Choosing Segregation? The Determinants of Racial Gaps in School Choice and the Integrating Effects of Peer Exposure *

Viola Corradini† Clemence Idoux‡

This version: May 26, 2026

Abstract

Combining administrative data with novel survey data from New York City, we document large racial gaps in the value-added and racial composition of school choices, and show that they primarily stem from differences in awareness of high-quality schools and preferences over school demographics. Exploiting quasi-experimental variation in middle school peer composition, we find that exposure to other-race peers leads students to choose and enroll in more integrated high schools, and leads Black and Hispanic students to enroll in higher-value-added schools. These peer effects operate by broadening awareness of high-quality schools and reducing preferences for demographically similar peers.

*We are thankful to the New York City Department of Education's Enrollment Research and Policy office for graciously sharing data, and to participants at the Northeast Economics of Education Workshop, URPP workshop, Michigan State University, UC Riverside, UC Irvine, Columbia University, and MIT 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. An older version of this paper was circulated under the title "Overcoming Racial Gaps in School Preferences: The Effect of Peer Diversity on School Choice". 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.

†Columbia Business School. Email: viola.corradini@columbia.edu

‡University of California San Diego. Email: cidoux@ucsd.edu

1 Introduction

Access to more advantaged schools and peers has lasting benefits for children. Moving to better neighborhoods during childhood raises long-run earnings (Chetty and Hendren, 2018; Chetty et al., 2016b,a), and transferring to wealthier, predominantly White school districts improves test scores and college enrollment (Bergman, 2018; Setren, 2024; Billings et al., 2014). Large urban school districts could generate similar gains without requiring families to relocate or cross district boundaries. Many districts have adopted centralized school choice systems that allow students to apply to schools outside their residential neighborhoods, which should, in principle, equalize access to high-quality schools and reduce school segregation across income and race.

New York City Public Schools offers perhaps the most favorable conditions for centralized choice systems to deliver on their promise: students can apply to any public high school in the city, the student population is diverse, and an extensive public transportation network makes many schools geographically accessible. Yet despite these features, the district remains one of the most segregated school systems in the United States (Cohen, 2021). This paper shows that racial differences in school choice are a key part of the explanation, then asks why these differences arise and what can be done to address them.

We begin by documenting substantial racial differences in families' application choices. Black and Hispanic applicants apply to schools where the share of White and Asian students is about 50% lower than in the schools chosen by White and Asian applicants. These gaps are not explained by differences in student achievement: among applicants with identical test scores, Black and Hispanic families choose schools with a White and Asian share 38% lower than their White and Asian peers. Nor are they explained by residential segregation: differences in the demographic composition of chosen schools are 30% larger than differences in the composition of schools within a 15-minute commute. Families tend to choose schools that are more racially homogeneous than the options near their homes, indicating that choice behavior itself is a key driver of segregation. These gaps extend beyond demographics: Black and Hispanic families also choose schools with value-added that is 0.1 test score standard deviations lower on average, more than a third of a standard deviation in school quality, contributing to inequality in access to educational opportunities. These patterns represent an important caveat for centralized choice systems, which may fail to equalize access or reduce segregation when family preferences differ sharply by race. They may also weaken the competitive pressure on schools predominantly serving Black and Hispanic students

to invest in quality, since such pressure depends on families rewarding higher value-added through their choices. Similar racial gaps have been documented in applications to selective colleges (Black et al., 2015; Hoxby and Avery, 2012; Dillon and Smith, 2017; Campbell et al., 2022), suggesting that the mechanisms we identify may extend beyond K–12 schooling.

Understanding the origins of these choice gaps is crucial for assessing their welfare implications and determining which policy interventions would be most effective in narrowing them. One possibility is that families make different choices because they have different preferences over school characteristics. An alternative explanation is that families differ in the information they possess about schools. Information frictions may prevent some households from learning about high-quality schools or understanding their chances of admission (Allende et al., 2019; Andrabi et al., 2017; Hastings and Weinstein, 2008; Hoxby and Turner, 2015). Distinguishing between these explanations is challenging because observed school choices reflect the joint influence of preferences and information.

We are uniquely positioned to address this challenge: we combine administrative data with a new survey of roughly 3,000 parents and guardians of high-school applicants in the 2022—2023 application cycle. The survey measures families' awareness of schools, beliefs about school characteristics and admission probabilities, aspirations for higher education, and perceptions of discrimination. It also includes a vignette experiment that elicits preferences over school attributes, including academic quality, safety, and the racial composition of peers, while holding other characteristics constant. These data allow us to separately measure information, beliefs, and preferences rather than inferring them solely from observed applications.

We find that racial choice gaps arise primarily from two sources: differences in the set of schools families are aware of and preferences over peer composition. Black and Hispanic families are less likely to know about schools enrolling higher shares of White and Asian students, as well as high value-added schools. They also have less accurate beliefs about school academic performance, measured by college enrollment and graduation rates. On the preference side, the vignette experiment reveals a preference for schools where students would not be part of a racial minority, especially among White and Asian applicants. This preference persists even when respondents receive clear information about schools' academic performance, suggesting it reflects homophily rather than statistical discrimination. By contrast, families across racial groups report similar preferences for academic quality, safety, and college outcomes, and we find no systematic cross-race differences in beliefs about admission chances or in how those beliefs translate into application behavior.

We quantify the relative contribution of each channel by estimating a model of school choice that integrates survey measures with administrative data on ranked applications, school characteristics, and admission probabilities. The decomposition indicates that differences in preferences over peer composition are the primary driver of racial choice gaps in school demographics and value-added, explaining, explaining about two-thirds and one-half of these gaps, respectively. Differences in school awareness account for roughly one-third of the racial choice gaps, while inaccurate beliefs about school characteristics play a more modest role.

Understanding what drives racial choice gaps is useful only insofar as it points to effective remedies. Information interventions may help address differences in families knowledge of school options, but changing preferences over school characteristics is more challenging, and it is less clear how policy can intervene along this margin. We provide new evidence on one single policy-relevant margin that may address both determinants at once: exposure to racially diverse peers earlier in students' educational trajectories. Attending majority-White and Asian schools reduces the raw racial gap in the demographic composition of chosen schools by about one quarter, and reduces the portion of the gap unexplained by differences in achievement and residence by roughly three quarters. Similarly, it closes about 40% of the raw gap, and about 90% of the unexplained gap, in the value-added of chosen schools. These effects are large in absolute terms and also relative to other interventions documented in the literature: attending a majority–White and Asian middle school increases the value-added of chosen schools more than direct interventions that inform families about school quality or school performance, and its effects on the peer composition of schools students choose or attend exceed those of various information interventions (Allende et al., 2019; Patrick Agte et al., 2024; Ainsworth et al., 2023; Hastings and Weinstein, 2008; Campos, 2024). These findings are consistent with prior research indicating that homophily may decline with repeated interaction across social groups (Carrell et al., 2019; Lowe, 2021; Paluck et al., 2019; Rao, 2019; Sacerdote, 2001, 2011) and that information about opportunities often spreads through peer networks (Campos, 2024).

To identify the causal effect of attending racially diverse middle schools on high school choices, we exploit the randomization embedded in the New York City school assignment mechanism. Conditional on students' preferences and school priorities, the centralized matching algorithm randomly assigns seats when programs are oversubscribed. This feature generates quasi-experimental variation in the racial composition of middle school peers. Using instrumental variables methods that build on Abdulkadiroğlu et al. (2017), Abdulka-

diroglu et al. (2022), and Angrist et al. (2022), we thus find that attending middle schools with larger shares of White and Asian students induces Black and Hispanic students to apply to high schools with higher value-added and larger shares of White and Asian students, narrowing a significant portion of the racial choice gap. These changes translate into differences in actual school assignments, leading Black and Hispanic students exposed to more White and Asian peers in middle school to enroll in higher-quality high schools. Exposure to other-race peers also shifts the choices of White and Asian students toward more diverse schools, though these effects are less than half as large and do not extend to school quality.

