NBER WORKING PAPER SERIES

OVERCOMING RACIAL GAPS IN SCHOOL PREFERENCES: THE EFFECT OF PEER DIVERSITY ON SCHOOL CHOICE

Clemence M. Idoux ⓡ
Viola Corradini ⓡ

Working Paper 33179
http://www.nber.org/papers/w33179

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2024

We are thankful to the New York City Department of Education's Enrollment Research and Policy office for graciously sharing data, and to participants of the MIT behavioral and labor lunch seminars for helpful comments. We are also grateful to participants at the URPP workshop, Michigan State University, and UC Riverside seminars for feedback. Thanks to Eryn Heying, Talia Gerstle, and Jim Shen for dependable administrative support, and to Erika Trevino and Amanda Schmidt for excellent research assistance. This paper reports on research conducted under data-use agreements between MIT, the project's principal investigators, and the New York City Department of Education. This research was made possible by grants from the George and Obie Shultz Fund at MIT and the Wharton Behavioral Lab at the University of Pennsylvania. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Overcoming Racial Gaps in School Preferences: The Effect of Peer Diversity on School Choice
Clemence M. Idoux ⓡ Viola Corradini
NBER Working Paper No. 33179
November 2024
JEL No. C26, I2, I28

ABSTRACT

Differences in school choice by race contribute to school segregation and unequal access to effective schools. Conditional on test score and district of residence, Black and Hispanic families consistently choose schools with fewer white and Asian students, lower average achievement, and lower value-added. This paper combines unique survey data and administrative data from New York City to identify the determinants of racial disparities in school choice and shows that attending a more diverse middle school can mitigate racial choice gaps. Instrumental variable estimates show that middle school students exposed to more diverse peers apply to and enroll in high schools that are also more diverse. These effects particularly benefit Black and Hispanic students who, as a result, enroll in higher value-added high schools. A post-application survey of guardians of high school applicants suggests that most cross-race differences in choice stem from information gaps and homophily in preferences for school demographics. The survey results also reveal that exposure to diverse middle school peers reduces racial differences in choices by addressing these underlying determinants: it increases preferences for peer diversity and broadens the range of known school options.

Clemence M. Idoux ⓡ
Department of Economics
University of California, San Diego
9500 Gilman Drive
LaJolla, CA 92093
and NBER
cidoux@ucsd.edu

Viola Corradini ⓡ
Columbia Business School
viola.corradini@columbia.edu

A data appendix is available at http://www.nber.org/data-appendix/w33179

1 Introduction

In large urban school districts, school choice breaks the connection between residential segregation and schooling and is often offered as a pathway to accessing better schools without the need for physical relocation. Attending schools in wealthier neighborhoods could be an alternative to relocating to low-poverty neighborhoods during childhood, which research has shown to have long-term positive effects (Chetty and Hendren, 2018; Chetty et al., 2016b,a). Consistently, recent research on inter-district busing programs suggests that disadvantaged students transferring to wealthier, predominantly white school districts experience significant improvements in test scores and college enrollment rates (Bergman, 2018).[1] In addition, wealthy school peers might influence long-term outcomes by shaping future behaviors, engagement with information, and decision-making (Conley and Udry, 2010; Cai et al., 2015; Campos, 2023; Golub and Sadler, 2016; Sacerdote, 2011, 2001; Epple and Romano, 2011).

This paper examines the effects of early-grade school peer diversity on one special decision: educational choices. Education decisions have important consequences on lifetime earnings, but they differ by race and socio-economic status. Even when given equal options, disadvantaged families often opt for less prestigious and academically challenging schools compared to their more privileged counterparts (Hoxby and Turner, 2015; Chetty et al., 2020, 2023; Carlana et al., 2022). These differences in school choices contribute to enduring achievement gaps and school segregation (Cohen, 2021; Laverde, 2020; Idoux, 2021).

In this paper, we combine novel survey data with administrative records from New York City (NYC) to examine the determinants of racial and ethnic disparities in school decisions and whether and why exposure to diverse peers in early grades reduces them. NYC provides an ideal context for investigating these questions: the city offers a wide array of school options, but Black and Hispanic families apply to lower-quality and more racially segregated high schools than white and Asian students living in the same neighborhood and with similar test scores.

Our first finding is that middle school diversity plays an important role in shaping high school choices. To address selection bias, we leverage the randomization in the NYC school assignment mechanism to estimate the causal impact of attending schools with higher enrollments of students from different racial backgrounds on subsequent high school choices. Conditional on an applicant's preferences and school priorities, the NYC choice algorithm randomizes seat assignments, thereby manipulating the middle school peer racial make-up

[1] However, attending a farther-away school does not necessarily lead to improved academic outcomes if students do not attend higher-quality schools (Angrist et al., 2022).

independently of potential outcomes. Our estimation strategy exploits this variation, building on the propensity score and instrumental variables methods developed in Abdulkadiroğlu et al. (2017b), Abdulkadiroglu et al. (2022), and Angrist et al. (2022) and extending them to the case where the endogenous variable is a function of the school of enrollment and observed covariates of all the students.

The instrumental variable (IV) estimates show that attending a more diverse middle school significantly affects high school choices and enrollment. Based on these IV estimates, Black and Hispanic students who attend majority white and Asian middle schools select high schools which enroll 6.6 percentage points (p.p.) more white and Asian students and have higher value-added of 0.036 test score standard deviations (σ). These effects correspond to a 24% and 36% reduction in the raw differences in demographic composition and value-added of preferred school choices between Black and Hispanic applicants and white and Asian applicants, which we refer to as "racial choice gaps".[2] They translate into a 73% and 90% reduction of the portion of these racial choice gaps that is unexplained by differences in distance to schools or test scores.

High school choices of white and Asian students, instead, are less affected by middle school peer diversity: enrollment in a majority Black and Hispanic middle school increases the Black and Hispanic peer share of the average high school choice by 2.9 p.p. and has no effect on chosen value-added. Importantly, these shifts in high school choices also affect the schools that applicants are offered in the centralized match. Attending a majority white and Asian middle school leads Black and Hispanic students to enroll in high schools with 6 p.p. more white and Asian students and 0.034 σ higher value-added.

To understand the underlying drivers of racial differences in school choices and how diverse peers narrow these gaps, we conduct a comprehensive post-application survey among 3,000 parents and guardians of high-school applicants during the 2022-2023 application cycle. The survey examines a range of factors that may influence high school choice, including preferences for specific school attributes, awareness of available school options, perceptions of academic quality, beliefs about admission probabilities to competitive schools, aspirations for higher education, and perceptions about discrimination. Unlike previous studies that solely rely on school applications to infer preferences for schools, we use parents' direct answers about their preferences, beliefs and perceptions to better understand what families truly value in schools, even when their school applications reflect misinformation about schools,

[2]Throughout, we use the term "race" to encompass "race and ethnicity". Likewise, the expressions "racial gaps" or "racial differences" refer to distinctions between Black and Hispanic students versus white and Asian students.

misperceptions of admission chances, or strategic reporting. The survey also includes a vignette study to separately estimate preferences for school characteristics such as high-achieving peers, school safety, and racial composition of the student body, that are often correlated with other unobservable school characteristics outside experimental settings.

Our second main finding is that racial disparities in school choice primarily arise from differences in the set of schools families are aware of and from a preference for schools with specific racial compositions, all else being equal. On average, Black and Hispanic families are less likely to know about majority-white and Asian schools or high value-added schools compared to their white and Asian counterparts with similar test scores and residing in the same neighborhoods. In parallel, results from the vignette study indicate a preference for schools where the racial composition aligns with one's own group. This preference persists even among white and Asian respondents who receive clear information about a school's academic performance, indicating that it is not solely a product of statistical discrimination.

We rule out several other potential explanations for differences in school choices across racial groups. Families do not differ in their reported preferences for school attributes; school safety, academic progress, and college enrollment and graduation rates are prioritized across all racial and socio-economic groups. Furthermore, when evaluating parents' information accuracy regarding these school attributes, we find that all groups show similar biases in their assessments. Similarly, we do not observe racial differences in higher education aspirations, perceptions of admission chances at competitive programs, or confidence in one's own academic performance among applicants with similar middle school test scores.

The unique opportunity to link the survey responses to administrative records that encompass students' school enrollment and application history allows us to examine why middle school diversity affects high school choices. Specifically, we investigate how middle school peers may affect the two key determinants of racial choice gaps uncovered by our survey: differences in known school options and homophily. Access to a diverse network of parents could reduce information imbalances, and in fact survey participants emphasize the significant role of interactions with other parents in informing their high school choices. Additionally, prior engagement with diverse peers could increase preferences for interactions with different demographic groups (Rao, 2019; Lowe, 2021; Carrell et al., 2019). Our estimates support these hypotheses. Our final main finding is that Black and Hispanic families whose children attend middle schools with a higher proportion of white and Asian students become aware of a broader range of high schools, especially those with high achievement levels and high value-added. Additionally, attending a more diverse middle school reduces homophily across

all demographic groups. In contrast, IV estimates of middle school peer effects on academic achievement are unlikely to account for the observed changes in application patterns, as they do not significantly increase the perceived likelihood of admission to more selective and sought-after schools.

This paper builds on research that considers how interactions with peers of different backgrounds may impact social attitudes and beliefs (Corno et al., 2019; Boisjoly et al., 2006; Carrell et al., 2019; Rao, 2019; Paluck et al., 2019). We add to this literature by showing that peers from diverse racial and ethnic backgrounds can influence future educational choices. We further unpack the mechanisms driving these positive peer effects, revealing how school integration shapes preferences for interracial contact, as well as how it affects information and beliefs about one's academic standing in a real and high-stakes setting. Our paper also contributes to the literature on information frictions in school choice (Kapor et al., 2020; Arteaga et al., 2021; Ainsworth et al., 2022; Patrick Agte et al., 2024; Campos, 2023) and their unequal impact by socio-economic status (Hastings and Weinstein, 2008; Hoxby and Turner, 2013, 2015; Allende et al., 2019; Pathak and Sönmez, 2008; Corradini, 2024). Leveraging a unique combination of survey data and administrative records, our study directly documents the existence of information disparities and shows that peer networks help to reduce them. The survey further enables us to demonstrate that beliefs about admissions impact school application behavior even when the allocation mechanism is strategy-proof, although sophistication and biases in admission beliefs play a lesser role in explaining racial disparities in school applications. Finally, our findings highlight the importance of path-dependence in school choice. As such, we speak to the literature on the impact of school integration policies (Idoux, 2021; Laverde, 2020; Bjerre-Nielsen and Gandil, 2020), uncovering a potential dynamic effect mostly overlooked so far.[3] Our results suggest that policies reducing school segregation in earlier school grades could lower school segregation and achievement inequality in later grades through changes in demand for schools.

[3]Hahm and Park (2022) considers dynamic effects of integration reforms primarily incorporating expected changes in test scores that affect future admission chances.

2 Background and Data

2.1 The NYC school assignment system

Enrollment in NYC public schools is determined by a centralized school assignment system at the entry grade of each school level. To enroll in pre-kindergarten, kindergarten, sixth grade and ninth grade, students and their families must submit applications through a centralized admission system run by NYC Public Schools (NYCPS). The assignment process is similar for each entry grade. Applicants are asked to rank academic programs by order of preference. School programs rank applicants based on eligibility and priority rules announced before families submit their school preferences.[4] The centralized admission system combines the information and makes a single school offer to each applicant using the student-proposing deferred acceptance (DA) algorithm.

School eligibility and admission criteria are typically based on residential location and academic achievement. Geographic eligibility and admission criteria are more stringent at lower grade levels. At the elementary level, 85% of schools only admit students in their school zone, and the remaining 15% non-zoned schools still give priority to students in their zone. NYC middle schools are intended to serve students residing in their local district, with 83% of middle school programs having zone or district eligibility requirements across the city's 32 local districts.[5] On the contrary, high schools are open to all students in the city, with only approximately 39% of schools giving priority to students residing in their borough or zone. High school and middle school programs instead equally rely on academic admission criteria: approximately a third of the middle and high school programs rank individual students based on prior grades, auditions, essays, and behavioral measures.[6]

To support families in the application process, NYCPS provides access to a personalized website, which only includes schools to which the applicant is eligible. The website is comprised of an information page about each school, which includes a list of offered programs, courses, and extracurricular activities; the performance of enrolled students on standardized tests; admission priorities and selection criteria for each of its programs; the number of applicants per seat and the priority of the last admitted applicant in the prior year. NYCPS also

[4] A school may operate more than one program.

[5] Of the remaining middle school programs, 14% are borough-wide programs, and only the remaining 3% are city-wide programs, with 23% of these programs giving priority to applicants residing in or attending schools in specific districts.

[6] Elementary school programs do not consider academic performance in admission, except for Gifted and Talented programs, which have a separate audition process.

issues annual school reports that list enrolled student demographics, teacher characteristics, and statistics about student performance and school environment. During the application cycle for enrollment in 2023-2024, applicants had also access to their random lottery number on their application profile for the first time.

2.2 Data

We combine two sources of data. The first is administrative data provided by NYC Public Schools (NYCPS) on student school choices, enrollment and test scores, between school years 2013-2014 and 2022-2023. The second source is a survey we conducted, in partnership with NYCPS, among guardians of students applying to enroll in high school in the fall of 2023. A key feature of our data is the possibility of linking survey answers to administrative data covering applicants' entire schooling history within NYCPS. We describe each source in greater detail below.

