_i, D_i, R_i; \Delta)$, which in turn depend on the choice model parameters as described in Appendix B.1.

The two-step score bootstrap adjusts inference for the extra uncertainty introduced by the first-step estimates while avoiding the need to recalculate $\hat{\Delta}$ or to analytically derive the influence of $\hat{\Delta}$ on $\hat{\Gamma}$. The first step directly applies the approach in Kline and Santos (2012) to the choice model estimates. This approach generates a bootstrap distribution for $\hat{\Delta}$ by taking repeated Newton-Raphson steps from the full-sample estimates, randomly reweighting each observation's score contribution. The bootstrap estimate of Δ in trial $b \in \{1...B\}$ is:

$$\hat{\Delta}^b = \hat{\Delta} - \left[\sum_i \left(\frac{\partial^2 \log \mathcal{L}\big(R_i|X_i, D_i; \hat{\Delta}\big)}{\partial\Delta\partial\Delta'}\right)\right]^{-1} \times \sum_i \zeta_i^b \left(\frac{\partial \log \mathcal{L}(R_i|X_i, D_i; \hat{\Delta})}{\partial\Delta}\right),$$

where the ζ_i^b are *iid* random weights satisfying $E\left[\zeta_i^b\right] = 0$ and $E\left[(\zeta_i^b)^2\right] = 1$. We draw these weights from a standard normal distribution.

Next, we use an additional set of Newton-Raphson steps to generate a bootstrap distribution for $\hat{\Gamma}$. The second-step bootstrap estimates are:

$$\hat{\Gamma}^b = \hat{\Gamma} - \left[\sum_i W_i(\hat{\Delta})W_i(\hat{\Delta})'\right]^{-1} \times \sum_i \left[-\zeta_i^b W_i(\hat{\Delta})(Y_i - W_i(\hat{\Delta})'\hat{\Gamma}) - W_i(\hat{\Delta}^b)(Y_i - W_i(\hat{\Delta}^b)'\hat{\Gamma})\right].$$

The second term in the last sum accounts for the additional variability in the second-step score due to the first-step estimate $\hat{\Delta}$. We construct standard errors and conduct hypothesis tests involving Γ using the distribution of $\hat{\Gamma}^b$ across bootstrap trials.

B.3 Empirical Bayes Shrinkage

We next describe the empirical Bayes shrinkage procecure summarized in Section 4.2. Value-added or control function estimation produces a set of school-specific parameter estimates, $\left\{\hat{\theta}_j\right\}_{j=1}^{J}$. Under the hierarchical model (10), the likelihood of the estimates for school j conditional on the latent parameters θ_j and the sampling variance matrix Ω_j is:

$$\mathcal{L}\left(\hat{\theta}_j|\theta_j, \Omega_j\right) = (2\pi)^{-T/2} \left|\Omega_j\right|^{-1/2} \exp\left(-\tfrac{1}{2}(\hat{\theta}_j - \theta_j)'\Omega_j^{-1}(\hat{\theta}_j - \theta_j)\right),$$

where $T = \dim(\theta_j)$. We estimate Ω_j using conventional asymptotics for the value-added models and the bootstrap procedure described in Section B.2 for the control function models. Our approach therefore requires school-specific samples to be large enough for these asymptotic approximations to be accurate.

An integrated likelihood function that conditions only on the hyperparameters is:

$$\begin{aligned}\mathcal{L}^I(\hat{\theta}_j|\mu_\theta, \Sigma_\theta, \Omega_j) &= \int \mathcal{L}(\hat{\theta}_j|\theta_j, \Omega_j)dF(\theta_j|\mu_\theta, \Sigma_\theta) \\ &= (2\pi)^{-T/2} \left|\Omega_j + \Sigma_\theta\right|^{-1/2} \exp\left(-\tfrac{1}{2}(\hat{\theta}_j - \mu_\theta)' \left(\Omega_j + \Sigma_\theta\right)^{-1} (\hat{\theta}_j - \mu_\theta)\right).\end{aligned}$$

EB estimates of the hyperparameters are then

$$\left(\hat{\mu}_\theta, \hat{\Sigma}_\theta\right) = \arg\max_{\mu_\theta, \Sigma_\theta} \sum_j \log \mathcal{L}^I\left(\hat{\theta}_j|\mu_\theta, \Sigma_\theta, \hat{\Omega}_j\right),$$

where $\hat{\Omega}_j$ estimates Ω_j.