Schools enrolling different demographic groups may differ along many dimensions; we therefore provide evidence that the effects on high school choices operate through exposure to other-race peers. Instrumental variables estimates with multiple endogenous variables confirm that exposure to other-race peers, rather than other middle school characteristics correlated with peer composition, drives these effects. Accounting for other middle school characteristics, such as peers' baseline achievement, the student-teacher ratio, the availability of a certified counselor, or school value-added, does not affect the estimated effect of exposure to White and Asian peers. Combining survey data with administrative data on students' educational trajectories further reveals the mechanisms at work: exposure to more other-race middle school peers reduces preferences for schools where most students share one's own racial background and broadens the set of schools Black and Hispanic families are aware of. These are the two main determinants of racial choice gaps, indicating that peer exposure operates precisely through the channels driving the gaps in the first place.

Finally, our evidence speaks to a broader question about the effects of school integration. We find that attending a majority-White and Asian middle school improves Black and Hispanic students' 7^{th}- and 8^{th}-grade math test scores by 0.086 and 0.141 standard deviations, respectively. Test score gains are even larger in high school, with Regents Math scores increasing by 0.191 standard deviation, as the direct achievement effects of attending a majority-White and Asian middle school compound with changes in high school enrollment toward higher value-added schools. Importantly, we show that middle school test score gains cannot explain the observed changes in high school application behavior, indicating that integration affects choices through channels beyond academic preparation. Our findings therefore suggest that integration has direct positive effects on achievement, as well as large downstream effects on educational choices that shape future academic outcomes.

This paper contributes to a growing literature studying the determinants of school choices and how they vary by race and socioeconomic status. Prior work has documented socioeco-

nomic differences in applications to selective colleges (Hoxby and Avery, 2012; Black et al., 2015; Dillon and Smith, 2017; Campbell et al., 2022; Chetty et al., 2023) and to higher-quality schools (Allende et al., 2019; Laverde, 2026). A strand of this literature has studied whether information interventions can improve students' choices (Hastings and Weinstein, 2008; Ainsworth et al., 2023; Campos, 2024; Bergman et al., 2020; Patrick Agte et al., 2024), particularly among low-SES students (Hoxby and Turner, 2015; Corcoran et al., 2018; Cohodes et al., 2025; Corradini, 2024), while others have examined tutoring and mentoring interventions aimed at improving aspirations (Carlana et al., 2022).

We contribute to this literature in several ways. We focus on differences in school choice across race — a uniquely important dimension in the U.S. context — and use our survey data to measure their determinants with an unusually high level of detail. On the information side, we are the first to our knowledge to show that awareness of schools plays a central role in generating racial choice gaps. We also find that information frictions alone cannot fully account for racial disparities, as preferences for demographically similar peers, reflecting homophily on both sides of the racial divide, are an independent and important driver. This connects to a separate strand of the literature documenting strong preferences for higher achieving, more advantaged and demographically similar peers (Rothstein, 2006; Hastings et al., 2008; Card et al., 2008; Baum-Snow and Lutz, 2011; Abdulkadiroğlu et al., 2020; Schachner, 2021; Hailey, 2022; Damon Clark et al., 2026). We show that while homophily by race is a key driver of choice gaps, these preferences are themselves shaped by prior exposure to diverse peers, a finding that bridges the literature on school choice with the contact theory literature (Corno et al., 2022; Boisjoly et al., 2006; Carrell et al., 2019; Rao, 2019; Paluck et al., 2019). Our analysis of how middle school peer composition shapes later choices also relates to contemporaneous work by Hahm and Park (2022), who show that attending higher-quality middle schools in New York City leads students to apply to higher-achieving high schools. We establish that racial composition is a key dimension of middle school peer environments, and we identify the specific channels through which it shapes subsequent choices.

This paper also contributes to the large literature on peer effects in education and how the social environment shapes students' longer-run outcomes (Epple and Romano, 2011). Recent work has highlighted the role of social networks in driving persistent gaps in economic outcomes (Chetty et al., 2022; Cattan et al., 2025; Barrios-Fernández et al., 2024). We add to this literature by documenting a novel mechanism through which peer environments shape longer-run trajectories: rather than operating solely through academic achievement, the

margin most studied in the literature, exposure to diverse peers shapes education choices, with downstream consequences for future achievement that exceed the direct effects on test scores.

Finally, we speak to the literature on racial segregation in schools and the effects of desegregation policies (Guryan, 2004; Reber, 2010; Billings et al., 2014; Bergman, 2018; Böhlmark and Willén, 2020; Bjerre-Nielsen and Gandil, 2024; Anstreicher et al., 2022; Setren, 2024). Most of this literature studies inter-district busing and desegregation programs from the 1960s through the 1980s. We offer a contemporary perspective from a large urban district with a centralized school choice system — a setting where institutional barriers to integration are, in principle, low. We document that persistent segregation nonetheless arises from choice behavior itself, and provide novel evidence that the effects of integration extend well beyond test scores to shape the educational choices that determine students' longer-run opportunities.

2 Background and Data

2.1 Data

We combine two data sources to study the drivers of racial differences in school choice. The first is administrative data from NYC Public Schools (NYCPS) covering student applications, enrollment, and test scores from 2013–2014 to 2024–2025. The second is a survey we conducted, in partnership with NYCPS, of guardians of students applying to high school for fall 2023. The survey captures key determinants of school choice, such as information and beliefs, not available in the administrative data. 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 includes the universe of students who apply to or enroll in NYC public schools. Entry into NYC Public Schools (NYCPS) is governed by a centralized assignment system at key transition grades: pre-kindergarten, kindergarten, sixth grade, and ninth grade. At each entry point, families submit ranked lists of school programs, while programs rank applicants according to eligibility criteria and priority rules announced in advance, which may be based on residential location and academic achieve-

ment.[1] The centralized admission system then issues a single school offer to each applicant through the student-proposing deferred acceptance (DA) algorithm.

For our analysis, we track consecutive cohorts of students who apply for 6th-grade seats in public middle schools from 2015–2022 and who later, three years after entry, apply for 9th-grade seats in traditional public high schools from 2018–2025.[2] The administrative records from the centralized middle and high school matches allow us to observe students' rank-ordered school lists, priority categories, and offers - along with enrollment files identifying the school each student ultimately attends. Application and enrollment records are linked to student demographics, standardized state test scores in Math and ELA from 4th to 8th grades, Regents Algebra scores (typically taken in spring of 9th grade), and SAT scores (typically taken in 11th grade).

We use administrative data to analyze students' school choices and to construct measures of middle and high school characteristics. We measure school quality using value-added models (VAM) estimated on students' Math state test scores.[3] We adopt the *Risk-Controlled Value-Added (RC VA)* approach of Angrist et al. (2024), which adjusts for students' school assignment propensity scores.[4] Value-added is reported in test-score standard-deviation units (σ), centered so that the citywide mean school quality is zero. We observe middle and high school racial and ethnic composition and measure school peer quality as the average baseline Math achievement of students in the year prior to enrollment in the school. We also construct measures of high school selectivity and popularity. Selectivity is defined as a binary indicator for schools that screen applicants using test scores, auditions, or other assessments, and popularity as the ratio of rejected to accepted applicants.[5] Value-added is time-invariant, whereas population characteristics and popularity vary by year.

Column (1) of Table 1 includes summary statistics for the sample of middle school applicants who were also observed applying to 9th 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.