Administrative data The administrative data covers all students who either enrolled in or applied to a NYC public middle school or high school through the centralized school matches. Our sample focuses on applicants seeking 6^{th} grade seats in public middle schools for enrollment in 2015 to 2020 and who three years later (2018-2023) apply for a 9^{th} grade seat in traditional public high schools within NYC. Applicants who only apply to NYC specialized (exam) and charter schools are omitted from the applicant file.

NYC match data include applicants' rank order lists of schools, for both the middle and the high school application, priorities, and school assigned. Enrollment data indicate the school where the student enrolled in each year after assignment. Application and enrollment data are linked with student demographics, standardized state test scores in math and ELA from assessments in 4^{th} grade and 7^{th} grade, and scores on the Regents Algebra Exam, taken mostly in 9^{th} grade as well as SAT, taken mostly in 11^{th} grade.

From the administrative data, we are able to measure key middle and high school characteristics. First, we measure school quality using value-added models (VAM) capturing the contribution of schools to student achievement. In particular, we adopt a recent methodological improvement in the school VAM literature introduced by Angrist et al. (2021) and referred to as *Risk-Controlled Value-Added (RC VA)*. [7] Our primary measure of high school

[7] Compared to standard methods, this method controls for students' school assignment propensity scores, derived from student applications and the priority status assigned by schools at the time of high school application.

achievement is students' Regents exam scores in Math, which are required for graduation in New York State. We report value-added measures in units of test score standard deviations (σ), centered around the mean value-added of high schools in the city, which is normalized to zero. Second, we measure student-body composition, including the racial and ethnic make-up of students in both middle and high schools, as well as the average baseline achievement of high school students based on their 8^{th} grade Math standardized test scores. Finally, we create measures of high school selectivity and popularity. Selectivity is measured by a binary indicator for high schools that screen students based on test scores, auditions, or other assessments. Popularity is calculated as the ratio of rejected to accepted applicants at each school.[8] While value-added measures are fixed over time, statistics on popularity and demographics vary by year.

Column (1) of Table 1 includes summary statistics for the sample of middle school applicants who were also observed applying to 9^{th} grade seats within NYC public schools. The sample is racially diverse and includes many low socio-economic status (SES) students (72% are eligible for subsidized lunch). The average student attends a middle school where 60% of peers are Black or Hispanic. On average, students rank 8 high school programs on their list. Top choices tend to be of higher quality, more selective, and more popular than the average school in the city. On average, applicants' top 3 program choices have a Regents Math VA which is 0.13 σ higher than the average school in the city. In addition, 67% of applicants rank at least one screened program among their top three choices, which are in higher demand than the average school in the city.[9]

Survey data We supplement the administrative data with responses from a post-application survey of guardians of 9^{th}-grade applicants, which we conducted in partnership with NYC Public Schools. The survey took place from February 17 to March 6, 2023, after applicants had submitted their high school applications but before they received their match offers. It was sent electronically to the email addresses provided during the application process and only parents and guardians were permitted to respond.[10] Respondents could answer in English, Spanish, or Chinese. Upon completion of the survey, participants that had answered at least one survey question were sent a 10-dollar Amazon gift card.

[8] For schools with multiple programs, we construct a school-level measure of popularity by taking a weighted average of program-level rejection-to-acceptance ratios, weighted by each program's seat capacity.

[9] Screened programs reject on average 4.15 applicants for each admitted applicant, compared to a rejection rate of 1.36 for the average school in the city.

[10] 75% of the survey respondents reported that parents and guardians played an essential role in their student's high school selection.

We selected 21,401 potential participants who were general education high school applicants enrolled in NYC public middle schools, for whom we could observe baseline test scores. Of the participants, 17% completed some questions of the survey, and 15% – referred to as respondents in Table 1 – answered over half of the questions. As shown in column (5) of Table 1, survey respondents are more likely to be white or Asian and less likely to be low-income compared to the general NYC high school applicant population (column (1)). Furthermore, respondents' students scored higher on tests than the average NYC student.

The survey examined various dimensions of the choice process for families, including sources of information, essential school characteristics, knowledge of school options and their features, perceptions of admission probabilities and their influence on choice, perceptions of discrimination and its impact on decision-making, and educational aspirations. Additionally, the survey conducted a vignette experiment, described in more details in section 5.1.2, which aimed at disentangling families' relative preferences for different school characteristics and uncovering potential statistical discrimination. The complete survey, as presented to participants, together with detailed information on the construction and randomization of the questions, is available in Online Appendix C.

2.3 Race and School Choice

NYC is among the most segregated school districts in the United States (Frankenberg et al., 2019). Because geographic priorities are more frequently used in admission rules for earlier grades, racial segregation is unsurprisingly more pronounced in elementary and middle schools. Appendix Figure B1 compares overexposure to Black and Hispanic peers for students of different races and grades. Overexposure is defined as the difference between the percentage of Black and Hispanic students enrolled in a school and the percentage of Black and Hispanic students at the same grade level within the city. Across all grade levels, Black and Hispanic students attend schools which enroll disproportionately more Black and Hispanic students than the city's average of 68%. Their over-exposure to Black and Hispanic peers is also more marked in earlier grade levels. Black and Hispanic students attend elementary schools with 18 p.p. more Black and Hispanic students than the elementary school population, middle schools with 15 p.p. more Black and Hispanic students than the middle school population, and high schools with 11 p.p. more Black and Hispanic students than the high school population. In contrast, white and Asian students typically go to high schools with 23 p.p. fewer Black and Hispanic peers than the city average.

While school-side factors may contribute to segregation through admission and eligibility

criteria, demand-side factors are as important. Past research has documented that school choices differ by socio-economic status and ethnicity, with students from poorer families typically applying to schools with lower outcomes in terms of test scores and lower inputs in terms of quality and overall resources (Carlana et al., 2022; Laverde, 2020; Allende et al., 2019; Abdulkadiroğlu et al., 2017a). In NYC too, differences in school choice across student race explain why segregation by race and income remains high (Idoux, 2021) and access to school quality and resources often differs by race or socio-economic group (Corradini, 2024).

In Figure 1, we show that school choices vary by race even for students with similar baseline achievement. Specifically, while everyone chooses schools with a demographic composition similar to their own race, Black and Hispanic students select less selective and lower-quality schools. Each panel compares a different dimension of the high school listed first by applicants of different races based on their 7^{th} grade Math test scores.[11] Across all racial groups, applicants with higher test score prefer schools that enroll more white and Asian students and students with higher baseline achievement. These schools are also more likely to screen applicants on academic achievement and have higher value-added on average. Nonetheless, for any given level of baseline achievement, white and Asian applicants favor these characteristics more than Black and Hispanic applicants.

Panel A of the figure shows that white and Asian students' first-choice schools have on average 20 p.p. more white and Asian students and enroll students with 0.25 σ higher 8^{th} grade test scores than the first choices of Black and Hispanic students with similar middle school test scores. Similarly, Panel B shows that white and Asian students are 10 p.p. more likely than Black and Hispanic students with the same test scores to rank as their first choice a school that screens on academic performance, and they favor schools with 0.06 σ higher Math value-added on average. As a result, the median Black or Hispanic applicant in the baseline test score distribution applies to schools of comparable quality to those chosen by white or Asian applicants whose test scores are 35 percentiles lower.

These racial gaps in preferences for school attributes are approximately constant across 7^{th} grade test scores and are not explained by differences in residential locations. Controlling for district of residence explains half of the gap in the share of white and Asian students (conditional on test scores), but only 25% of the differences in peer baseline math achievement, and none of the difference in probability of applying to screened programs. Importantly, differences in distance to schools account for only 20% of the gap in value-added unexplained

[11] To reflect the information that 8^{th} graders had access to at the time of their high school application, school characteristics are computed using the demographics of the 9^{th} grade cohort enrolled in each school at the time of application.

by middle school test scores, as illustrated in the chart on the left in Panel A of Figure B3. This graph presents the racial gap in value-added for first-choice schools, conditional on baseline achievement and district of residence, suggesting that race affects school choice independently of both test scores and geographic residence.

Consistent with the notion that race influences school preferences, we find that racial gaps in school choices are smaller among students attending more racially diverse middle schools. Figure 2 compares the characteristics of applicants' first-choice high schools based on whether they come from majority white and Asian or majority Black and Hispanic middle schools.[12] The figure reveals two key insights.

First, the racial composition of students' middle school peers significantly shapes the demographics of their high school choices. Black and Hispanic students from majority white and Asian middle schools select high schools with 25 percentage points more white and Asian students compared to other Black and Hispanic students. Conversely, white and Asian students from predominantly Black and Hispanic middle schools tend to choose high schools with 15 percentage points fewer white and Asian peers.

Second, Black and Hispanic applicants who attend majority white and Asian middle schools tend to select high schools with higher value-added and student baseline achievement, and they are also more likely to apply to screened programs if they have high test scores. In contrast, the preferences of white and Asian students for these attributes do not vary as much with the racial composition of their middle school peers. The chart on the right in Panel A of Figure B3 shows how the racial gap in chosen value-added varies across middle schools when controlling for students' district of residence. On average, the choice gap is more than halved for students in majority-white middle schools, while it remains substantial for students from non-majority-white middle schools.[13]

3 Middle School Effects on the Racial Choice Gap

The descriptive evidence in the previous section suggests that middle school diversity plays an important role in shaping school preferences and reducing racial gaps in school choices. Diversity in earlier grades may attenuate preferences for more homogeneous peers in high schools, which might drive part of the differences in application behavior across racial groups.

[12] Figure B2 presents a version of the figures based on 4^{th} grade Math test scores, considering that 7^{th} grade scores could be influenced by middle school demographics. The main patterns observed remain similar.

[13] Panel B of Figure B3 also presents a version of the figures based on 4^{th} grade Math test scores. The main patterns remain similar.

Additionally, middle schools may contribute to leveling the playing field for families across income and race by providing more equitable access to information and creating opportunities for families from varied backgrounds to interact. Indeed, middle schools are a place where parents share information: 75% of respondents reported discussing high school applications with other parents at their student's middle school at least once, 26% engaged in such discussions more than five times, and over a quarter indicated other parents as one of their most important information sources. Middle schools also serve as an institutional source of information about high school applications by organizing information sessions and providing guidance from school staff during the application process.[14]

This discussion motivates us to study the causal effects of middle school demographics on high school choice. We are interested in measuring the effect of exposure to other-race peers in middle school, as measured by the parameter α in the following regression:

$$Y_i = \alpha C_i + X_i'\Gamma + u_i \tag{1}$$

C_i is a measure of contact with other-race peers in the middle school where i enrolls, X_i is a vector of controls, and u_i is a regression residual. Y_i measures an attribute of student i's high school choices, such as the average value-added of the schools ranked within their top three selections. In most of the analysis, we focus on students' top three high school choices, as three-quarters of students are assigned to one of these top choices, and 90 percent of students submit at least three choices.

In most analyses, we group students into two categories based on race or ethnicity: White and Asian students, and Black and Hispanic students. C_i represents a measure of contact with other-race peers in the same middle school grade as student i. For simplicity, when we refer to "other-race peers", we mean Black and Hispanic peers for White or Asian students, and White and Asian peers for Black or Hispanic students. In some specifications, C_i indicates the leave-one-out share of other-race peers in students' middle school, while in others, it indicates having a majority (above 50%) of other-race middle school peers.

Appendix Table A1 presents OLS estimates of α in equation (1), where C_i represents either the presence of a majority of other-race peers (top rows) or the share of other-race peers in middle school (bottom rows). Consistent with Figure 2, the OLS estimates show that Black and Hispanic students who attend middle schools with a greater proportion of white

[14] 16% of families cite middle school sessions as one of their most important sources of information, while 26% of respondents overall and over 30% of low-income, Black, and Hispanic families rely on middle school staff as one of their main sources of information about high schools.

and Asian peers tend to choose high schools that are more popular, with more white and Asian students, and higher-quality. Conversely, white and Asian students with a higher share of Black and Hispanic peers in middle school tend to select high schools with fewer white and Asian students, lower peer achievement, and similar selectivity, though of relatively lower quality. These patterns, however, may be influenced by selection bias if students with particular high school preferences tend to sort into middle schools with different racial compositions. In the following section, we present our instrumental variable approach to identify the causal effects of attending middle schools with a more diverse set of peers.

3.1 Instrumental Variables Framework

The tie-breaking in the middle school match algorithm results in school assignments that are partially determined at random for some students. Our econometric framework takes advantage of this feature to identify the causal effect of exposure to classmates of a different race in middle school.

School offers are a function of applicant preferences and priorities, which we refer to as applicant type θ_i, and the set of tie-breaking variables. Tie-breakers include a common lottery number used by unscreened schools and a set of non-lottery tie-breakers (such as test scores) used by screened schools. This means that school assignment differences for students with the same value of θ_i and proximity to non-lottery cutoffs are due solely to the tie-breaking embedded in the match.