By standard arguments, the posterior distribution for θ_j given the estimate $\hat{\theta}_j$ is

$$\theta_j|\hat{\theta}_j \sim N\left(\theta_j^*, \Omega_j^*\right) \ ,$$

where

$$\theta_j^* = \left(\Omega_j^{-1} + \Sigma_\theta^{-1}\right)^{-1}\left(\Omega_j^{-1}\hat{\theta}_j + \Sigma_j^{-1}\mu_\theta\right),$$

$$\Omega_j^* = \left(\Omega_j^{-1} + \Sigma_\theta^{-1}\right)^{-1}.$$

We form EB posteriors by plugging $\hat{\Omega}_j$, $\hat{\mu}_\theta$ and $\hat{\Sigma}_\theta$ into these formulas.

Table A1. Sample restrictions

	All cohorts	2003-2004	2004-2005	2005-2006	2006-2007
	(1)	(2)	(3)	(4)	(5)
All NYC eighth graders	368,603	89,671	93,399	94,015	91,518
In public school	327,948	78,904	83,112	84,067	81,865
With baseline demographics	276,797	68,507	67,555	68,279	72,456
With address data	275,405	67,644	67,377	68,108	72,276
In preference sample	270,157	66,125	66,004	67,163	70,865
In Regents math sample	155,850	40,994	41,022	39,177	34,657
In PSAT sample	149,365	31,563	37,502	39,480	40,820
In high school graduation sample	230,087	56,833	56,979	57,803	58,472
In college sample	173,254	0	56,979	57,803	58,472

Notes: This table displays the selection criteria for inclusion in the final analysis samples. Preference models are estimated using the sample in the fourth row, and school effects are estimated using the samples in the remaining rows.

Table A2. Correlations of peer quality and treatment effect parameters for Regents math scores, value-added model

	Peer quality	Value-added parameters						
		ATE	Female	Black	Hispanic	Sub. lunch	Log tract inc.	Math score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ATE	0.544							
	(0.041)							
Female	0.114	0.200						
	(0.078)	(0.085)						
Black	0.009	-0.035	-0.206					
	(0.070)	(0.077)	(0.132)					
Hispanic	0.041	-0.065	-0.301	0.960				
	(0.071)	(0.080)	(0.136)	(0.014)				
Subsidized lunch	0.076	-0.178	0.018	0.017	0.090			
	(0.090)	(0.100)	(0.147)	(0.152)	(0.156)			
Log census tract income	-0.235	-0.076	-0.208	-0.114	-0.024	-0.332		
	(0.107)	(0.127)	(0.179)	(0.194)	(0.201)	(0.214)		
Eighth grade math score	-0.050	0.069	-0.087	-0.058	-0.034	-0.066	-0.139	
	(0.065)	(0.069)	(0.099)	(0.095)	(0.099)	(0.119)	(0.145)	
Eighth grade reading score	-0.579	-0.412	-0.056	-0.037	-0.031	0.139	0.201	0.144
	(0.068)	(0.083)	(0.129)	(0.130)	(0.135)	(0.139)	(0.186)	(0.109)

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 value-added model controlling for observed characteristics.

Table A3. Joint distribution of peer quality and treatment effect parameters for PSAT scores/10

		Peer quality	Control function parameters								
			ATE	Female	Black	Hispanic	Sub. lunch	Log tract inc.	Math score	Reading score	Pref. coef.
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Mean		0	0	-0.092	-0.246	-0.256	-0.091	0.003	1.020	1.102	0.001
		-	-	(0.011)	(0.031)	(0.031)	(0.012)	(0.011)	(0.017)	(0.012)	(0.001)
Standard deviation		0.846	0.490	0.114	0.416	0.418	0.111	0.066	0.250	0.147	0.017
		(0.034)	(0.022)	(0.013)	(0.027)	(0.028)	(0.014)	(0.017)	(0.013)	(0.011)	(0.001)
Correlations:	ATE	0.850									
		(0.018)									
	Female	-0.438	-0.515								
		(0.087)	(0.089)								
	Black	-0.212	-0.309	0.196							
		(0.067)	(0.067)	(0.133)							
	Hispanic	-0.246	-0.327	0.301	0.961						
		(0.066)	(0.066)	(0.130)	(0.011)						
	Subsidized lunch	-0.388	-0.311	0.174	-0.114	-0.024					
		(0.094)	(0.106)	(0.169)	(0.148)	(0.149)					
	Log census tract income	-0.130	-0.104	0.123	-0.114	0.035	-0.471				
		(0.149)	(0.164)	(0.252)	(0.232)	(0.233)	(0.220)				
	Eighth grade math score	0.735	0.681	-0.246	-0.038	-0.088	-0.189	-0.208			
		(0.035)	(0.043)	(0.105)	(0.087)	(0.087)	(0.113)	(0.180)			
	Eighth grade reading score	0.160	0.326	-0.297	-0.173	-0.172	-0.037	0.170	0.238		
		(0.071)	(0.074)	(0.123)	(0.106)	(0.105)	(0.143)	(0.227)	(0.090)		
	Preference coefficient (ψ_j)	0.345	0.200	-0.284	-0.151	-0.116	-0.294	0.250	0.060	-0.122	
		(0.065)	(0.074)	(0.114)	(0.090)	(0.089)	(0.120)	(0.196)	(0.084)	(0.094)	

Notes: This table shows the estimated joint distribution of peer quality and school treatment effect parameters for PSAT scores divded by 10. 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.