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

[2] Applicants who only apply to specialized (exam) or charter schools are excluded. We also exclude the 2021 6th-grade cohort due to missing baseline test scores during COVID.

[3] We use Regents Algebra exams for high school VAM, required for graduation in New York State, and 7th grade state standardized Math test scores for middle school VAM.

[4] Propensity scores are derived from students' applications and the priority status assigned by schools at the time of high school application.

[5] For schools with multiple programs, popularity is the seat-capacity-weighted average of program-level rejection-to-acceptance ratios.

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 0.13σ 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.[6] High school choices vary markedly by race, a pattern explored in greater detail in the next section.

Survey data In partnership with NYC Public Schools, we fielded a post-application survey in February 2023 targeting guardians of 9$^{\text{th}}$-grade applicants. The survey was administered after families submitted high school applications but before match results were released. It was distributed electronically to the email addresses provided during the application process, and only parents or guardians were eligible to respond.[7] The survey was available in English, Spanish, and Chinese. Respondents who completed at least one question received a $10 Amazon gift card.

The survey examined various dimensions of the choice process for families, including sources of information, 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 included a vignette experiment, described in more detail in Section 3.2.1, in which respondents chose between fictitious schools. The experiment was designed to disentangle families' relative preferences for different school characteristics and to assess whether preferences for student-body composition reflected statistical discrimination. We repeated the vignette experiment in a second survey conducted among parents of 2025 high school applicants.[8] The complete survey and detailed information on the construction and randomization of the questions are available in Online Appendix C.

For the 2023 survey round, we selected 21,401 potential participants who were guardians of students enrolled in NYC public middle schools who applied to general education high school seats in 2023 and for whom baseline test scores were available. Of these, 17% completed at least part of the survey, and 15% - the respondent sample in Table 1 - answered more than half of the questions. As shown in column (5) of Table 1, respondents are more

[6]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.

[7]75% of the survey respondents reported that parents and guardians played an essential role in their student's high school selection.

[8]The demographic composition of this round closely mirrored that of the 2023 survey. Appendix Table A1 reports summary statistics comparing the respondent populations across the two surveys.

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
Peer Quality in top 3 choices	0.24	0.21	0.05	0.44	0.34	0.19	0.50
Popularity of top 3 choices	4.31	4.18	3.53	5.14	5.91	4.43	7.39
Lists a screened program in top 3 choices	0.71	0.69	0.60	0.81	0.78	0.68	0.88
Math VA of 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 2022-2023 and then successively applied to high school for enrollment in 2018-2019 to 2025-2026. 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 restricts to applicants who (i) have non-degenerate risk of school assignment and (ii) received a middle school offer in the middle school match. Column 5 to 7 include applicants who responded to our survey in 2023.

likely to be White or Asian and less likely to be low-income than the full population of NYC high school applicants (column (1)). Respondents' students also scored higher on standardized tests than the citywide average.

Although survey respondents are a selected group - likely more resourced and more

engaged parents - Panel C of Table 1 shows that racial differences in their high school choices, reported in columns 6–7, closely resemble those observed in the experimental applicant sample, reported in columns 3–4.[9] In both samples, White and Asian families choose schools that are more selective and more popular, and that have higher value added, larger shares of White and Asian students, and higher baseline student achievement. The next section describes in greater detail how high school choices vary by student race.

2.2 Segregation and Choice

School segregation and the NYC School Assignment System NYC is one of the few places in the United States where reducing school segregation is plausibly within reach: it has a highly diverse student population, an extensive public transportation network, and a centralized school choice system with the potential to weaken the link between residential and school segregation. Despite this centralized school choice system, NYC remains among the most segregated school districts in the country (Frankenberg et al., 2019).

Appendix Figure B1 compares overexposure to Black and Hispanic peers, defined as the difference between the share of Black and Hispanic students in a school and the corresponding citywide grade-level share, by students' race and grade. Across all grades, Black and Hispanic students attend schools where the share of Black and Hispanic peers exceeds the city average of 68%. Segregation is more pronounced in earlier grades. On average, Black and Hispanic students attend elementary schools with 18 percentage points more Black and Hispanic peers than the elementary population, middle schools with 15 points more than the middle school population, and high schools with 11 points more than the high school population. In contrast, White and Asian students attend elementary, middle, and high schools with 28, 23, and 23 percentage points fewer Black and Hispanic students, respectively, than the citywide average.

This pattern is consistent with the greater use of residence-based priorities in admissions at earlier grade levels. While the centralized admissions process operates similarly across entry grades—applicants rank programs, and programs rank applicants based on eligibility and priority rules—the criteria shaping access differ. High schools admit students citywide, and only 39% apply any borough- or zone-based priority. In contrast, residential location continues to play a central role for younger students. At the elementary level, 85% of schools admit only students living within their attendance zone, and the remaining 15% — though

[9]The full applicant sample also exhibits similar patterns across racial groups.

formally non-zoned — still prioritize zoned students. Access to middle schools is similarly constrained: during our study period, 83% of programs limited enrollment to residents of the same local district.[10] Academic admission criteria may further contribute to segregation if Black and Hispanic students receive lower scores. Approximately one-third of middle and high school programs rank applicants based on prior grades, auditions, essays, or behavioral measures.[11]

Segregation by Choice? While school admission and eligibility rules may shape segregation, we show that students' choices themselves are a major driver of high school segregation, and ultimately also generate disparities in access to high-quality schools.

We examine high schools, which rarely prioritize applicants based on residence. Even with minimal geographic screening, school segregation could still mirror residential patterns if students prefer shorter commutes. Figure 1, however, shows that residential segregation in NYC is less pronounced than segregation in high school choices. The figure compares the share of White and Asian students in applicants' top-ranked schools to the demographic composition of schools within a 15-minute public transit commute of their home and of schools reachable within the same commute time as their first-choice option —approximately 37 minutes on average for Black and Hispanic students and 38 minutes for White and Asian students.[12] This comparison isolates whether students' first choices differ meaningfully from equally proximate alternatives.

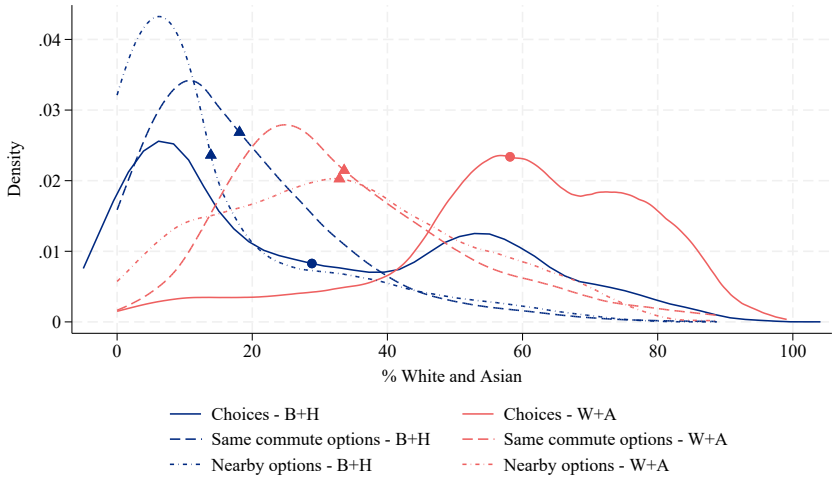
White and Asian students rank, at the top of their lists, schools enrolling a substantially higher share of White and Asian students (59% on average) than both schools accessible within an equivalent commute and those within a 15-minute radius, where the average share is approximately 34%. The distribution for Black and Hispanic students is instead bimodal: most rank schools with disproportionately low White and Asian representation, below the share in same-commute alternatives, while others place majority-White and Asian schools at the top of their lists. These patterns indicate that choice behavior itself, not just where families live, is a key driver of school segregation: racial gaps in chosen schools exceed racial

[10] Of the remaining programs, 14% were borough-wide and 3% were citywide, with 23% of these giving priority to applicants living in or attending schools in specific districts. There are 32 districts in NYC. Starting in 2025, all middle schools accept applications from outside their district, making the process more similar to the citywide high school choice system.