Angrist et al. (2022) shows that the causal effect of any ordered school characteristic, such as peer racial make-up, can be estimated via a 2SLS regression that instruments the enrolled school characteristic with the offered school characteristic and controls for the expected value of the instrument. We adopt a similar method. We instrument the share of other-race peers in the middle school of enrollment with the other-race peer share in the offered school, controlling for the other-race peer share that a students should expect based on their school rankings and priorities. The instrument's expected value controls for systematic differences in potential outcomes between applicants who are offered schools with different racial compositions. Our framework, however, adapts that of Angrist et al. (2022) to account for the fact that the racial makeup of a school depends not only on each student's individual offer but also on the complete set of offers. To circumvent this issue, we compute the potential school racial make-up which uses students' offer distributions instead of realized offers in the construction of the instrument. The remainder of this section describes the empirical strategy in more detail.

For each applicant i, we estimate the probability of assignment to each middle school s in the market. This assignment probability, or propensity score, can be written as:

$$\varphi_s(\theta_i, \tau_i(\delta_N)) = E[D_i(s)|\theta_i, \tau_i(\delta_N)]$$

where $D_i(s)$ indicates an offer at school s. This probability is a function of the applicant type θ_i and indicators for proximity to cutoffs for non-lottery programs, denoted by $\tau_i(\delta_N)$ and determined by a data-driven bandwidth, δ_N. In the large-market theoretical framework outlined in Abdulkadiroglu et al. (2022), the propensity score $\varphi_s(\theta_i, \tau_i(\delta_N))$ depends only on a few match-determined parameters and is easily tabulated from data on the match.

Next, we define the *potential* leave-one-out share of other-race peers in school s as the expected share of other-race peers in school s before resolving any uncertainty over tiebreakers:

$$c_i^P(s) = \frac{\sum_{j \neq i} O_i(j) \cdot \varphi_s(\theta_j, \tau_j(\delta_N))}{\sum_{j \neq i} \varphi_s(\theta_j, \tau_j(\delta_N))}$$

where $O_i(j)$ is a dummy equal to 1 if student j is of a different race than student i. This quantity incorporates the uncertainty in assignment of all students in the match, as the expectation is taken with respect to the probability distributions of school offers of all students.

Potential other-race peer shares will typically differ from realized other-race peer shares, $c_i(s) = \frac{\sum_{j \neq i} O_i(j) \cdot E_j(s)}{\sum_{j \neq i} E_j(s)}$, which are computed using enrollment decisions $E_j(s)$. This discrepancy arises from both the uncertainty in the match and factors such as imperfect offer compliance, student dropouts, and the late enrollment of students who did not participate in the match.

The instrument for the realized share of other-race peers in the school of enrollment, $C_i = \sum_s E_i(s) c_i(s)$, is the potential share of other-race classmates in the middle school offered through the match,

$$Z_i = \sum_s D_i(s) c_i^P(s)$$

.

The expectation of the instrument is derived by taking the expectation over the potential other-race peer share of all schools in student i's middle school application list:

$$\mu_i := E[Z_i|\{\theta_j\}, R_i] = \sum_{s \in S} \varphi_s(\theta_i, \tau_i(\delta_N)) c_i^P(s)$$

As shown in Angrist et al. (2022), conditioning on μ_i ensures the instrument is valid, as

$\epsilon_i \perp Z_i|\mu_i$. Intuitively, μ_i controls for any variation in offered peer race that is due to applicant type θ_i. Any remaining variation is due solely to the tie-breaking randomness in the match.

The research design deployed here is thus a two-stage least squares (2SLS) procedure that uses Z_i to instrument for C_i, controlling for the expected other-race share μ_i. The causal effect of interest is an estimate of coefficient β in the 2SLS system:

$$Y_i = \beta C_i + \kappa_2 \mu_i + \sum_s g_s(R_{is}) + X_i'\Gamma_2 + \epsilon_i \tag{2}$$

$$C_i = \gamma Z_i + \kappa_1 \mu_i + \sum_s h_s(R_{is}) + X_i'\Gamma_1 + \nu_i. \tag{3}$$

Because β might differ by race, we estimate this system of equations separately by race. $g_s(\cdot)$ and $h_s(\cdot)$ are local-linear functions of non-lottery-school tie-breakers R_{is}.[15] These functions employ the bandwidth used to define $\tau_i(\delta_N)$.[16] First and second stage models also include a set of baseline covariates, denoted by X_i.[17]

In addition to the ordered treatment based on the share of other-race peers, the reported estimates also consider a Bernoulli treatment for enrolling in a middle school where the majority of peers are of a different race, denoted by $M_i = \mathbb{I}\{C_i > 0.5\}$. For these estimates, the instrument for M_i is an indicator for being offered a middle school where the offered potential other-race peer share is above 50%. Formally:

$$Z_i^M = \mathbb{I}\{Z_i > 0.5\}.$$

Similarly, the relevant control function for Z_i^M is:

$$\mu_i := E[Z_i^M|\{\theta_j\}, R_i] = \sum_{s \in S} \varphi_s(\theta_i, \tau_i(\delta_N))\mathbb{I}\{c_i^P(s) > 0.5\}.$$

[15] The control functions are as specified in Abdulkadiroglu et al. (2022),

$$g_s(R_{is}) = \omega_{1s}a_{is} + \kappa_{is}\left[\omega_{2s} + \omega_{3s}(R_{is} - T_s) + \omega_{4s}(R_{is} - T_s)\mathbb{I}(R_{is} > T_s)\right].$$

where a_{is} indicates whether applicant i applied to school s, and $\kappa_{is} = a_{is} \times \mathbb{I}(T_s - \delta_s < R_{is} < T_s + \delta_s)$ selects applicants in a bandwidth of size δ_s around an admission cutoff at each school s, T_s.

[16] The bandwidths used here are estimated as suggested by Calonico et al. (2014). Bandwidths are computed separately for each test score variable; we use the smallest of these for each program. We set $\delta_N = 0$ for non-lottery programs with fewer than 5 applicants in the bandwidth who are either below or above the tie-breaker cutoff.

[17] Baseline covariates consist of dummies for female, special needs, free or reduced price lunch, and limited English proficiency, baseline math and ELA scores, and year of application dummies.

The sample consists of middle school applicants with non-degenerate variation or risk for the continuous instrument. That is, the analysis is restricted to applicants who have risk of being assigned to two or more schools with different other-race peer compositions. Appendix Table A2 describes the restrictions applied to construct this experimental sample with greater detail. Columns (2)-(4) in Table 1 compare demographics, the middle school peer racial mix, and high school choices of students in the experimental samples to those of the universe of students observed applying to both middle school and high school in NYC in the study period. While Black and Hispanic students are slightly over-represented in the experimental sample, the sample appears to be quite similar to the population of applicants in column (1).

Appendix Tables A3 and A4 report a set of results meant to validate our research design. Panel A of both tables checks whether differential attrition may lead to selection bias. Both tables show that the likelihood of observing high school choice outcomes is unrelated to the majority other-race offer (Table A3) and offered other-race share instruments (Table A4). Panel B evaluates covariate balance by reporting regressions of covariates on the instruments, controlling for the corresponding μ_i and functions of non-lottery program tiebreakers. For the discrete instrument in Appendix Table A3, the estimates show differences in baseline English test scores for Hispanic applicants. Nonetheless, the magnitude of this difference seems unlikely to lead to substantial omitted variables bias. For the continuous instrument in Appendix Table A4, the estimates show no statistically significant relationships between the offered share of other-race peers and baseline covariates. In any case, all 2SLS estimates are from models that include the baseline covariates listed in the table as controls.

3.2 2SLS Estimates on School Choices

Table 2 reports the 2SLS estimates of attending a majority-white and Asian middle school (top rows in each panel) and attending a middle school with a 10 p.p. higher share of white and Asian peers (bottom rows) on Black and Hispanic high school choices. Panel A considers only students' top three choices and Panel B averages across all the choices in an applicant's list.[18] To account for a change in the number of choices, the table also reports the effect on the length of rank order lists.

Attending a middle school with a higher proportion of white and Asian peers significantly alters the overall application profile of Black and Hispanic students. First, as a result of

[18]The two separate panels disentangle whether exposure to diverse peers affects students' overall preference profile or only students' marginal preferences for the programs they are less likely to attend.

attending a majority-white and Asian middle school, Black and Hispanic applicants prefer high schools that enroll more white and Asian students. Majority-white middle schools induce Black and Hispanic applicants to list schools among their top 3 choices with 4.8 p.p. fewer Black students and 6.6 p.p. more white and Asian students. Similarly, an increase by 10 percentage points in the share of white and Asian middle school peers decreases the share of Black students in top school choices by 1 percentage point and increases the share of white and Asian students by 1.3 percentage points.[19] Attending white and Asian middle schools also induces Black and Hispanic students to rank schools enrolling higher-achieving peers, plausibly because white and Asian students tend to have higher test scores. All these effects are significant at the 5% level.

Second, attending middle schools with more white and Asian peers encourages Black and Hispanic students to choose higher-quality and more selective schools. In particular, attending majority-white middle schools (middle schools with 10 p.p. more white students) increases the average value-added of Black and Hispanic students' top three choices by 0.04 σ (0.01 σ). The magnitude of these effects is only slightly smaller when considering all high school choices. In contrast, the racial composition of middle school peers has little effect on the likelihood of applying to programs that use screened admission methods and on the total number of ranked schools. Because the length of the list is unaffected, these effects are due to substitution rather than the addition of schools to the list.

The changes in Black and Hispanic students' high school choices induced by attending majority-white middle schools are significant, especially when considered against average choice gaps conditional on achievement and district of residence. Attending a majority-white and Asian middle school reduces the raw racial choice gap in school value-added by 36% and the choice gap unexplained by differences in achievement or residence by approximately two-thirds. Similarly, it narrows respectively one and three quarters of the raw and of the unexplained choice gap in school demographic composition.

By contrast, the high school choices of white and Asian students are less influenced by the racial composition of their middle schools, as shown in Table 3. Middle school demographics do not significantly impact the selectivity or quality of high school choices for these students. Additionally, when analyzing the demographic composition of their top high school choices,

[19] Appendix Table A5 presents models that separately analyze the effects on Black and Hispanic students. The bivariate model uses endogenous regressors to separately measure the effects of exposure to peers from the two other ethnic groups. The estimates indicate that exposure to middle school peers from a particular ethnicity encourages students to rank high schools with a greater share of that ethnicity and a lower share of their own, with similar patterns observed for white and Asian peer exposure across Black and Hispanic students.

other-race peer effects are less than half as strong as those observed for Black and Hispanic applicants. For instance, attending majority-Black and Hispanic middle schools only raises the proportion of Black or Hispanic students in white and Asian applicants' top three choices by 2.9 percentage points. The effects are somewhat larger when considering all choices, which contrasts with the patterns observed for Black and Hispanic students. Given that list length, if anything, tends to shorten, these findings suggest that white and Asian applicants primarily adjust their lower-ranked high school choices in response to greater middle school peer diversity. In comparison, interactions with white and Asian peers appear to most strongly influence the top choices of Black and Hispanic students.

A comparison of the 2SLS estimates with the OLS estimates in Table A1 reveals that OLS estimates of peer effects are not extremely biased, especially for Black and Hispanic students. Specifically, the 2SLS estimates of exposure to white and Asian peers are less than 40% lower than the OLS estimates for the racial composition of school choices and are not statistically distinguishable for peer achievement and school value-added.

Finally, Table 4 shows that attending middle schools with more white and Asian peers influences not only the choices but also the school offers for Black and Hispanic students. Black and Hispanic applicants attending majority-white and Asian middle schools are matched to high schools with 6.1 percentage points more white and Asian students and 0.034 σ higher value-added. Although the effects on the characteristics of school offers are somewhat smaller than those on school choices, they suggest that attending more diverse middle schools did not lead these students to rank schools beyond their reach. However, this increase in high school value-added, peer diversity, and peer achievement is accompanied by a 5 percentage point rise in the likelihood of being unmatched, consistent with the lower match rates among white and Asian students. By contrast, middle school demographics have minimal impact on the value-added or racial composition of high school offers for white and Asian students, reflecting the smaller effects observed in their choices.

In summary, the 2SLS analysis indicates that attending majority-white and Asian middle schools leads Black and Hispanic students to choose and match with higher-quality schools. In contrast, middle school demographics do not influence the school quality of choices and offers for white and Asian students. Additionally, exposure to other-race peers in middle school reduces the preference for attending high schools with a higher enrollment of students from one's own race across demographic groups, particularly among Black and Hispanic applicants. Therefore, increasing middle school diversity could serve as a lever for high school desegregation by influencing students' choices. The next section discusses the potential

channels through which middle school peer diversity may impact high school choice.

4 Conceptual Framework

We use a simple school choice framework to shed light on what may drive racial differences in school choices, and why exposure to different peers reduces these gaps. Applicants, indexed by i, choose which high schools to apply to solve a portfolio choice problem as in Chade and Smith (2006). They choose a ranked-ordered list (ROL) of schools $R \in \mathcal{R}_i$, where $\mathcal{R}_i$ comprises the sets resulting from all the k-permutations of A_i, the set of schools the applicant is aware of.[20] Each ROL can be mapped to a lottery over high schools whose weights depend both on the ordering of schools in the list and on applicant beliefs about admission probabilities. Hence, applicants choose their ROL to maximize their expected utility, which can be written as:

$$\max_{R \in \mathcal{R}_i} \sum_{s \in A_i} p_{is}(R, \widehat{q}_i) E[u_{is}(\theta_i, X_{is}) | \mathcal{I}_i] - c_i(R) \tag{4}$$

The utility that student i gets from attending school s, $u_{is}(\theta_i, X_{is})$, depends on the student's preferences for the school attributes X_{is}, including distance from the school, parameterized by the vector of preferences θ_i. Because students may have imperfect knowledge about school attributes, they form expectations about u_{is} according to their (potentially inaccurate) beliefs. These beliefs constitute I_i, the information about school attributes available to student i at the time of application. $p_{is}(R, \widehat{q}_i)$ denotes the subjective probability of *assignment* to school s. Due to the nature of deferred acceptance, applicants' probabilities of *admission* at schools are independent of their rank-order lists, and we assume that applicants are aware of this property. Consequently, $p_{is}(R, \widehat{q}_i)$ depends separately on the choice of R and on the subjective belief of the likelihood of admission at every school, $\widehat{q}_i$. Subjective beliefs about admission probabilities are based on applicants' understanding of the school's admission criteria, and their assessment of their relative ranking in terms of priority and test scores among other applicants to the school.