Table A4. Joint distribution of peer quality and treatment effect parameters for high school graduation

		Peer quality (1)	ATE (2)	Female (3)	Black (4)	Hispanic (5)	Sub. lunch (6)	Log tract inc. (7)	Math score (8)	Reading score (9)	Pref. coef. (10)
			Control function parameters								
Mean		0	0	0.063	-0.006	-0.013	-0.013	0.003	0.132	0.062	-0.001
		-	-	(0.004)	(0.007)	(0.008)	(0.003)	(0.003)	(0.003)	(0.002)	(0.000)
Standard deviation		0.100	0.043	0.048	0.090	0.103	0.023	0.020	0.034	0.027	0.006
		(0.004)	(0.008)	(0.004)	(0.007)	(0.007)	(0.003)	(0.005)	(0.002)	(0.002)	(0.000)
Correlations:	ATE	0.587									
		(0.136)									
	Female	-0.067	-0.537								
		(0.070)	(0.171)								
	Black	-0.229	-0.308	-0.073							
		(0.069)	(0.194)	(0.140)							
	Hispanic	-0.170	-0.234	-0.083	0.959						
		(0.067)	(0.197)	(0.134)	(0.012)						
	Subsidized lunch	0.153	-0.162	0.125	0.177	0.287					
		(0.094)	(0.236)	(0.169)	(0.180)	(0.175)					
	Log census tract income	0.105	0.168	-0.520	-0.140	-0.256	0.328				
		(0.103)	(0.263)	(0.182)	(0.226)	(0.222)	(0.224)				
	Eighth grade math score	-0.397	-0.624	0.069	-0.164	-0.121	0.073	-0.034			
		(0.060)	(0.165)	(0.097)	(0.108)	(0.106)	(0.123)	(0.125)			
	Eighth grade reading score	-0.571	-0.581	-0.128	0.193	0.093	-0.179	0.137	0.470		
		(0.059)	(0.178)	(0.111)	(0.135)	(0.132)	(0.146)	(0.151)	(0.103)		
	Preference coefficient (ψ_j)	0.624	0.445	0.119	-0.106	-0.049	0.043	-0.157	-0.245	-0.467	
		(0.044)	(0.180)	(0.084)	(0.089)	(0.086)	(0.114)	(0.110)	(0.079)	(0.078)	

Notes: This table shows the estimated joint distribution of peer quality and school treatment effect parameters for high school graduation. 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.

Table A5. Joint distribution of peer quality and treatment effect parameters for college attendance

		Peer quality	Control function parameters								
			ATE	Female	Black	Hispanic	Sub. lunch	Log tract inc.	Math score	Reading score	Pref. coef.
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Mean		0	0	0.076	-0.010	-0.009	-0.008	0.002	0.118	0.064	-0.002
		-	-	(0.003)	(0.009)	(0.009)	(0.003)	(0.003)	(0.003)	(0.002)	(0.000)
Standard deviation		0.099	0.052	0.031	0.121	0.122	0.025	0.021	0.031	0.026	0.005
		(0.051)	(0.029)	(0.005)	(0.008)	(0.010)	(0.006)	(0.005)	(0.009)	(0.006)	(0.002)
Correlations:	ATE	0.855									
		(0.002)									
	Female	0.016	-0.308								
		(0.016)	(0.030)								
	Black	-0.067	-0.488	0.180							
		(0.013)	(0.027)	(0.183)							
	Hispanic	-0.122	-0.460	0.055	0.951						
		(0.015)	(0.028)	(0.183)	(0.050)						
	Subsidized lunch	-0.057	0.141	0.392	-0.620	-0.478					
		(0.023)	(0.030)	(0.305)	(0.334)	(0.285)					
	Log census tract income	-0.208	0.133	-0.241	-0.734	-0.692	0.385				
		(0.030)	(0.037)	(0.229)	(0.190)	(0.188)	(0.263)				
	Eighth grade math score	-0.198	-0.189	0.241	-0.021	-0.016	0.210	-0.442			
		(0.023)	(0.028)	(0.074)	(0.071)	(0.067)	(0.079)	(0.118)			
	Eighth grade reading score	-0.289	-0.112	-0.256	-0.346	-0.401	-0.082	0.278	0.258		
		(0.123)	(0.123)	(0.087)	(0.079)	(0.085)	(0.115)	(0.185)	(0.191)		
	Preference coefficient (ψ_j)	0.768	0.503	0.177	0.107	0.073	-0.005	-0.197	-0.064	-0.309	
		(0.059)	(0.059)	(0.093)	(0.085)	(0.084)	(0.106)	(0.158)	(0.103)	(0.185)	

Notes: This table shows the estimated joint distribution of peer quality and school treatment effect parameters for college attendance. 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.