[11] Elementary schools do not consider academic performance in admissions, except for Gifted and Talented programs, which use a separate audition process.

[12] The number of nearby schools also does not differ sharply by race: White and Asian students can reach an average of 18 high schools within a 30-minute public transit commute (2.2 within 15 minutes), compared to 25 (3.6 within 15 minutes) for Black and Hispanic students.

Figure 1: % White & Asian in School Choices and in Proximate Schools



Notes: Solid lines plot the distribution of the share of White and Asian students in applicants' top-ranked schools, by student race. Dashed lines plot the distribution of the share of White and Asian students in schools reachable within the same commute time as the top-ranked school, and dash–dot lines plot the distribution of the share of White and Asian students in schools within a 15-minute commute. Filled circles and triangles indicate the mean of each distribution.

gaps in geographically accessible schools.

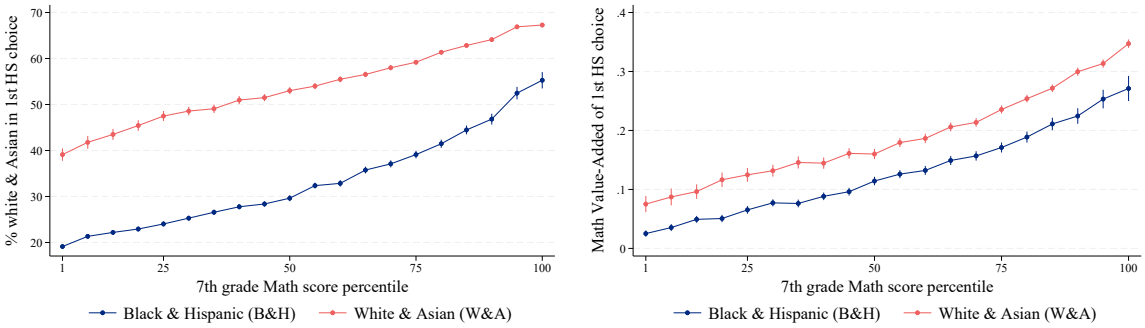
Another explanation for racial differences in school choices is that they could by driven by differences in academic achievement rather than by race per se: students may sort into schools aligned with their own academic ability, particularly where high schools explicitly screen on grades or test scores. Panel (a) of Figure 2, however, shows that choices differ substantially by race even among students with identical baseline achievement. The figure compares top-ranked schools across applicants' race and 7th-grade math scores.[13] Although higher-achieving students prefer schools with higher shares of White and Asian students, White and Asian applicants select schools with roughly 20 percentage points more White and Asian students than those chosen by Black and Hispanic students of comparable academic ability, equivalent to 0.9 between-school standard deviations in the White and Asian share. The racial gap is large compared to differences across the entire achievement distribution: White and Asian students in the bottom quartile of 7th-grade Math scores choose schools with a similar White and Asian share to that of Black and Hispanic students in the top quartile.

These differences in choices matter not only for segregation but also for access to high-

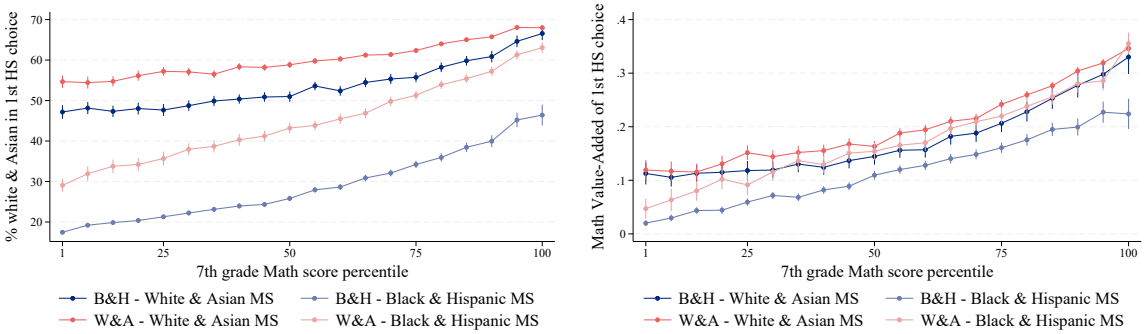
[13]To reflect the information available at the time of application, school characteristics are based on the demographic composition of the incoming 9th-grade cohort.

Figure 2: Differences in High School Choices by Race

(a) Conditional on 7th Grade Achievement



(b) Conditional on 7th Grade Achievement and Middle School Demographics



Notes: This figures plots the average White and Asian share (left) and value-added (right) of the high school ranked first by applicants' race and middle school test score percentile. Panel (b) distinguishes students by 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 is restricted to applicants applying to high schools in years 2018, 2019, 2020, 2023, 2024, and 2025. The 2021 and 2022 cohorts are not included because 7th grade test scores are not available due to COVID.

quality schools. Conditional on baseline achievement, Black and Hispanic students select schools with an approximately constant 0.06σ lower Math value-added (0.22 of a standard deviation in school quality), and these gaps persist after controlling for district of residence, confirming they are not explained by residential segregation.[14] This gap in value-added follows directly from the positive correlation between school value-added and White and Asian shares, combined with Black and Hispanic students' lower likelihood of choosing schools with higher White and Asian shares, rather than from a lower propensity to choose high value-

[14]Panel (a) of Appendix Figure B2 plots these racial gaps by baseline achievement, controlling for district of residence. This is, if anything, a conservative test: students living in different neighborhoods may be at equal distance to the same school, but be exposed to different socioeconomic environments, social networks, and other factors that shape school choices.

added schools, conditional on school demographics. [15] Because mean baseline achievement and selectivity are also positively correlated with the share of White and Asian students, racial differences in choices extend across all of these dimensions, as shown in Appendix Figure B3.

Taken together, this descriptive evidence suggests that school segregation and racial disparities in access to high-quality schools reflect differences in choices, not only differential admission to the same schools. White and Asian students actively choose schools with higher value-added, higher peer quality, greater academic screening, and higher non-minority shares relative to Black and Hispanic students living in the same neighborhoods and with identical test scores. Race thus emerges as a distinct dimension shaping school choice, one that operates independently of academic achievement, residential location, and even socioeconomic background.[16]

Peer Environment and High School Choice It is not only students' own race that matters: the racial composition of their middle school peers does too. Panel (b) of Figure 2 illustrates this pattern. Two findings stand out. First, the demographics of students' high school choices vary strongly with the demographics of their middle school peers: Black and Hispanic students from majority White and Asian middle schools select high schools with roughly 25 percentage points more White and Asian students than those from majority Black and Hispanic middle schools, while White and Asian students from majority Black and Hispanic middle schools select schools with about 15 percentage points fewer White and Asian students.[17] Second, middle school peer composition also affects access to high-quality schools: under-matching is less pronounced among Black and Hispanic students from majority White and Asian middle schools, who choose higher value-added and more selective schools with higher baseline achievement. In contrast, White and Asian students' preferences for these school attributes vary little with middle school composition.[18] While selection into

[15] This is confirmed by a student-school level regression of an indicator for school choice on school value-added and school racial composition: conditional on racial composition, the correlation between choice and value-added does not differ by race.