Finally, applicant i incurs a cost $c_i(R)$, which captures any psychological or monetary cost that a student might face when forming her ROL, given their information and beliefs. For instance, listing highly-selective programs may induce a psychological cost when students

[20] Each R is a strictly ordered set where the ordering of elements in R corresponds to student i's expressed-preference order.

anticipate being disappointed if they are not granted admission. This cost is likely small, but rationalizes why applicants may not include programs for which their admission chances are slim or why applicants may submit very short lists if they are almost certain of being granted admission to one of their top choices.

In this framework, differences in choices across demographic groups arise from the different components of applicants' objective functions:

1. *Differences in preferences* - θ_i. Applicants may put different weight on different school features, even when these attributes are perfectly observed.

2. *Differences in information.* These may take two forms:

   (a) *Differences in awareness sets* - A_i. Applicants are likely only aware of a subset of the more than 400 high schools in the city, and this subset may differ for each individual.

   (b) *Differences in beliefs about school attributes* due to differences in information accuracy or information processing - $E[.|I_i]$. [21]

3. *Differences in perceived probabilities of admissions* - $\widehat{q}_i$. Different students may hold varying beliefs about their admission chances. These differences may accurately reflect differences in individual priorities and test scores, or they may stem from biases in students' subjective assessments of their chances, even when admission probabilities are similar.[22] Applicants' socio-economic and racial background might affect their degree of optimism and confidence in their relative ability and thus influence their perceived chances of admissions.

Peers can shape school choices by influencing each component of applicants' expected utility. Information spreads through social networks, and preferences for schools can also be shaped by social interactions. In particular, positive interactions with school peers from a different race or ethnicity can help reduce prejudice (Rao, 2019; Lowe, 2021; Carrell et al., 2019; Paluck et al., 2019) and change preferences over the demographic composition of schools.

[21] A few existing studies look at school choice responses to information disclosure about school attributes, concluding that families are imperfectly informed about school attributes such as school value-added (Ainsworth et al., 2022; Bergman et al., 2020; Andrabi et al., 2017; Allende et al., 2019; Campos, 2023; Corradini, 2024)

[22] Kapor et al. (2020) and Arteaga et al. (2021) find that beliefs about admission chances differ from rational expectations values using survey evidence in a similar context.

Additionally, peers in earlier school grades may impact applicants' perceived admission probabilities in two ways: peer effects on test scores can alter actual admission chances, and peer interactions can shape perceived probabilities by influencing self-assessments of relative ability. For example, Corno et al. (2019) found that Black students in racially mixed classrooms were less likely to overestimate white peers' GPAs than those in predominantly Black classrooms, a change attributed to improvements in self-image and reduced stereotype threat. The next section examines which channels contribute to racial choice gaps and peer effects on high school selection.

5 Evidence from the Post-Application Survey

Following the framework outlined above, we designed a survey to capture differences in preferences for school characteristics, information about the existence and attributes of schools, and beliefs about admission probabilities. Sections 5.1 - 5.3 analyze survey responses to identify which of these factors are the main drivers of racial disparities in school choices. Section 5.4 concludes with suggestive evidence that exposure to peers of different races helps narrow these gaps by influencing the channels our survey highlights as most significant.

5.1 Racial Gaps in Preferences for School Characteristics and Peers

5.1.1 Stated Preferences

We begin by examining whether applicants of different races prioritize different school attributes. When asked to identify the most important school characteristics, respondents from all racial groups provided similar answers. As shown in Appendix Figure B4, at least 20% of respondents cited the same six attributes as among their top three: safety, academic progress, college and graduation rates, commuting time, the number of AP classes offered, and whether their students would feel they belong. Panel (b) illustrates racial differences in the proportion of respondents rating each attribute as important. The first coefficient indicates the raw cross-race difference, while the second accounts for district of residence and middle school test scores. The differences in race are not statistically significant for the three most frequently mentioned attributes: safety, academic progress, and college and graduation rates.

Aligned with their emphasis on academic achievement, respondents also hold high aspirations for their children's education. Panel (a) of Figure B5 shows that over half agree that attending college is essential for success, and 87% want their children to obtain at least a four-year degree. Educational aspirations among different racial groups are similar; after controlling for district of residence and baseline test scores, Black and Hispanic respondents are just as likely as white and Asian respondents to desire a four-year college education for their children. However, they are 7 percentage points less likely to view college as important for success, which may reflect differences in personal experiences. This evidence indicates that, after accounting for differences in middle school test scores, families across demographic groups share similar school selection criteria and aspirations for their children's education. This finding implies that differences in aspirations or emphasis on academic performance do not explain racial disparities in school choice among students with similar test scores.

5.1.2 Revealed Preferences: Vignette Experiment

The stated rankings of school characteristics reported above do not capture differences in intensity of preferences and may suffer from social desirability bias. In particular, families may be reluctant to express their preferences regarding the demographic makeup of schools. Although few families explicitly reported prioritizing school demographic composition in their choices, the pronounced differences in the demographics of schools selected by students of different races in Figure 1 suggest that this factor may be important to families. It can be challenging, however, to infer preferences for school demographics based on actual school choices, as observed school characteristics may be correlated with unobserved factors. To investigate the causal impact of specific school features on family choices, we conducted a vignette experiment within the survey. The vignette experiment enables us to manipulate school characteristics independently while keeping all other aspects of schools constant.

The experiment consisted of two parts. In the first part, we elicited respondents' cardinal preferences for hypothetical schools by asking them about the likelihood on a scale from 1 to 6 of including two hypothetical schools in their application list. In the second part, we elicited respondents' ordinal preferences for hypothetical schools by asking them to rank two sets of three hypothetical schools. In both parts of the experiment, respondents were told that their student would have high admission chances at any of these hypothetical schools, which were described as identical except for their safety rating, academic performance rating, and racial composition. The experiment employed a factorial design to randomly combine these characteristics, as shown in Appendix Figure B6.

Hypothetical schools had either high-safety or low-safety ratings, and either a balanced racial composition representative of the school district, a majority of Black students, a majority of Hispanic students, or a majority of white or Asian students. In addition, 60% of students were assigned to receive precise information about the schools' academic performances, while 40% received imprecise information. The precise academic information consisted of the 4-year graduation rate and the college and career program enrollment rate.[23] The imprecise information consisted of the share of students that earned enough credits in 9^{th} grade to be on track for graduation, an imprecise signal about graduation rates and college enrollment rates. Participants who received imprecise information were also always presented with a school with median academic performance, while students in the precise information treatment arm were randomly assigned to see high-performing or low-performing schools. The factorial design resulted in 16 unique combinations for the precise-information treatment arm and 8 unique combinations for the imprecise-information treatment arm. The two schools for the first part of the experiment were randomly selected without replacement from these unique combinations. For the second part of the experiment, two sets of three distinct schools were randomly chosen without replacement. Figure 3 shows an example of the vignettes as seen by survey participants while Table A6 outlines the information presented to participants in each treatment arm.[24]

To analyze the vignette experiment, we model respondents' utility to attend any of the hypothetical school as:

$$u_{is} = \sum_{r \in \{b,w\}} (\alpha_r + \beta_r X_s) \times Z_r(i) + e_{is}$$

.

X_s includes the characteristics of school cards: high safety level, majority-black, majority-hispanic, majority-white and Asian, and high academic performance. $Z_r(i)$ indicates whether the respondent is white or Asian ($r = w$) or Black or Hispanic ($r = b$). Thus, α_r captures, separately for each racial group, the average utilities respondents would derive for attending a low-safety, racially-balanced and low-achievement hypothetical school compared to their outside option, while β_r capture the additional utility or disutility from higher safety or academic ratings or a different demographic composition. Finally, $e_{is} \sim N(0, \sigma^2)$ are

[23] Information on school safety, graduation rates, and college enrollment rates are the school performance indicators displayed to families on the application website.

[24] To minimize the salience of the experimental design to respondents, small numbers were added to or subtracted from the values shown to respondents for each metric.

independent and identically distributed utility shocks.

We combine the absolute preferences for schools and relative rankings of schools provided by respondents to estimate their respective weights for different school characteristics.[25] Respondents' reports of their likelihood of applying to the schools provides a normalization for the scale and location of the utility. A respondent indicating a likelihood of a (on a scale from 1 to 6) to list a school implies that $u_{is} \in [a - 0.5, a + 0.5]$. The full parameter vector $\theta = (\beta, \sigma)$ is estimated using a Gibbs sampler to maximize the likelihood of observing the responses to both questions.

Table 5 presents the estimates in likelihood units. Columns (1) and (2) show that school academic and safety ratings are the primary factors that influence families' school choices for all respondents, in line with their stated preferences for school characteristics. Holding all else constant, a high academic rating increases utility by approximately 1.3 and 1.4 points for Black and Hispanic and white and Asian respondents respectively, while a high safety rating increases it by approximately 0.7 points for both respondent groups.

Preferences for a school's racial composition are influenced by the availability of information on academic performance, suggesting that these preferences may partly reflect statistical discrimination. In this context, respondents infer academic performance based on a school's demographics. Comparing the first two columns with the last two in Table 5 shows that respondents with less precise information about school performance are more swayed by racial composition, often disadvantaging majority-Black schools in their rankings. This pattern holds across racial groups: without detailed information, Black and Hispanic respondents are 0.22 points more likely to list racially balanced schools compared to any other school, while white and Asian respondents show a stronger preference for schools with peers of their own race.

However, white and Asian students' preferences for a school's demographic makeup are not fully explained by statistical discrimination. Even when informed about a school's academic performance, they are 0.27 points more likely to favor a majority white or Asian school and 0.28 and 0.44 points less likely to list a majority Hispanic or Black school, respectively, over a racially balanced school.[26] In contrast, there is no indication that Black and Hispanic respondents consider school demographics when they have more precise information about academic performance.

[25] We exclude a small number of respondents whose rankings of cards exhibit inconsistencies across questions.

[26] The table also reveals that white and Asian respondents may have better outside options, as they are 0.55 points less likely to list the reference school compared to Black and Hispanic respondents.

Consistent with stated preferences, the vignette experiment suggests that racial differences in school choices are unlikely to stem from differences in preferences for achievement levels or safety, as all respondents highly value these attributes. Instead, these gaps may be driven by homophily: white and Asian respondents consistently favor schools with more white and Asian students, while Black and Hispanic respondents are less likely to choose majority white or Asian schools when their information on academic performance is limited.[27]

5.2 Racial Gaps in Information about Schools

The conceptual framework highlights two distinct aspects of family information that may explain differences in application behavior: the schools families are aware of and the accuracy of their information on specific school characteristics. We examine each in turn.

5.2.1 Awareness Sets

With over 400 high schools in the city, it is unlikely that families are familiar with all of them, and awareness of their existence may vary across racial groups. To measure school "awareness sets", we asked respondents to identify which schools they recognized from a list of ten. The schools in the list were selected to be relatively close to the respondent's home, popular, and diverse in characteristics. The specific schools shown to each respondent were randomized based on their district of residence.

Panel (a) of Figure 4 shows the share of schools recognized by respondents for different types of schools (left), and racial differences in school recognition (right). The first coefficient for each school type in the right figure reflects raw racial differences, while the second adjusts for district of residence and baseline achievement. On average, respondents were familiar with about one-third of the schools listed, regardless of race. However, Black and Hispanic families were significantly less likely to recognize schools with high white and Asian enrollment, high popularity, high quality or high academic performance, as measured by college enrollment, and graduation rates.[28] Even after controlling for district and baseline achievement, they

[27]This preference for racially similar peers may reflect concerns about potential discrimination by peers, a concern more commonly cited by Asian families: 24% of them believe their student could face racial discrimination from peers, compared to 12% of white respondents. Black and Hispanic families are also slightly more likely to fear peer discrimination than white families, differences are not statistically significant. Among the 23% of respondents reporting concerns about discrimination by peers or teachers, 70% noted that these concerns influenced their high school choices.

[28]High-demand schools were selected among the most popular schools in a respondent's borough, as measured by the share of applicants rejected to applicants accepted. A school is defined as having a high or a low peer quality, value-added, and racial composition if it is respectively in the top or bottom 25%

were 8% less likely to recognize schools with high white and Asian enrollment or high college enrollment rates, and 7% less likely to know schools with high value-added. Conversely, they were 5% more likely to know low value-added schools and 26% more likely to recognize schools with predominantly Black and Hispanic students.[29]

5.2.2 Accuracy of Information

We then explore differences in knowledge about specific school characteristics using two sets of questions. In the first set, we ask respondents to directly compare two schools, such as identifying which has a higher graduation rate, with the outcome of interest being the probability of correctly ranking the schools above the 50% accuracy rate expected by chance. In the second set, respondents evaluate a single school relative to others in the borough by indicating, for instance, the quartile of the graduation rate distribution it falls into; here, the outcome of interest is the correlation between the actual quartile ranking and respondents' beliefs. The schools chosen for these questions were relatively close to the respondent's home, popular, and varied in characteristics, though they were not necessarily in the respondent's awareness set. Additionally, the paired schools differed substantially in the characteristic being compared to ensure the questions were neither overly challenging nor ambiguous.