Table A6. Joint distribution of peer quality and treatment effect parameters for log college quality

		Peer quality	Control function parameters								
			ATE	Female	Black	Hispanic	Sub. lunch	Log tract inc.	Math score	Reading score	Pref. coef.
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Mean		0	0	0.048	-0.038	-0.034	-0.006	0.000	0.103	0.058	-0.002
		-	-	(0.003)	(0.006)	(0.007)	(0.003)	(0.002)	(0.002)	(0.002)	(0.000)
Standard deviation		0.097	0.063	0.027	0.082	0.085	0.022	0.011	0.031	0.019	0.004
		(0.037)	(0.020)	(0.003)	(0.006)	(0.006)	(0.005)	(0.003)	(0.008)	(0.008)	(0.001)
Correlations:	ATE	0.929									
		(0.008)									
	Female	0.132	0.109								
		(0.017)	(0.029)								
	Black	-0.060	-0.244	-0.017							
		(0.015)	(0.029)	(0.154)							
	Hispanic	-0.229	-0.332	-0.126	0.947						
		(0.015)	(0.026)	(0.070)	(0.045)						
	Subsidized lunch	-0.073	0.044	0.259	-0.341	-0.215					
		(0.037)	(0.039)	(0.106)	(0.109)	(0.092)					
	Log census tract income	0.216	0.242	-0.580	-0.640	-0.672	0.061				
		(0.065)	(0.053)	(0.137)	(0.116)	(0.106)	(0.101)				
	Eighth grade math score	0.533	0.735	0.428	-0.132	-0.130	0.122	-0.254			
		(0.038)	(0.041)	(0.098)	(0.129)	(0.095)	(0.068)	(0.169)			
	Eighth grade reading score	0.295	0.470	-0.036	-0.258	-0.267	-0.362	0.107	0.462		
		(0.410)	(0.432)	(0.089)	(0.080)	(0.080)	(0.092)	(0.154)	(0.098)		
	Preference coefficient (ψ_j)	0.750	0.617	0.124	0.032	-0.064	-0.074	0.027	0.305	0.173	
		(0.215)	(0.245)	(0.007)	(0.021)	(0.011)	(0.025)	(0.021)	(0.038)	(1.233)	

Notes: This table shows the estimated joint distribution of peer quality and school treatment effect parameters for college quality. 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.

Table A7. Preferences, peer quality, and math effects, matched first choice model

	(1)	(2)	(3)	(4)
Peer quality	0.367		0.400	0.409
	(0.053)		(0.054)	(0.067)
ATE		0.209	-0.058	-0.036
		(0.045)	(0.043)	(0.045)
Match effect				-0.092
				(0.049)
N	21684			

Notes: This table reports estimates from regressions of school-by-covariate cell mean utility estimates on peer quality and Regents math treatment effect parameter estimates. Mean utilities, peer quality, and treatment effects are scaled in standard deviation units. 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 come from an OLS regression of Regents math scores on school indicators interacted with covariates, with controls for distance and fixed effects for first choice schools. Treatment effect parameters are empirical Bayes posterior means. 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.

Table A8. Preferences, peer quality, and math effects, alternative measures of popularity

	Log first-choice share		Minus log sum of ranks	
	Value-added	Control function	Value-added	Control function
	(1)	(2)	(3)	(4)
Peer quality	0.487	0.505	0.036	0.035
	(0.071)	(0.063)	(0.005)	(0.005)
ATE	-0.009	-0.051	-0.001	-0.003
	(0.045)	(0.042)	(0.003)	(0.003)
Match effect	-0.091	-0.200	-0.004	-0.013
	(0.043)	(0.048)	(0.003)	(0.004)
N	15892		21684	

Notes: This table reports estimates from regressions of measures of school popularity on peer quality and Regents math treatment effect parameter estimates. Peer quality and treatment effects are scaled in standard deviation units. 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 parameters are empirical Bayes posterior means. Columns (1) and (3) report results from value-added models, while columns (2) and (4) report results from control function models. The dependent variable in columns (1) and (2) is the log of the share of students in a covariate cell ranking each school first, and the dependent variable in columns (3) and (4) is minus the log of the sum of ranks for students in the cell. Unranked schools are assigned one rank below the least-preferred ranked school. All regressions include cell indicators. Standard errors are double-clustered by school and covariate cell.