[16] We do not directly observe income, but use eligibility for free or reduced-price lunch (FRPL) as a proxy. Racial gaps in both the White and Asian share and value-added of top-ranked high schools are approximately twice as large as the corresponding gaps by FRL status, suggesting that race shapes school choice above and beyond socioeconomic differences.

[17] Replicating the analysis using 4th-grade math scores does not meaningfully change the results. These results are available upon request.

[18] Again, these differences across middle schools are not driven by residential location: they remain similar after controlling for district of residence (Appendix Figure B2, Panel (b)). Replicating these figures using 4th-grade math scores on the x-axis does not meaningfully change the results. Results are available upon

middle schools may account for part of these differences, in Section 4 we show that they reflect causal effects of middle school peer composition.

That racial choice gaps vary with the school environment suggests that preferences may be malleable, influenced by information or peers, and therefore potentially responsive to policy. Understanding what drives these differences is essential to design effective interventions, but administrative data alone cannot capture families' information, beliefs, or perceptions. In the next section, we make progress on this front using a comprehensive survey designed to measure these factors.

3 Determinants of the Racial Choice Gap

3.1 Conceptual Framework

To shed light on the forces driving racial differences in school choice, we adopt a simple school choice framework. Applicants, indexed by i, choose which high schools to apply to by solving a portfolio choice problem as in Chade and Smith (2006). Each applicant chooses a ranked-ordered list (ROL) of schools $R \in \mathcal{R}_i$, where $\mathcal{R}_i$ is the set of all k-permutations of A_i, the set of schools known to applicant i.[19] Each ROL can be mapped to a lottery over schools with probabilities determined by both the order of schools in the list and the applicant's beliefs about admission chances. Applicants thus choose their ROL to maximize expected utility, which can be written as:

$$\max_{R \in \mathcal{R}_i} \sum_{s \in A_i} p_{is}(R, \widehat{q}_i) u_{is}(\theta_i, E[X_{is}|\mathcal{I}_i]) - c_i(R) \tag{1}$$

The utility that student i derives from attending school s, denoted $u_{is}(\theta_i, X_{is})$, depends on the student's preferences over school attributes X_{is}, including distance from the school, parameterized by the preference vector θ_i. Because students may have imperfect knowledge of school attributes, they form expectations about them based on their (potentially inaccurate) beliefs. These beliefs constitute $\mathcal{I}_i$, the information available to student i at the time of application.

The weight assigned to each school's utility, $p_{is}(R, \widehat{q}_i)$, represents the subjective probability that student i is *assigned* to school s. Under the deferred acceptance mechanism,

request.

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

applicants' probabilities of *admission* to each school are independent of the rank-order list they submit, and we assume that applicants are aware of this property. Thus, $p_{is}(R, \widehat{q}_i)$ depends separately on the choice of R and on applicants' subjective beliefs $\widehat{q}_i$ about *admission* probabilities. These beliefs reflect students' understanding of each school's admission criteria as well as their own perceived standing relative to other applicants (e.g., in terms of priority status or test scores).

Finally, each applicant incurs a cost $c_i(R)$ when submitting her ROL. This cost captures the psychological burden associated with each ROL, given the applicant's information and beliefs. For example, listing highly selective programs may generate a psychological cost if students anticipate disappointment from rejection. Although likely small, this cost rationalizes why applicants may exclude schools where their admission chances are slim, or why they may submit short lists when they are nearly certain of 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 - $E[X_{is}|\mathcal{I}_i]$. [20]

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.[21] Applicants' socio-economic and racial background might affect their

[20] 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., 2023; Bergman et al., 2020; Andrabi et al., 2017; Allende et al., 2019; Campos, 2024; Corradini, 2024)

[21] Kapor et al. (2020) and Arteaga et al. (2022b) find that beliefs about admission chances differ from rational expectations values using survey evidence in a similar context.

degree of optimism and confidence in their relative ability and thus influence their perceived chances of admissions.

3.2 Survey Evidence

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. Subsections 3.2.1 - 3.2.3 analyze survey responses to identify which of these factors are important to explain racial differences in school choices.

3.2.1 Preferences for School Characteristics and Peers

Isolating families' preferences for school characteristics, such as school demographics, using administrative data on school choices is challenging because observed school features may be correlated with unobserved factors. To address this issue, we embedded a vignette experiment in our survey, following prior studies that use this approach to measure preferences for school racial composition (Tedin and Weiher, 2004; Billingham and Hunt, 2016; Haderlein, 2022; Hailey, 2022). This method allows us to manipulate school characteristics independently while holding all other school attributes constant.

We elicited both cardinal (on a scale from 1 to 6) and ordinal (ranked) preferences for hypothetical schools. Each respondent rated their likelihood of applying to two hypothetical schools on a 1–6 scale and ranked two sets of three schools in order of preference. The hypothetical schools were described to respondents as identical except for their safety rating, academic performance rating, and racial composition.[22] The experiment used a factorial design to randomly combine these characteristics, as illustrated in Appendix Figure B4.

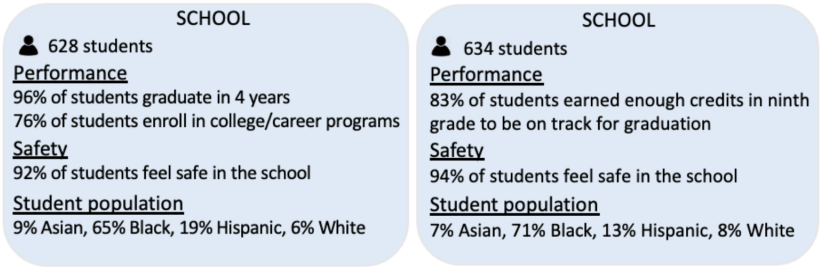
Hypothetical schools varied in safety (high vs. low) and racial composition: balanced and representative of the district, majority Black, majority Hispanic, or majority White and Asian. A randomly selected 60% of respondents to the 2023 survey also received precise academic performance information, including the four-year graduation rate and college and career program enrollment rate.[23] The remaining 40% received only imprecise information, defined as the share of students earning enough 9th-grade credits to be on track for graduation, which is only weakly correlated with graduation and college enrollment rates.

[22] We adapt the prompt of the question used in Hailey (2022).

[23] Safety, graduation rates, and college enrollment are school performance indicators displayed to families on the application portal. We repeated the vignette experiment with high school applicants' guardians in 2025. In 2025 everyone received precise academic performance information.

Participants in the imprecise-information group were presented with a school of median performance, while those in the precise-information group were randomly assigned high- or low-performing schools. Figure 3 shows an example vignette as seen by participants, and Table A2 summarizes the information presented in each treatment arm.[24] The factorial design resulted in 16 unique combinations for the precise-information treatment arm and 8 unique combinations for the imprecise-information treatment arm.[25]

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).

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

$$u_{is} = a_{r(i)} + b_{r(i)} X_s + e_{is}$$

.

X_s includes the characteristics of school cards: high safety level, majority-Black, majority-Hispanic, majority-White and Asian, and high academic performance. a_r captures, separately for each racial group $r \in \{White/Asian, Black/Hispanic\}$, the average utilities respondents would derive for attending a low-safety, racially-balanced and low-achievement hypothetical school compared to their outside option, while b_r captures 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 independent and identically distributed utility shocks.

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

[25] The two schools for the cardinal preferences part of the experiment were randomly selected without replacement from these unique combinations. For the ordinal preferences part, two sets of three distinct schools were randomly chosen without replacement.