As before, Panels (b) and (c) of Figure 4 present average outcomes on the left and racial differences on the right. Overall, families are imperfectly informed about the school characteristics they prioritize, tending to be better informed about easily observable factors, like a school's location, and less informed about harder-to-assess measures, such as value-added. Families' knowledge is most accurate regarding commuting distances, followed by college graduation rates and peer quality, measured by the average 7th grade Math achievement of enrolled students. In contrast, families are only 8 percentage points better than random guessing at correctly ranking two schools by value-added, and the correlation between their beliefs about school value-added and the school quartile ranking is not significantly different from zero. They are similarly uninformed about the number of AP classes schools offer.

of schools in that dimension. High peer quality schools enroll students whose average 7th grade Math test scores are at least 0.04σ higher, and are on average 0.35σ higher, than the mean. High (Low) value-added schools increase (decrease) test scores by at least 0.18σ (0.20σ) relative to the mean school VA in the city, and on average by 0.35σ (0.34σ). High white and Asian (Black and Hispanic) schools have a student body that is at least 27% (95%) , and on average is 46% (97%), white or Asian (Black or Hispanic).

[29]Black and Hispanic families may be aware of different school options partly because they use different information sources as shown in Figure B7. Compared to white and Asian families, they rely less on family and friend networks and more on institutional resources, such as middle schools and the NYCPS website, and often turn to more time-intensive sources like individual high school sessions or non-official websites.

Differences in information accuracy across racial groups are unlikely to drive differences in application behavior. In fact, there are generally no significant disparities in the accuracy of respondents' beliefs about school characteristics. Compared to their white and Asian counterparts, Black and Hispanic respondents' beliefs are only marginally less correlated with the actual quartile of schools in terms of peer quality and the offering of AP classes.

One reason for the low information accuracy may be that the selected schools were not necessarily within students' awareness sets.[30] However, even when restricting the sample to schools definitely within a respondent's awareness set, accuracy improves only for easily observed attributes. Panel (a) of Table A7 reports the average probability of correctly answering pairwise comparison questions: for all questions (first row), for questions where one school is certainly known (second row), and for questions where both schools are certainly known (third row). A school is considered known to a respondent if it appeared in the awareness set question and was recognized or if the school was ranked in their high school application. Panel (b) similarly presents the rank-rank correlation coefficient when limiting the sample to schools the applicant is known to recognize. While information accuracy improves for easily observed attributes—such as commuting time, peer achievement levels, college and graduation rates — it does not improve for school value-added. We conclude that differences in the accuracy of information about school value-added do not explain why white and Asian students apply to higher-quality schools; instead, differences in awareness of the existence of these schools appear to be a more relevant factor.

5.3 Racial Gaps in Admission Beliefs

Even if students from the same neighborhood and with similar test scores, in reality, have equal chances of admission, differences in perceptions of admission chances at competitive schools could contribute to racial disparities in applications. When deciding where to apply, in fact, applicants may exclude certain programs from their lists if they believe their chances of admission are too low.[31] The survey provides direct evidence that applicants avoid applying to their preferred schools when they perceive low odds of acceptance, although this behavior cannot explain why Black and Hispanic students are less likely to apply to popular and high-performing schools.

[30] We decided not to only include schools in the respondent's awareness set as we did not want to condition on an outcome but rather we wanted to capture how families form beliefs based on basic cues such as the school name, borough, and district, which are easily accessible from browsing the school directory.

[31] This behavior is observed in deferred acceptance mechanisms when applicants face any application cost. Idoux (2021) provides evidence supporting this claim in the context of NYC.

Panel A of Table 6 shows that beliefs about admission chances strongly influence whether applicants apply to their dream school - the program they would choose if admission were guaranteed - and this effect is similar across racial groups. Students who believe they are certain of acceptance are 26 percentage points more likely to apply than those who feel they have no chance, only 37% of whom applied to their dream school (control mean).[32] Beliefs about admission chances are also positively associated with the likelihood of applying to a high-demand program, though the correlation is lower, as not all respondents necessarily prefer these schools.

Panel B shows that Black and Hispanic respondents have a comparable understanding of their actual admission chances to that of white and Asian respondents. For both groups, a one-point increase in actual admission chances translates to a 17 percentage point rise in beliefs of gaining admission to high-demand or dream schools and a 12 percentage point increase in the likelihood of applying to dream schools. The patterns are similar even after controlling for 7^{th} grade test scores. Thus, differences in pessimism about admission chances for preferred programs do not appear to drive racial gaps in applications.

Panel C suggests that similar beliefs about admission chances across racial groups may partly result from comparable perceptions of students' relative academic standing compared to peers at high-demand schools. On average, respondents are optimistic about their children's academic ranking: white and Asian students in the city's bottom tercile are perceived as being near the median (with a control mean of approximately 2 on a scale from 1 = bottom third to 3 = top third). Applicants in the bottom tercile, relative to students at high-demand schools, are also viewed as slightly above the median in these higher-achieving environments. Among Black and Hispanic respondents, optimism about their children's relative performance is, if anything, slightly higher at the lower end of the citywide distribution and becomes even more positive for high-demand schools as their children's test scores increase. These findings suggest that Black and Hispanic families are not deterred from applying to competitive programs due to a lack of confidence in admission chances or academic ability.

5.4 Peer Effects Mechanisms

Our findings so far suggest that racial gaps in school choices are likely due to differences in awareness of higher-quality options and preferences for schools with demographically similar

[32] Additionally, over one-third of survey respondents stated that they changed their application after observing their random lottery number.

student bodies, rather than differences in preferences for high-quality schools, accuracy of information, or beliefs about admission chances. In this final section, we examine whether exposure to a more diverse middle school peers reduces racial differences in school choice by addressing these underlying factors. Specifically, by linking survey responses to data on students' middle school applications and offers, we explore whether interactions with peers from different racial backgrounds help reduce inter-group prejudice and shape information, test scores, and beliefs.

Preferences Over School Demographics and Perceptions of Discrimination Recent studies find that positive interactions with individuals of different ethnic backgrounds reduce inter-group or inter-racial prejudice (Rao, 2019; Carrell et al., 2019; Lowe, 2021; Corno et al., 2019; Boisjoly et al., 2006; Paluck et al., 2019). Our vignette experiment allows us to isolate the impact of attending middle schools with a higher proportion of students from different racial backgrounds on preferences for the demographic composition of future classmates. To examine how middle school diversity affects racial preferences for high schools, we re-estimate respondents' preferences in the vignette experiment based on whether their student attended a majority-white and Asian middle school. These estimates, presented in Table 7, are suggestive as they rely on raw peer variation rather than instrumental variation.

On average, respondents with students attending majority-white and Asian middle schools prefer hypothetical high schools that are majority white and Asian over racially balanced schools. In contrast, other Black and Hispanic respondents tend to favor racially balanced schools over white and Asian schools, while other white and Asian respondents show no strong preference between racially balanced and majority-white and Asian schools. One interpretation of these findings is that interactions with families of other races in earlier grades may reduce taste-based discrimination. Alternatively, it may reduce statistical discrimination, leading households to rely less on race when inferring school academic performance.

Although we lack the statistical power to fully distinguish between these explanations, our survey provides some evidence that attending a middle school where most peers are of a different race may reduce fear of discrimination, potentially reducing inter-group prejudice and taste-based discrimination. Appendix Table A8 presents OLS and IV estimates of the effects of enrolling in middle schools where the majority of peers are from a different race on measures of perceived discrimination by respondent race.[33] Results show mostly null

[33]Outcomes are binary indicators reflecting agreement with statements about perceived discrimination, such as the likelihood of discrimination by peers or teachers, whether decisions were influenced by fear of discrimination, and whether students would fit well with peers of other races.

effects, except for Asian students, who report the highest perceived discrimination across dimensions. For Asian respondents, attending a majority Black and Hispanic middle school is associated with a lower likelihood of reporting that high school choices were influenced by fear of discrimination and a higher likelihood of agreeing with the statement, "*My student would fit well in a school where the majority of peers are from a different race*".

Information Next, we examine whether attending majority-white middle schools helps close the information gaps identified in section 5.2.1. Table 8 presents OLS and IV estimates of the effect of attending a majority-white and Asian middle school on the likelihood of being familiar with different types of schools. The dependent variables and school types are consistent with those in Panel (a) of Figure 4.

The OLS estimates show that Black and Hispanic applicants attending majority-white middle schools are familiar with a different set of schools compared to other Black and Hispanic students. Specifically, they are more likely to recognize schools that are popular, higher-performing, of better quality, and with larger enrollments of white and Asian students. In contrast, differences in school knowledge across middle schools are much smaller for white and Asian students. These patterns are consistent with findings from administrative data on school applications, suggesting that awareness of school options not only contributes to racial gaps in school choices but may also account for a significant portion of peer effects on choice. The IV estimates further confirm that attending majority-white middle schools causally increases the probability that Black and Hispanic students are aware of high-quality schools and schools with more white and Asian students, by 22 and 28 percentage points, respectively.

Test Scores and Beliefs of Admission Chances Lastly, we examine the effects of attending schools where the majority of peers are from a different race on middle school achievement. Higher test scores could increase students' chances of admission to selective programs and might explain a greater propensity to apply. We measure the impact of middle school diversity on student achievement using the same 2SLS approach as in our analysis of peer effects on high school choices. Table 9 presents the effects of middle school peer race on standardized state Math test scores for both middle and high school.[34]

Attending middle schools with a higher enrollment of white and Asian students has positive effects on the test scores of Black and Hispanic students, showing an increase of

[34]There is minimal concern about test-taking affecting the results. Appendix Table A9 shows that middle school peers have no significant impact on the likelihood of taking standardized tests.

0.14 σ by 8th grade. In contrast, middle school demographics have no impact on the test scores of white and Asian students. The positive effects of middle school peers are even more pronounced in high school Math test scores, as these also reflect the impact of middle school peers on high school choices and the increased value-added of the high schools attended.

However, the increase in 7th grade test scores is not substantial enough to significantly change admission chances or beliefs about admission to popular high school programs. Appendix Table A10 presents IV estimates of peer effects on achievement, admission chances, and beliefs about these outcomes based on the survey sample. Even within this smaller sample, we find that white and Asian school peers boost test scores and citywide academic ranking for Black and Hispanic students. However, these improvements are not sufficient to raise their standing relative to peers in competitive programs or to change actual admission probabilities for high-demand or "dream school" programs. Although test score gains lead to more positive beliefs about one's academic standing relative to other students in the city, they do not affect beliefs about admission chances to high-demand programs — which is consistent with the minimal increase in actual admission chances to those programs.

These results, along with the null peer effect on the likelihood of applying to screened high school programs, suggest that changes in the perceived probability of admission to competitive programs due to higher test scores are unlikely to explain the effects of middle school peers on Black and Hispanic students' application behavior. While it is possible that the increase in test scores influenced Black and Hispanic applicants' preferences for schools enrolling more academically focused students—since students may prefer schools with peers at a similar achievement level to avoid mismatch—this channel is also quantitatively minor. A simple regression shows that a 7th-grade test score increase of 0.086 σ (the effect of attending majority-white schools) raises the baseline Math achievement of students enrolled in Black and Hispanic applicants' top three choices by only 0.014 σ, or 11% of the corresponding middle school peer effect on choices. Similarly, this test score increase would be associated with a rise of 0.44 percentage points in the share of White and Asian students in Black and Hispanic applicants' top three choices and an increase of 0.005 σ in the value-added of their top choices—representing just 7% and 14%, respectively, of the middle school peer effect on school choices.

Summary In summary, we conclude that attending middle schools enrolling more white and Asian peers likely influences Black and Hispanic students' high school choices by (1) changing the type of high schools they know and (2) reducing aversion to attending schools

where the majority of students are white or Asian. Additionally, while majority-white and Asian middle schools have significant positive effects on Black and Hispanic students' middle school test scores, these gains are unlikely to explain the reduction in choice gaps, as they are not large enough to impact admission probabilities or beliefs about admission chances.

6 Conclusions

We document large racial differences in high school choices, even when we compare otherwise similar students living in the same neighborhood and with similar test scores. Black and Hispanic applicants, on average, choose schools of lower quality and with a lower enrollment of white and Asian students. Understanding the roots of these differences, and what works in reducing them, is important because these choice patterns amplify achievement gaps and drive racial segregation in schools.

Combining administrative data and novel survey evidence, we show that these differences are driven by a combination of preferences for the racial composition of schools and information disparities, in the form of limited awareness of school options. While Black and Hispanic students are familiar with fewer majority-white and high-quality schools, attending majority-white and Asian middle schools reduces these gaps and, in turn, influences their high school choices, making them more similar to those of their white peers.