We combine the absolute preferences for schools and relative rankings of schools provided by respondents to estimate their respective weights for different school characteristics.[26] Respondents' reports of their likelihood of applying to the schools provide a normalization for the scale and location of the utility. A respondent indicating a likelihood of x (on a scale from 1 to 6) to list a school implies that $u_{is} < 1.5]$ for $x = 1$, $u_{is} \in [x - 0.5, x + 0.5]$ for $x \in \{2, 3, 4, 5\}$ and $u_{is} > 5.5]$ for $x = 6$. The full parameter vector $\vartheta = (a, b, \sigma)$ is estimated using a Gibbs sampler to maximize the likelihood of observing the responses to both questions.

Table 2 reports the estimates in units of the 1-6 likelihood-of-applying scale from the cardinal preferences experiment. Columns (1) and (2) show that school academic performance is the strongest determinant of school choice across all respondents. Holding other attributes constant, a high academic rating raises utility by roughly 1.1 point for Black and Hispanic respondents and by about 1.4 points for White and Asian respondents. Safety also increases utility for both groups, although its effect is smaller than that of academic performance.

Table 2: 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	3.06***	2.29***	3.52***	3.04***
	(0.07)	(0.06)	(0.11)	(0.10)
High-academics	1.08***	1.42***		
	(0.05)	(0.05)		
High-safety	0.46***	0.65***	1.32***	1.08***
	(0.05)	(0.05)	(0.10)	(0.09)
Majority Black	-0.28***	-0.59***	-0.45**	-0.69***
	(0.07)	(0.06)	(0.14)	(0.13)
Majority Hispanic	-0.15**	-0.31***	-0.37***	-0.64**
	(0.07)	(0.06)	(0.14)	(0.12)
Majority White	-0.12	0.31***	-0.38***	0.37***
	(0.07)	(0.07)	(0.13)	(0.13)
N respondents	3,702		811	

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.

Preferences over peer composition also matter, particularly among White and Asian re-

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

spondents. Column (2) shows that White and Asian respondents prefer majority-White and Asian schools to racially balanced schools, and racially balanced schools to majority-Black or majority-Hispanic schools. These demographic preferences are sizable: White families' preference for a majority-White school over a majority-Black school is larger than the effect of moving from a low-safety to a high-safety school. In contrast, Black and Hispanic respondents tend to favor racially balanced schools over all other school types, and place approximately half as much weight on racial composition as they do on safety.

The strength of demographic preferences depends on the information families receive about academic quality. When performance data are imprecise, respondents rely more heavily on racial composition, suggesting that demographic preferences in part reflect statistical discrimination: families may infer school performance from student demographics. Comparing Columns (1)–(2) with (3)–(4) reveals that limited information amplifies aversion to majority-Black and majority-Hispanic schools across all racial groups. Black and Hispanic respondents also penalize more the majority-White schools when they lack detailed performance information.

However, the preferences of White and Asian respondents cannot be explained solely by statistical discrimination. Even when schools' academic performance is made explicit, White and Asian respondents continue to favor majority-White or Asian schools and disfavor majority-Hispanic or majority-Black schools. With precise information, they rate majority-White/Asian schools roughly 0.31 points higher and majority-Hispanic and majority-Black schools 0.31 and 0.59 points lower, respectively, relative to racially balanced schools.[27] In contrast, Black and Hispanic families only exhibit aversion to majority-White/Asian schools when they lack reliable academic information, indicating that demographic preferences play a more limited role for them once uncertainty about academic quality is reduced.

Taken together, the vignette experiment points to homophily in preferences for school demographics, especially among White and Asian respondents, rather than differences in how families value school safety or academic performance, as a key driver of racial differences in actual school choices. Reported priorities are consistent with these findings. When asked directly, respondents identify largely similar factors as most important. Panel (a) of Appendix Figure B5 lists the most frequently cited school attributes, their overall selection rates, and racial differences in the share rating each as important, along with confidence intervals.[28] For

[27] Table 2 also suggests that White and Asian respondents may have stronger outside options, as they rate the reference school 0.77 points lower than Black and Hispanic respondents.

[28] The first coefficient reflects raw cross-race differences; the second conditions on district of residence and middle school test scores.

the three most frequently mentioned attributes (safety, value-added, and graduation/college outcomes) selection rates do not differ significantly across racial groups. Similarly, Panel (b) shows that, although Black and Hispanic families are 11 percentage points less likely to view college as essential for success, they are just as likely as White and Asian parents to hope their children will attend a four-year college once baseline achievement is controlled for. Where differences emerge is in attitudes toward belonging and peer composition. Black and Hispanic families place less weight on "sense of belonging", which may help explain why homophily is stronger among White and Asian respondents.[29]

3.2.2 Information about Schools

Differences in families' school choices outside the vignette experiment may also stem from differences in information. The conceptual framework highlights two distinct aspects of information: the schools families are aware of and the accuracy of their information on specific school characteristics. We examine each in turn.

Awareness Sets

To measure "awareness sets", we asked respondents to identify which schools they recognized from a list of ten. The specific schools shown to each respondent were randomized based on the student's district of residence, so that they were selected to be relatively close to the respondent's home, popular among students in that district, and diverse in characteristics.

Panel (a) of Figure 4 shows the share of schools recognized for different types of schools (numbers on the left), and racial differences in school recognition (coefficients on the right). The darker-shade coefficient for each school type measures raw racial differences controlling for school fixed effects, while the lighter-shade also conditions on distance to the school and student baseline Math achievement.[30]

On average, respondents are familiar with about one-third of the schools listed, regardless of race. However, Black and Hispanic families are significantly less likely to recognize schools with high White and Asian enrollment, high popularity, high value-added or high peer quality.[31] Even conditional on distance to the school and baseline achievement, they

[29] Concerns about peer treatment may further reinforce these patterns. As reported in Appendix Figure B6, Asian respondents are the most likely to anticipate discrimination by peers (24% versus 12% among White respondents). Black and Hispanic families also report slightly greater concern than White families, although these differences are not statistically significant. Among the 23% of families who report concerns about discrimination by peers or teachers, 70% indicate that these concerns influence their school choices.

[30] Only controlling for distance explains a very small share of racial gaps in awareness sets, while controlling for students' baseline achievement roughly explains 50% of racial gaps in awareness.

[31] High-demand schools were selected among the most popular schools in a respondent's borough, as

are 10% less likely to recognize schools with high White and Asian enrollment, and 6% less likely to know schools with high value-added. Conversely, they are 19% more likely to know low value-added schools and 36% more likely to recognize schools with predominantly Black and Hispanic students.[32]

Accuracy of Information

We then examine differences in knowledge about school characteristics using two sets of questions. Our preferred measure asks respondents to evaluate a single school by selecting the quartile they believe best represents its position in the borough on a given characteristic; the outcome of interest is the correlation between the actual quartile ranking and respondents' beliefs.[33] The second measure asks respondents to compare pairs of schools - for example, identifying which has a higher graduation rate or better safety rating - with a 50% accuracy rate expected by chance.[34]

Panel (b) of Figure 4 presents results using our preferred measure.[35] Families are imperfectly informed about school characteristics, and tend to be better informed about easily observable factors, such as commuting distance, than harder-to-assess measures like safety and value-added. Knowledge is most accurate for commuting distances, followed by college graduation rates and peer quality (measured by average 7th grade Math achievement). By contrast, beliefs about school value-added and safety are only weakly correlated with actual quartile rankings, and families appear uninformed about AP course offerings.

Black and Hispanic respondents hold less accurate beliefs than White and Asian respondents about some academic attributes, such as college enrollment rates and AP course offerings. However, their information appears to be similarly accurate for school value-added,

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% 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).

[32]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 the NYCPS website. They are also less likely to use more time-intensive sources, such as individual high school sessions or non-official websites.

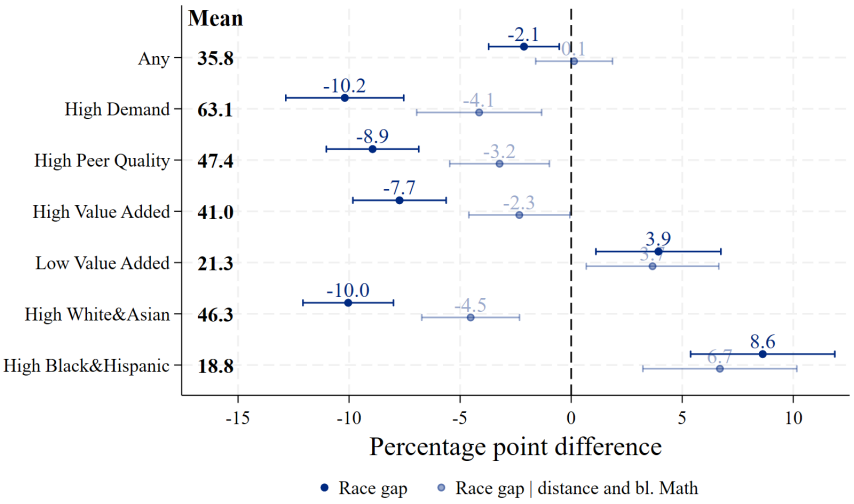
[33]To increase precision, we repeated this set of questions in 2025. In the 2025 survey we did not restrict to schools in the same borough and asked about percentile positions within the city instead.

[34]The two schools in each pair were selected to differ meaningfully along the relevant dimension, ensuring questions were neither overly challenging nor ambiguous.

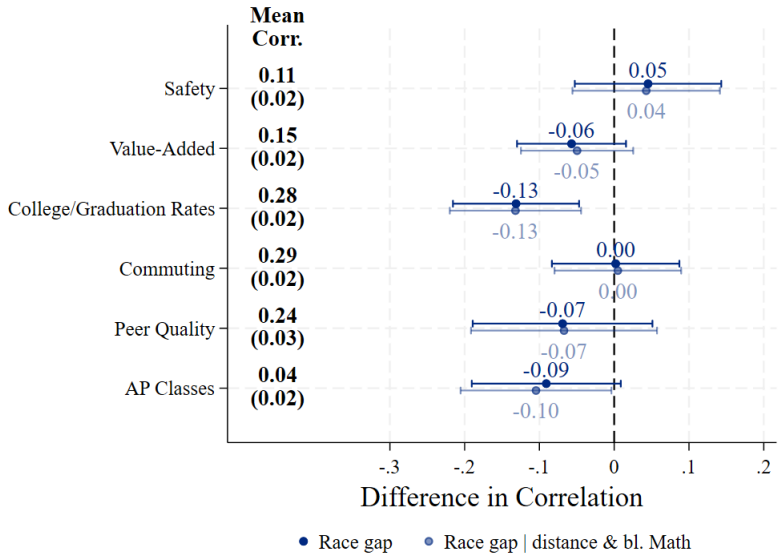
[35]Appendix figure B8 compares these to results using pairwise school comparisons, which are significantly noisier, partly due to the smaller sample size.

Figure 4: Differences in Information About Schools

(a) Awareness of schools



(b) Accuracy of Information - Corr(Beliefs, School Characteristics)



Notes: This figure examines racial differences in awareness and accuracy of information about schools. Panel (a) reports the share of schools known to respondents, by school type. Panel (b) shows the correlation between actual school characteristics and respondents' believes about them, measuring both as quartile ranks. For each measure, the numbers on the left show overall mean outcomes and the bars on the right show average within-school racial differences, reported both raw and conditional on distance and students' baseline Math achievement. Error bars represent 95% confidence intervals. This figure uses data from survey questions Q9 and Q10a-g.

peer quality, safety, and commuting distance. Although these information gaps could contribute to observed differences in school choices, beliefs among White and Asian respondents

are also only weakly correlated with actual school rankings, and beliefs across all groups tend to cluster toward the middle, suggesting widespread imperfect information. Overall, racial differences in the accuracy of beliefs about school characteristics appear less pronounced and less pervasive than racial differences in awareness of school options.

3.2.3 Beliefs about Admission Probabilities

The final component of our framework is beliefs about admission probabilities. When forming application lists, families may strategically omit schools they perceive as out of reach.[36] If perceptions of admission chances differ by race, such differences could contribute to racial disparities in applications even among students with similar achievement and residential location.

Panel (a) of Figure 5 and Panel A of Table 3 show that perceived admission chances in fact play a role in whether students apply to their dream school, the program they would choose if acceptance were guaranteed. Applicants who believe admission is certain are 26 percentage points more likely to report applying than those who believe they have no chance. In the latter group, only 37% apply.[37] Beliefs also predict applications to high-demand schools, though less strongly, since not all students prefer these programs. Importantly, responses to beliefs are statistically similar across racial groups: for any given belief about admission chances, Black and Hispanic respondents are equally likely to apply as White and Asian respondents.

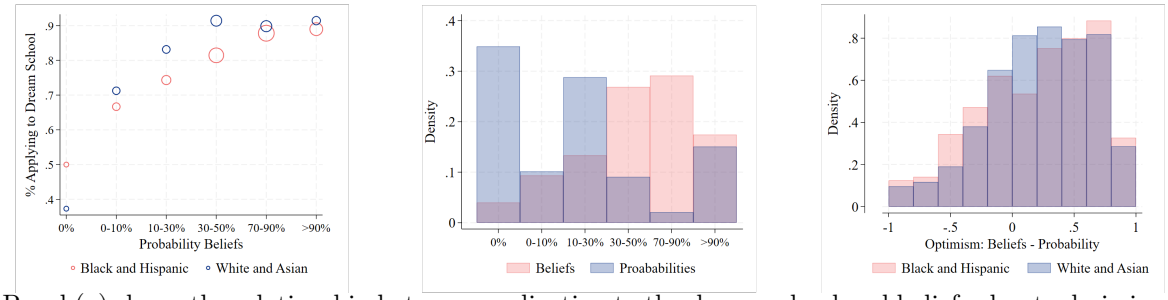
Panel B of Table 3 shows that families' understanding of their actual admission chances is broadly similar across race. A one–point increase in actual admission chances translates into a 24 percentage point increase in perceived admission probability and a 12 percentage point increase in the likelihood of applying to the dream school. These gradients remain similar after conditioning on baseline academic achievement. The application-to-dream-school gradient is similar across racial groups, while belief-updating gradient is somewhat smaller for Black and Hispanic respondents but remains positive. Panel (c) of Figure 5 similarly shows that the distribution of optimism (difference between perceived and actual chances) is comparable across race and if anything, slightly more right-skewed among Black and Hispanic respondents.

[36] This behavior arises in deferred acceptance settings when applicants face any application cost. Idoux (2021) provides evidence supporting this claim in the context of NYC.

[37] Consistent with beliefs shaping decisions, more than one-third of respondents reported revising their lists after observing their random lottery number, which was shown to applicants in the year we ran the survey.

Figure 5: Beliefs about Admission Chances and Application Behavior

(a) % Applying to Dream School (b) Beliefs vs. Actual Chances (c) Optimism by Race



Notes: Panel (a) shows the relationship between application to the dream school and beliefs about admission chances. Panel (b) plots the distribution of beliefs about admission chances and of actual admission chances, for the subset of schools respondents were asked about. Panel (c) plots the distribution of optimism across respondents' race. Actual admission chances depend on applicants' observed priorities, scores and random number; they are computed through simulations of the assignment mechanism by resampling applicants and random numbers.