We also find large information frictions in the form of inaccurate beliefs about school attributes and admission chances to high-demand programs, but these are not differential by race. These results highlight that the interventions trying to correct biased beliefs, which have often been the focus of previous studies, might not be the solution to unequal school choices. What seems to be first order is raising awareness about the existence of high-quality schooling opportunities.

Engagement with better-informed peers in earlier school years contributes to this objective, indicating that a potential strategy to promote changes in school choices and bridge information disparities could involve promoting integration in the early grades, which tend to exhibit higher levels of racial segregation. More broadly, these results show the importance of social interactions in shaping the frontier of possibilities that young adults consider when making choices, which may be consequential for settings even beyond high school choice.

Tables

Table 1: Summary Statistics

	Administrative Data Sample				Survey Respondents		
	MS Applicants	Experimental Sample					
	applying to HS	All	Black& Hispanic	White& Asian	All	Black& Hispanic	White& Asian
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Demographics and Baseline Scores							
Black	0.20	0.19	0.31	0.00	0.14	0.28	0.00
Hispanic	0.39	0.41	0.69	0.00	0.33	0.66	0.00
White	0.17	0.17	0.00	0.41	0.21	0.00	0.42
Asian	0.21	0.22	0.00	0.54	0.29	0.00	0.58
FRPL	0.71	0.72	0.82	0.58	0.64	0.75	0.53
Female	0.52	0.51	0.52	0.50	0.52	0.52	0.51
Ell	0.08	0.09	0.10	0.07	0.06	0.08	0.04
4th gr. Math	0.23	0.19	-0.12	0.65	0.51	0.15	0.87
4th gr. Ela	0.23	0.19	-0.05	0.53	0.46	0.18	0.74
Panel B: Middle School of Enrollment							
% Black&Hispanic peer in MS	61	61	76	38	54	71	37
Enrolled in MS with >50% white&Asian	0.38	0.36	0.15	0.68	0.47	0.21	0.73
Panel C: High School Choices							
Number of HS choices	8.2	7.9	8.2	7.5	8.9	8.9	8.9
% Black&Hispanic in top 3 choices	58	58	69	41	54	65	43
Mean baseline peer Math in top 3 choices	0.24	0.21	0.05	0.44	0.34	0.19	0.50
Mean popularity in top 3 choices	4.31	4.18	3.53	5.14	5.91	4.43	7.39
Lists a screened program among top 3 choices	0.71	0.69	0.60	0.81	0.78	0.68	0.88
Mean Math VA in top 3 choices	0.13	0.12	0.08	0.18	0.17	0.11	0.22
Panel D: High School Offers							
Assigned to 1st choice	0.47	0.47	0.52	0.40	0.42	0.50	0.34
Assigned to top 3 choices	0.76	0.77	0.80	0.71	0.72	0.79	0.65
N	256,127	118,291	70,337	47,954	3,628	1,817	1,811

Notes: The administrative data sample in columns 1 to 4 includes general education students who applied to middle school for enrollment in 2015-2016 to 2020-2021 and then successively applied to high school for enrollment in 2018-2019 to 2023-2024. Column 1 reports descriptive statistics for the sample of applicants with non-missing information on their demographics, 4th grade test scores, and residential address. Columns 2 to 4 restrict this sample to the experimental sample, which includes middle school applicants with non-missing baseline information who (i) have non-degenerate risk of school assignment and (ii) received a middle school offer in the middle school match. The survey respondents in column 5 to 7 include any survey participants who answered at least one survey question. The baseline scores are 4th grade scores from the NY state standardized assessments. High school popularity corresponds to the number of applicants rejected by the program divided by the number of accepted applicants (city-mean is 1.37). High school baseline peer Math refers to the average 8th grade Math state standardized test scores of students enrolled. Screened

programs are programs that admit students based on their middle school grades and/or auditions and essays. High school value-added uses state Math test scores (Regents exams) and follows the risk-controlled value-added computation in Angrist et al. (2021).

Table 2: 2SLS Estimates of Peer Effects on Black & Hispanic Applicants' HS Choices

	% Black (1)	% Hispanic (2)	% White&Asian (3)	Peer Math (4)	Popularity (5)	Screened (6)	Math VA (7)	Length of rol (8)
				Top 3 choices				
Majority white & Asian MS	-4.730***	-1.601**	6.543***	0.128***	0.341**	-0.002	0.036***	0.242
	(0.821)	(0.772)	(1.007)	(0.021)	(0.168)	(0.027)	(0.010)	(0.223)
Share white & Asian (10pp)	-0.996***	-0.341**	1.324***	0.021***	0.116***	0.009*	0.008***	0.014
	(0.158)	(0.158)	(0.192)	(0.004)	(0.034)	(0.005)	(0.002)	(0.042)
Mean	23.24	45.89	29.02	0.05	3.53	0.60	0.08	8.22
Race gap	10.11	17.97	-27.78	-0.39	-1.61	-0.21	-0.10	0.68
Race gap | 7^{th} gr scores	7.46	13.60	-20.90	-0.20	-0.38	-0.07	-0.04	1.08
Race gap | 7^{th} gr scores & district	3.46	5.99	-9.24	-0.16	-0.55	-0.07	-0.04	0.33
				All choices				
Majority white & Asian MS	-3.858***	-1.411**	5.516***	0.110***	0.258**	0.012	0.029***	
	(0.695)	(0.607)	(0.799)	(0.016)	(0.122)	(0.021)	(0.007)	
Share white & Asian (10pp)	-0.757***	-0.298**	1.048***	0.017***	0.076**	0.008**	0.007***	
	(0.132)	(0.122)	(0.148)	(0.003)	(0.025)	(0.004)	(0.001)	
Mean	24.04	46.77	27.38	-0.01	3.14	0.79	0.06	
Race gap	10.12	16.15	-26.00	-0.33	-1.32	-0.09	-0.09	
Race gap | 7^{th} gr scores	7.62	12.76	-20.24	-0.18	-0.26	0.00	-0.04	
Race gap | 7^{th} gr scores & district	3.61	4.91	-8.36	-0.13	-0.46	-0.02	-0.04	
N	70,337	70,337	70,337	70,337	69,952	70,337	70,326	70,337

Notes: This table reports 2SLS estimates of middle school demographic composition effects on the characteristics of Black and Hispanic applicants' high school choices. Panel A focuses on the top 3 choices, panel B includes all the choices. The sample includes students with non-degenerate risk of middle school assignment, who applied to Middle schools for enrollment in 2015-2016 to 2020-2021 and then successively applied to high school for enrollment in 2018-2019 to 2023-2024. All models control for application year, student demographic characteristics (ELL status, gender, poverty status, district of residence), and 4th grade Math and ELA test scores. High school popularity, screened status, peer Math, and High School Math VA are defined in the notes of Table 1. Robust standard errors in parenthesis.

Table 3: 2SLS Estimates of Peer Effects on White & Asian Applicants' HS Choices

	% Black (1)	% Hispanic (2)	% White&Asian (3)	Peer Math (4)	Popularity (5)	Screened (6)	Math VA (7)	Length of rol (8)
Top 3 choices								
Majority Black & Hispanic MS	1.281*	1.355*	-2.857**	-0.072**	0.194	0.014	-0.012	-0.733**
	(0.660)	(0.784)	(1.024)	(0.027)	(0.203)	(0.028)	(0.012)	(0.279)
Share Black & Hispanic (10pp)	-0.178	-0.430**	0.683**	0.016**	-0.037	-0.003	0.003	0.134**
	(0.133)	(0.165)	(0.209)	(0.005)	(0.043)	(0.006)	(0.002)	(0.057)
Mean	12.82	27.67	57.40	0.45	5.16	0.81	0.19	7.51
Race gap	10.11	17.97	-27.78	-0.39	-1.61	-0.21	-0.10	0.68
Race gap | 7^{th} gr scores	7.46	13.60	-20.90	-0.20	-0.38	-0.07	-0.04	1.08
Race gap | 7^{th} gr scores & district	3.46	5.99	-9.24	-0.16	-0.55	-0.07	-0.04	0.33
All choices								
Majority Black & Hispanic MS	1.643**	1.549**	-3.411***	-0.064**	0.129	-0.006	-0.008	
	(0.585)	(0.651)	(0.892)	(0.021)	(0.155)	(0.022)	(0.009)	
Share Black & Hispanic (10pp)	-0.345**	-0.382**	0.797***	0.014**	-0.015	0.001	0.002	
	(0.120)	(0.136)	(0.181)	(0.004)	(0.033)	(0.004)	(0.002)	
Mean	13.60	30.41	53.96	0.33	4.48	0.88	0.16	
Race gap	10.12	16.15	-26.00	-0.33	-1.32	-0.09	-0.09	
Race gap | 7^{th} gr scores	7.62	12.76	-20.24	-0.18	-0.26	0.00	-0.04	
Race gap | 7^{th} gr scores & district	3.61	4.91	-8.36	-0.13	-0.46	-0.02	-0.04	
N	45,739	45,739	45,739	45,739	45,685	45,737	45,716	45,739

Notes: This table reports 2SLS estimates of middle school demographic composition effects on the characteristics of white and Asian applicants' high school choices. Panel A focuses on top 3 choices, Panel B includes all the choices. The sample, controls and outcomes are as defined in the notes of Table 2. Robust standard errors in parenthesis.

Table 4: 2SLS Estimates of Peer Effects on Characteristics of Offered High School

	Matched (1)	Offered rank (2)	% Black (3)	% Hispanic (4)	% White&Asian (5)	Peer Math (6)	Popularity (7)	Screened (8)	Math VA (9)
Panel A: Black & Hispanic applicants									
Majority white & Asian MS	-0.054***	0.091	-4.759***	-1.101	6.094***	0.099***	-0.086	-0.026	0.034**
	(0.016)	(0.150)	(1.212)	(1.032)	(1.252)	(0.025)	(0.178)	(0.030)	(0.014)
Share white & Asian (10 pp)	-0.005*	0.003	-0.925***	-0.357*	1.257***	0.013**	0.080**	0.006	0.009**
	(0.003)	(0.029)	(0.228)	(0.204)	(0.221)	(0.005)	(0.033)	(0.005)	(0.003)
Mean	0.95	2.35	27.39	49.32	21.57	-0.14	1.35	0.22	-0.00
N	70,333	66,818	66,975	66,975	66,975	66,944	66,992	66,990	66,837
Panel B: White & Asian applicants									
Majority Black & Hispanic MS	0.008	-0.861***	1.683	0.002	-2.052	-0.035	0.842**	0.070*	-0.017
	(0.021)	(0.210)	(1.134)	(1.166)	(1.521)	(0.034)	(0.257)	(0.038)	(0.018)
Share Black & Hispanic (10pp)	0.001	-0.104**	0.390	0.459*	-0.905**	-0.007	0.168**	0.019**	-0.002
	(0.004)	(0.044)	(0.238)	(0.250)	(0.319)	(0.007)	(0.055)	(0.008)	(0.004)
Mean	0.92	2.91	15.32	32.11	50.58	0.25	2.86	0.44	0.12
N	45,734	42,032	42,281	42,281	42,281	42,279	42,297	42,292	42,198
Race gap	0.03	-0.55	11.67	16.98	-28.33	-0.38	-1.49	-0.21	-0.12
Race gap | 7^{th} gr scores	0.02	-0.31	8.82	13.60	-22.21	-0.21	-0.40	-0.04	-0.05
Race gap | 7^{th} gr scores & district	0.02	-0.30	4.44	5.52	-9.77	-0.14	-0.51	-0.08	-0.04

Notes: This table reports 2SLS estimates of middle school demographic composition effects on the characteristics of high school offers. Panel A focuses on Black and Hispanic applicants, while panel B focuses on white and Asian applicants. The sample, controls and outcomes are as defined in the notes of Table 2. Robust standard errors in parenthesis.

Table 5: Vignette Experiment Preference Estimates

	Precise Information		Imprecise Information	
	Black & Hispanic Respondents (1)	White & Asian Respondents (2)	Black & Hispanic Respondents (3)	White & Asian Respondents (4)
Constant	2.71***	2.17***	3.41***	3.04***
	(0.09)	(0.08)	(0.1)	(0.09)
High-academics	1.28***	1.44***		
	(0.07)	(0.06)		
High-safety	0.66***	0.74***	1.16***	1.08***
	(0.07)	(0.06)	(0.09)	(0.08)
Majority Black	-0.11	-0.44***	-0.28**	-0.7***
	(0.09)	(0.08)	(0.12)	(0.12)
Majority Hispanic	0.00	-0.28***	-0.16	-0.59**
	(0.09)	(0.08)	(0.12)	(0.11)
Majority White	-0.09	0.27***	-0.22*	0.36***
	(0.1)	(0.08)	(0.12)	(0.11)
N respondents	1,212		957	

Notes: This table reports preference estimates for the school vignette experiment, separately for respondents assigned to the treatment arm giving precise information about school academics (Columns 1-2) and for respondents assigned to the treatment arm giving imprecise information about school academics (Columns 3-4). The constant captures the absolute likelihood on a scale from 1 to 6 of listing the school. Preferences are estimated through Gibbs sampling using answers to survey questions Q17 and Q18.

Table 6: Beliefs About Admission Probability

	(1)	(2)	(3)	(4)
Panel A: Beliefs about admission probabilities and application behavior				
	Applied to "high-demand" school		Applied to "dream school"	
Admission belief	0.129**	0.084	0.269***	0.263***
	(0.064)	(0.063)	(0.039)	(0.039)
(Admission belief) × (Black or Hispanic)	-0.012	-0.015	-0.023	-0.023
	(0.101)	(0.099)	(0.060)	(0.060)
Black or Hispanic	-0.117*	-0.051	0.013	0.022
	(0.064)	(0.065)	(0.043)	(0.044)
7th grade test scores controls		X		X
Mean white&Asian, low belief	0.280	0.280	0.373	0.373
N	1,017	1,017	2,460	2,460
Panel B: Admission probabilities and application behavior				
	Admission belief		Applied to "dream school"	
Admission probability	0.168***	0.135***	0.123***	0.107***
	(0.028)	(0.028)	(0.039)	(0.039)
(Admission probability) × (Black or Hispanic)	0.005	0.008	-0.030	-0.029
	(0.028)	(0.028)	(0.043)	(0.043)
Black or Hispanic	0.008	0.035	0.016	0.026
	(0.032)	(0.032)	(0.046)	(0.046)
7th grade test scores controls		X		X
Mean white&Asian, low probability	0.487	0.487	0.849	0.849
N	3,402	3,402	2,431	2,431
Panel C: Beliefs about perfomance tercile				
	Within the City		Within High-demand school	
Actual performance tercile		0.357***		0.147***
		(0.036)		(0.026)
Actual performance tercile* Black&Hispanic		-0.103**		0.086**
		(0.044)		(0.043)
Black&Hispanic		0.231**		-0.130
		(0.115)		(0.084)
Mean white&Asian, 1st tercile		2.08		2.28
N		1,332		1,028

Notes: This table reports OLS estimates of the relationships between applicants' beliefs about admission chances and application behavior (Panel A), applicants' actual admission chances, beliefs and application behavior (Panel B) and applicants' relative performance and their beliefs about it (Panel C). The student relative performance is measured as the tercile in the city-wide distribution of test score (Column 1) or in the distribution of test scores within specific high-demand schools (Column 2). All models control for residential district fixed effects, and for high-demand or dream school fixed effects, except Column 2 of Panel C. Columns 2 and 4 of Panels B and C control for 7th grade test scores. Panel B also controls for applicants' random numbers, as actual admission probabilities estimates account for the uncertainty coming from the

lottery. Robust standard errors are reported in parenthesis, clustered at the student level for column 1 and 2 of panel B. Panels A and B use data from survey questions Q7a, Q7c (application and beliefs about chances in dream school), and Q13 (beliefs about chances in high-demand school). Panel C uses data from survey question Q11 (Column 2) and Q12 (Column 4).

Table 7: Vignette Experiment Preference Estimates by MS Demographics

	Black & Hispanic Respondents		***White & Asian Respondents***	
	(1)		*(2)*	
	Main coefficient	Interaction with (Majority White&Asian MS)	Main coefficient	Interaction with (Majority White&Asian MS)
Constant	2.78***	-0.63***	2.18***	-0.09
	(0.09)	(0.2)	(0.12)	(0.14)
Imprecise Info on Academics	0.81***	0.24	1.00***	-0.08
	(0.09)	(0.19)	(0.12)	(0.14)
High-Academics	1.26***	0.29*	1.52***	-0.11
	(0.08)	(0.17)	(0.11)	(0.13)
High-Safety	0.86***	0.05	0.97***	-0.16
	(0.06)	(0.13)	(0.09)	(0.1)
Majority Black	-0.22***	0.12	-0.50***	0.02
	(0.08)	(0.18)	(0.12)	(0.14)
Majority Hispanic	-0.11	0.26	-0.38***	0.02
	(0.09)	(0.19)	(0.12)	(0.14)
Majority White&Asian	-0.23***	0.46**	0.09	0.28**
	(0.09)	(0.19)	(0.12)	(0.14)
N Respondents	914		1,086	

Notes: This table reports how preference estimates in the school vignette experiment vary across applicants enrolled in majority white and Asian middle schools and applicants enrolled in other middle schools. Column 1 reports estimates for Black and Hispanic respondents, column 2 for white and Asian respondents. The constant captures the absolute likelihood on a scale from 1 to 6 of listing the school. Preferences are estimated through Gibbs sampling using answers to survey questions Q17 and Q18.

Table 8: OLS and 2SLS Estimates of Peer Effects on Consideration Sets

	Any school		High Demand		High Peer Quality		High VA		High white&Asian %	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(High White&Asian MS) × (Black&Hispanic)	0.05***	0.16	0.06**	0.02	0.08***	0.20	0.06**	0.23*	0.09***	0.29**
	(0.02)	(0.11)	(0.03)	(0.18)	(0.02)	(0.14)	(0.02)	(0.14)	(0.02)	(0.13)
High White&Asian MS	-0.02	-0.08	0.02	0.04	-0.02	-0.09	0.00	-0.15	-0.03	-0.14
	(0.01)	(0.08)	(0.02)	(0.11)	(0.02)	(0.09)	(0.02)	(0.10)	(0.02)	(0.09)
Black&Hispanic	-0.00	-0.00	-0.03	-0.03	-0.05***	-0.05**	-0.02	-0.02	-0.05***	-0.05***
	(0.01)	(0.01)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Mean white&Asian	0.380	0.380	0.680	0.680	0.520	0.520	0.440	0.440	0.510	0.510
N Respondents	2,569	2,569	2,569	2,569	2,569	2,569	2,565	2,565	2,568	2,568
N	25,690	25,690	5,138	5,138	12,344	12,344	10,161	10,161	12,926	12,926

Notes: This table reports OLS and 2SLS estimates of middle school demographic composition effects on survey respondents' awareness sets, and how these vary for applicants of different races. It uses data from survey question Q9. The dependent variables are binary outcomes indicating that respondents recognized a

school by name. Depending on the regression, the sample is restricted to schools in high-demand (columns 3-4), schools with high peer quality (columns 5-6), with high value-added (columns 7-8), or enrolling high shares of white and Asian students (columns 9-10). High-demand schools were selected among the most popular schools in a respondent's borough, as measured by the share of applicants rejected to applicants accepted. A school is defined as having high peer quality, value-added, and white and Asian student share if it is in the top 25% of schools in that dimension. All regressions control for residential district fixed effects, 4th grade test scores and commuting time to the school. Clustered standard errors at the student-level in parenthesis.

Table 9: 2SLS Estimates of Peer Effects on Test Scores

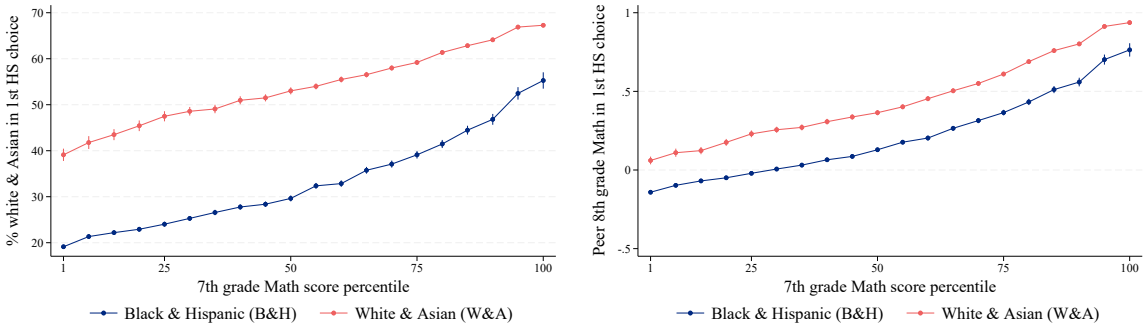
	Middle School Outcomes				***High School Outcomes***	
	6th grade Math	7th grade Math	8th grade Math	Regents Math	Regents Math	SAT Math
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Black & Hispanic students			
Majority white&Asian MS	0.031	0.086**	0.141**	0.146*	0.191**	0.018
	(0.040)	(0.041)	(0.059)	(0.077)	(0.092)	(0.063)
Share white&Asian (10pp)	-0.006	0.012	0.023*	0.025*	0.022	-0.003
	(0.008)	(0.008)	(0.012)	(0.014)	(0.018)	(0.013)
Mean	-0.11	-0.09	0.00	0.29	-0.17	-0.29
N	50,529	45,220	29,625	16,039	18,549	24,196
			Panel B: White & Asian Students			
Majority Black%&Hispanic MS	-0.021	0.022	0.109	-0.003	-0.019	-0.094
	(0.049)	(0.048)	(0.105)	(0.062)	(0.149)	(0.091)
Share Black&Hispanic (10pp)	-0.005	-0.003	0.036*	-0.002	-0.024	0.004
	(0.011)	(0.011)	(0.022)	(0.012)	(0.035)	(0.020)
Mean	0.73	0.75	0.78	0.95	0.47	0.74
N	32,302	30,212	14,909	17,214	7,739	18,241

Notes: This table reports 2SLS estimates of middle school demographic composition effects on 6th, 7th, and 8th grade state standardized test scores, and SAT and Regents Math test scores. Panel A restricts the experimental sample to Black and Hispanic students, while Panel B restricts the sample to white and Asian students. Due to COVID-related interruptions or the timing of test administration, certain test scores are unavailable for some cohorts. Column 1 excludes from the experimental sample cohorts enrolling in middle school in 2019 and 2020, column 2 the 2018 and 2019 cohorts, column 3 and 4 the 2018 and 2018, column 5 the 2016, 2017 and 2020 cohorts and column 6 excludes the 2018, 2019 and 2020 cohorts. The experimental sample and control variables are as defined in the notes of Table 2. Robust standard errors in parenthesis.

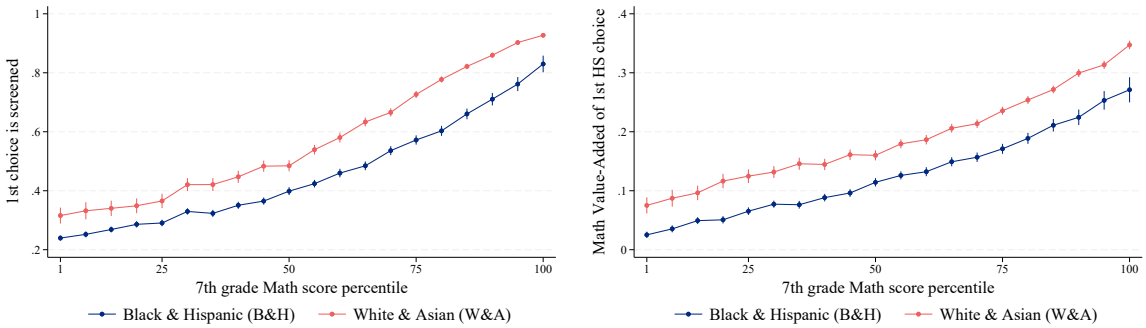
Figures

Figure 1: Differences in High School Choices by Race and Middle School Test Scores

(a) Differences in Student-Body Composition



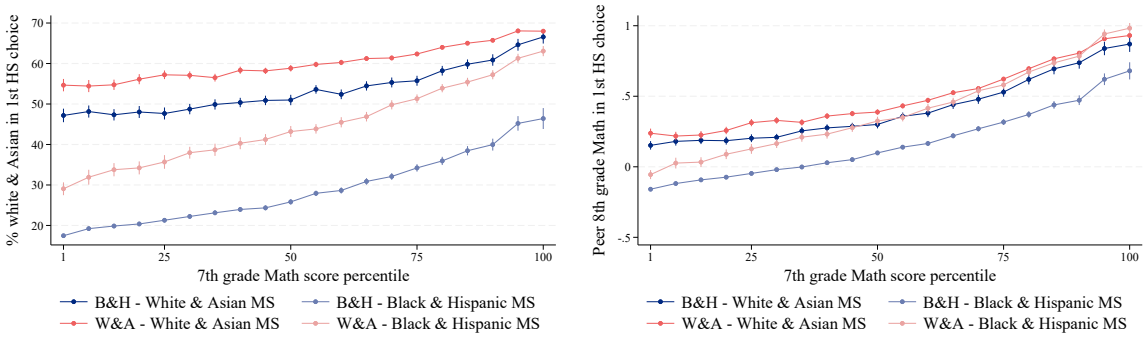
(b) Differences in School Selectivity and Quality



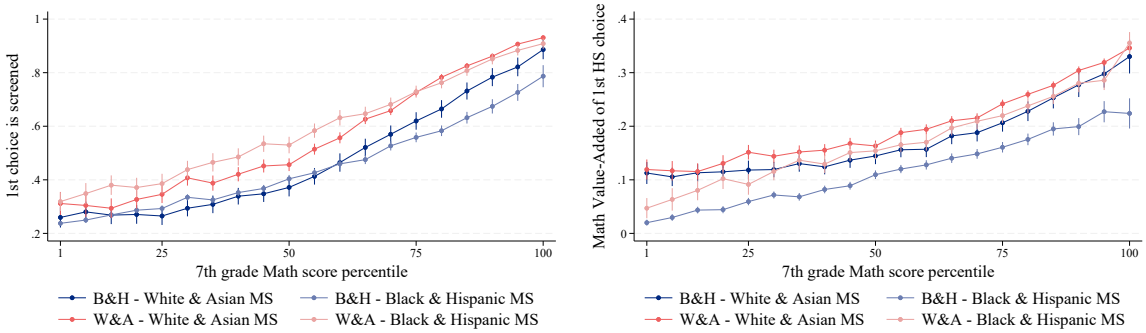
Notes: This figures plots the average characteristics of the high school ranked first by applicants' race and middle school test score percentile. Panel (a) considers the percentage of white and Asian students and the mean 8th grade math scores of students enrolled in the school. Panel (b) considers whether the school is screened and the school risk-controlled Math value-added. The sample is restricted to applicants applying to high schools in years 2018, 2019, 2020, and 2023. The 2021 and 2022 cohorts are not included because 7th grade test scores are not available due to COVID.

Figure 2: Differences in High School Choices Depending on Percentage of white Peers in Middle School

(a) Differences in peer composition

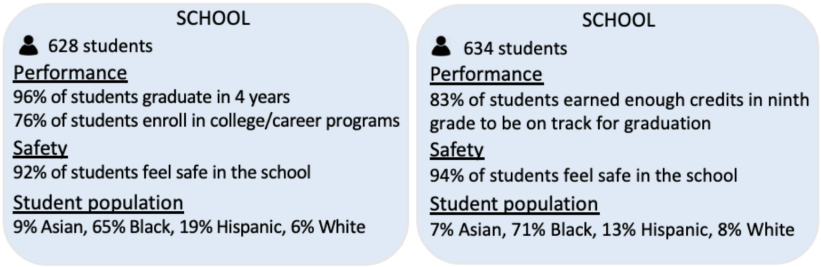


(b) Differences in school selectivity and quality



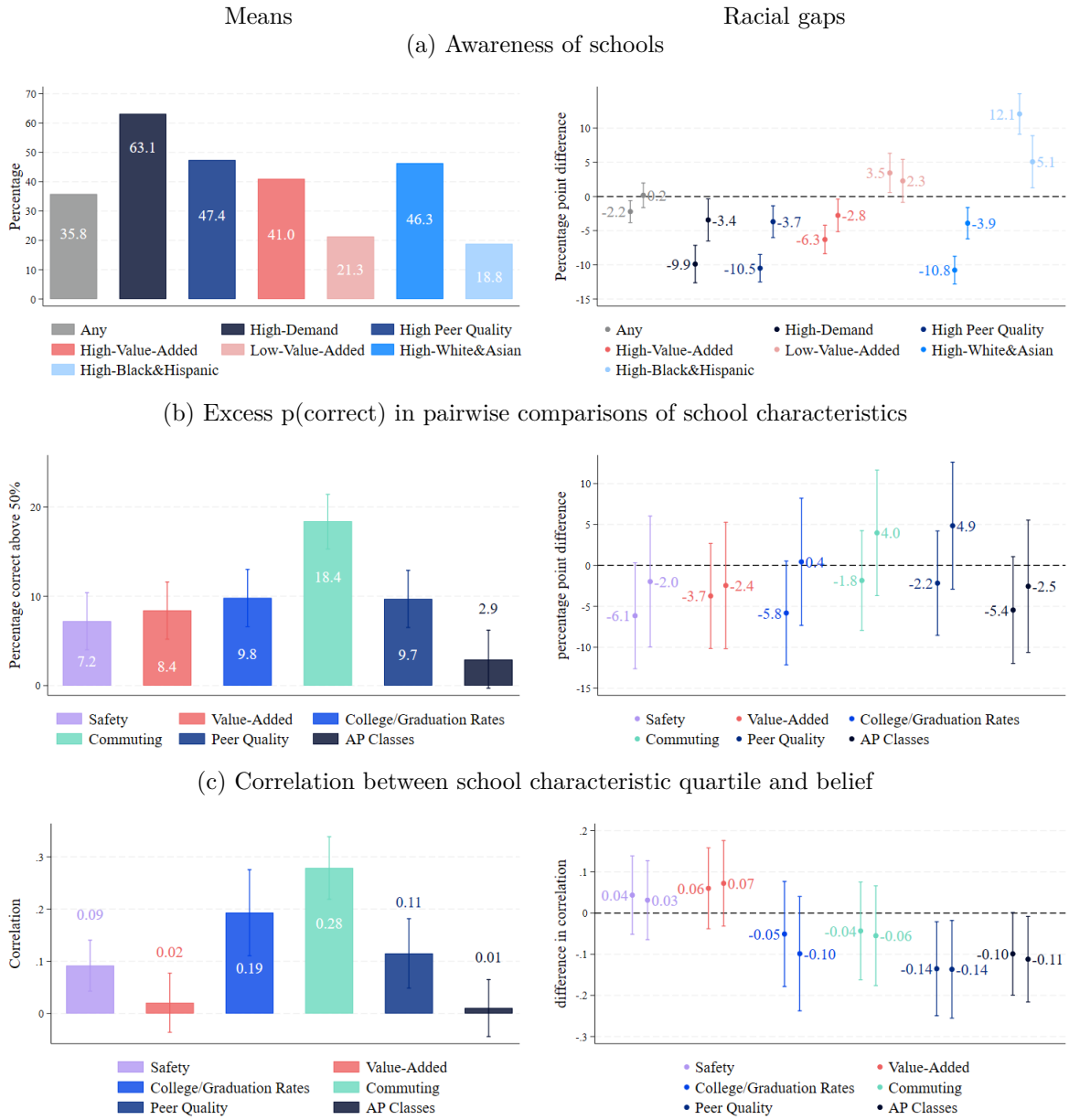
Notes: This figure plots the average characteristics of the high school ranked first by applicants' race, middle school test score percentile, and the racial composition of the middle school attended. High school choice characteristics are shown in lighter shades for students attending majority-white and Asian middle schools ($\geq$ 50% white and Asian enrollment) and in darker shades for those enrolled in majority-Black and Hispanic middle schools. The sample and high school characteristics considered are the same as in Figure 1.

Figure 3: School Cards for Vignette Experiment



Notes: This figure displays examples of two cards used in the vignette experiment. The left card provides precise academic information (Precise Information Treatment arm, received by 60% of survey participants), while the right card presents imprecise academic information (Imprecise Information Treatment arm, received by 40% of survey participants).

Figure 4: Differences in Information About Schools



Notes: This figure reports differences in information about school existence and characteristics. The outcome in Panel (a) is the share of schools respondents were aware of, for each type of school. The outcome in Panel (b) is the percentage of respondents (in excess of 50%) who correctly ranked two schools, for different school characteristics. The outcome in Panel (c) is the correlation of respondents' ranking beliefs with the true ranking of the school, for different school characteristics. The left figure in each panel reports mean outcomes among all respondents, while the right figure shows race differences in responses. In the figures on the right, the first bar depicts the raw percentage point difference across race while the second bar depicts the percentage point difference controlling for district of residence and middle school baseline test score. The capped lines display 95% confidence intervals. This figure uses data from survey questions Q9 and Q10a-g.

References

Abdulkadiroglu, A., Angrist, J., Narita, Y., and Pathak, P. A. (2022). Breaking Ties: Regression Discontinuity Design Meets Market Design. *Econometrica*, 90(1):117–151.

Abdulkadiroğlu, A., Agarwal, N., and Pathak, P. A. (2017a). The Welfare Effects of Coordinated Assignment: Evidence from the New York City High School Match. *American Economic Review*, 107(12):3635–3689.

Abdulkadiroğlu, A., Angrist, J. D., Narita, Y., and Pathak, P. A. (2017b). Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation. *Econometrica*, 85(5):1373–1432.

Ainsworth, R., Dehejia, R., Pop-Eleches, C., and Urquiola, M. (2022). Why do households leave school value added on the table? the roles of information and preferences. Technical report.

Allende, C., Gallego, F., and Neilson, C. (2019). Approximating the equilibrium effects of informed school choice. Technical report.

Andrabi, T., Das, J., and Khwaja, A. I. (2017). Report cards: The impact of providing school and child test scores on educational markets. *American Economic Review*, 107(6):1535–63.

Angrist, J., Gray-Lobe, G., Idoux, C. M., and Pathak, P. A. (2022). Still Worth the Trip? School Busing Effects in Boston and New York. Working Paper 30308, National Bureau of Economic Research. Series: Working Paper Series.

Angrist, J., Hull, P., Pathak, P., and Walters, C. (2021). Credible school value-added with undersubscribed school lotteries. Technical report.

Arteaga, F., Kapor, A., Neilson, C., and Zimmerman, S. (2021). Smart Matching Platforms and Heterogeneous Beliefs in Centralized School Choice. Technical Report w28946, National Bureau of Economic Research, Cambridge, MA.

Bergman, P. (2018). The Risks and Benefits of School Integration for Participating Students: Evidence from a Randomized Desegregation Program. *Working paper*.

Bergman, P., Chan, E. W., and Kapor, A. (2020). Housing search frictions: Evidence from detailed search data and a field experiment. Working Paper 27209, National Bureau of Economic Research.

Bjerre-Nielsen, A. and Gandil, M. H. (2020). Attendance Boundary Policies and the Limits to Combating School Segregation.

Boisjoly, J., Duncan, G. J., Kremer, M., Levy, D. M., and Eccles, J. (2006). Empathy or Antipathy? The Impact of Diversity. *American Economic Review*, 96(5):1890–1905.

Cai, J., De Janvry, A., and Sadoulet, E. (2015). Social networks and the decision to insure. *American Economic Journal: Applied Economics*, 7(2):81–108.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, 82(6):2295–2326.

Campos, C. (2023). Social interactions and preferences for schools:experimental evidence from los angeles. Technical report.

Carlana, M., La Ferrara, E., and Pinotti, P. (2022). Goals and gaps: Educational careers of immigrant children. *Econometrica*, 90(1):1–29.

Carrell, S. E., Hoekstra, M., and West, J. E. (2019). The Impact of College Diversity on Behavior toward Minorities. *American Economic Journal: Economic Policy*, 11(4):159–182.

Chade, H. and Smith, L. (2006). Simultaneous Search. *Econometrica*, 74(5):1293–1307.

Chetty, R., Deming, D. J., and Friedman, J. N. (2023). Diversifying Society's Leaders? The Causal Effects of Admission to Highly Selective Private Colleges. Working Paper 31492, National Bureau of Economic Research. Series: Working Paper Series.

Chetty, R., Friedman, J. N., Saez, E., Turner, N., and Yagan, D. (2020). Income AScergorsesgCatoilolengaesndinInthteergUennietreadtiSotnaatlesM obility. *Quaterly Journal of Economics*.

Chetty, R. and Hendren, N. (2018). The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects*. *The Quarterly Journal of Economics*, 133(3):1107–1162.

Chetty, R., Hendren, N., and Katz, L. F. (2016a). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, 106(4):855–902.

Chetty, R., Hendren, N., Lin, F., Majerovitz, J., and Scuderi, B. (2016b). Childhood environment and gender gaps in adulthood. *American Economic Review*, 106(5):282–88.

Cohen, D. (2021). NYC School Segregation Report Card: Still Last, Action Needed Now.

Conley, T. G. and Udry, C. R. (2010). Learning about a new technology: Pineapple in ghana. *American Economic Review*, 100(1):35–69.

Corno, L., Ferrara, E. L., and Burns, J. (2019). Interaction, stereotypes and performance. Evidence from South Africa. IFS Working Papers W19/03.

Corradini, V. (2024). Information and access in school choice systems: Evidence from new york city.

Epple, D. and Romano, R. E. (2011). Chapter 20 - peer effects in education: A survey of the theory and evidence. volume 1 of *Handbook of Social Economics*, pages 1053–1163. North-Holland.

Frankenberg, E., Ee, J., Ayscue, J. B., and Orfield, G. (2019). Harming our Common Future:. page 45.

Golub, B. and Sadler, E. (2016). 504Learning in Social Networks. In *The Oxford Handbook of the Economics of Networks*. Oxford University Press.

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

Hoxby, C. and Turner, S. (2013). Expanding College Opportunities for High-Achieving, Low Income Students. *Education Next,*, 13(4):66–73.

Hoxby, C. M. and Turner, S. (2015). What high-achieving low-income students know about college. *American Economic Review*, 105(5):514–17.

Idoux, C. (2021). Integrating new york city schools: The role of admission criteria and family preferences.

Kapor, A. J., Neilson, C. A., and Zimmerman, S. D. (2020). Heterogeneous Beliefs and School Choice Mechanisms. *American Economic Review*, 110(5):1274–1315.

Laverde, M. (2020). Unequal Assignments to Public Schools and the Limits of School Choice. *job market paper*, page 48.

Lowe, M. (2021). Types of contact: A field experiment on collaborative and adversarial caste integration. *American Economic Review*, 111(6):1807–44.

Paluck, E. L., Green, S. A., and Green, D. P. (2019). The contact hypothesis re-evaluated. *Behavioural Public Policy*, 3(2):129–158.

Pathak, P. A. and Sönmez, T. (2008). Leveling the playing field: Sincere and sophisticated players in the boston mechanism. *American Economic Review*, 98(4):1636–52.

Patrick Agte, Claudia Allende, Adam Kapor, Christopher Neilson, and Fernando Ochoa (2024). Search and Biased Beliefs in Education Markets. *NBER Working Paper 32670*.

Rao, G. (2019). Familiarity does not breed contempt: Generosity, discrimination, and diversity in delhi schools. *American Economic Review*, 109(3):774–809.

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly Journal of Economics*, 116(2):681–704.

Sacerdote, B. (2011). Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far? In *Handbook of the Economics of Education*, volume 3, pages 249–277. Elsevier.