Consistent with these results, parents across racial groups report similar perceptions of their child's academic competitiveness, indicating that the lower application rates of Black and Hispanic students to selective schools are unlikely to be driven by differences in beliefs about their child's academic strengths. Panel C shows that respondents tend to view their children as stronger than predicted by their middle school test scores: students in the bottom tercile of academic performance are perceived, on average, to be at the median (2 on a 1–3 scale). If anything, Black and Hispanic respondents exhibit slightly more optimism at similar achievement levels. Perceptions relative to high-performing peers at competitive schools follow a similar pattern: they rise with test scores, as do perceptions of citywide academic standing, and if anything the slope is somewhat steeper among Black and Hispanic parents.

Taken together, these results indicate that Black and Hispanic students are not less likely to apply to popular or selective schools because they underestimate their admission chances or respond differently to those chances. Conditional on real admission chances, all groups update beliefs similarly, and those beliefs translate into comparable increases in applications to "dream schools" or schools in high demand. Thus, gaps in application behavior to popular and selective schools are unlikely to stem from differences in strategic behavior or confidence in admission likelihood.

Table 3: 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	0.237***	0.202***	0.115***	0.096**
	(0.028)	(0.028)	(0.038)	(0.039)
(Admission probability) × (Black or Hispanic)	-0.077**	-0.066**	0.019	0.024
	(0.034)	(0.033)	(0.051)	(0.051)
Black or Hispanic	0.033	0.057**	-0.012	-0.004
	(0.028)	(0.027)	(0.040)	(0.040)
7th grade test scores controls		X		X
Mean white&Asian, low probability	0.435	0.435	0.813	0.813
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) X (Black or Hispanic)		-0.103**		0.086**
		(0.044)		(0.043)
Black or 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' perceived and real relative academic performance (Panel C). Students' relative performance is measured as the tercile in the city-wide distribution of test score (Column 2) or in the distribution of test scores within specific high-demand schools (Column 4). All models control for residential district, and for high-demand or dream school fixed effects, when relevant. Columns 2 and 4 of Panels A and B control for 7th grade test scores. 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).

3.3 Model Estimation

To quantify the contribution of each factor to the observed racial gaps, we estimate the model in equation (1). We proceed in four steps, corresponding to the four components of the applicant's problem: awareness sets, beliefs about school characteristics, beliefs about admission chances, and preferences. Throughout, we estimate all parameters separately by applicant race group and assume that the shocks governing awareness, beliefs, and preferences are independent.

Let $r(i) \in \{Black, Hispanic, White/Asian\}$ denote applicant i's race group. For each applicant-school pair, define

$$W_{is} \equiv \Big(1, d_{is}, Z_s\Big),$$

where d_{is} is commute distance and Z_s is a vector of school characteristics including the share of White and Asian students, the share of Hispanic students, college enrollment rate, value-added, safety, and peer quality.

Awareness sets. Let π_{is} indicate whether school s belongs to applicant i's awareness set. We model this probability as:

$$\pi_{is} = \mathbf{1}\{\pi^*_{is} > 0\}, \qquad \pi^*_{is} = W_{is}\lambda_{r(i)} + \nu_{is},$$

with $\nu_{is} \sim N(0,1)$. We estimate λ_r by probit maximum likelihood using the survey question that asks respondents whether they recognize each school shown to them.[38]

Beliefs about school characteristics. We model beliefs about school attributes (peer quality, college enrollment rate, safety, and value-added percentiles) using a hierarchical linear model with school and school-by-race random effects:

$$\widehat{X}^j_{is} = W_{is}\gamma^j_{r(i)} + \eta^j_s + \eta^j_{s,r(i)} + v^j_{is}, \qquad v^j_{is} \sim N(0, \sigma^2_j),$$

where j indexes the school attribute, η^j_s is a school-level random intercept, and $\eta^j_{s,r(i)}$ is a school-by-race random intercept. The school random effect captures the fact that some schools are systematically perceived as better or worse than their observables would predict, while the school-by-race random effects allow these perception deviations to vary across

[38] We choose the threshold for π^*_{is} so that the share of schools predicted to be in the awareness set within the mean commute distance of schools in the survey data matches the corresponding share in the survey data.

racial groups. We estimate this model by restricted maximum likelihood using the survey questions that elicit respondents' beliefs about school characteristics.

Beliefs about admission probabilities. We model subjective admission beliefs as a function of the applicant's actual admission probability and overall school popularity using a hierarchical linear model:

$$\widehat{q}_{is} = Q_{is}\zeta_{m(i)} + \xi_s + \xi_{s,m(i)} + \upsilon_{is}, \qquad \upsilon_{is} \sim N(0, \sigma_\upsilon^2),$$

where $m(i)$ denotes applicant i's minority status, ξ_s is a school random intercept, $\xi_{s,m(i)}$ is a school-by-minority-status random intercept, and Q_{is} includes a constant, the applicant's actual admission probability at school s and its square, a dummy for undersubscribed schools, and a dummy for highly oversubscribed schools. This model is estimated by restricted maximum likelihood using the survey questions that elicit perceived chances of admission.

Preferences and application costs. We specify the utility from attending school s, given beliefs about its characteristics, as:

$$u_{is} = \left(d_{is}, \widehat{X}_{is}, R_s\right)\theta_{r(i)} + \epsilon_{is}, \qquad \epsilon_{is} \sim N(0, 1),$$

where R_s denotes the racial composition of school s (percentile of White/Asian share and Hispanic share), assumed to be known by students, and ϵ_{is} is a normal preference shock with variance normalized to one as the scale normalization.

To estimate preferences, we follow the strategic ranking framework of Idoux (2021). Under this framework, applicants form their rank-ordered lists strategically, taking into account their beliefs about admission probabilities at each school. Specifically, an applicant lists school s at position k in her rank-ordered list if and only if the expected benefit of adding that school exceeds the cost of including it:

$$p_{isk} \cdot u_{is} \geq c_i,$$

where p_{isk} is the probability that school s is the first school to offer the applicant a seat given the higher-ranked schools on her list, and c_i is applicant i's cost of adding an additional school to her list. The applicant considers schools in order of decreasing utility and lists each school if and only if this condition is satisfied. This behavioral rule generates rank-ordered lists

that can be shorter than the maximum allowed length when the marginal benefit of listing additional schools falls below the application cost.

We model application costs as:

$$c_i \sim \text{TruncNormal}\Big(\bar{c}_{r(i)}, \sigma^2_{c,r(i)}, 0, +\infty\Big),$$

where $\bar{c}_r$ and $\sigma_{c,r}$ are the mean and standard deviation of the cost distribution for race group r, truncated below at zero.

We estimate θ_r, $\bar{c}_r$, $\sigma_{c,r}$, and the latent utilities u_{is} jointly using a Gibbs sampler, plugging in the estimates $\widehat{\pi}_{is}$, $\widehat{X}_{is}$, and $\widehat{q}_{is}$ from the first three steps. The sampler iterates between drawing preference parameters conditional on latent utilities and costs, drawing latent utilities conditional on preferences and costs (subject to the ordering constraints implied by observed rank-ordered lists), and drawing individual costs conditional on utilities (subject to the listing rule above). We report preference estimates rescaled so that the coefficient on commute distance equals -1, so all parameters are expressed in units of willingness to travel (minutes).

We also estimate a version of the model that combines data from the survey vignette experiment with administrative data on applicants' actual rank-ordered lists. The vignette experiment is useful because it holds fixed school characteristics other than safety, academic performance, and racial composition. The preference parameters estimated from the vignette experiment bear a causal interpretation: they represent relative preferences among school attributes that may otherwise be correlated with unobserved school characteristics. To align the utility specification with the vignette design, we modify the school attributes as follows. First, we replace continuous school characteristics with binary indicators matching the vignette format: we construct a high-safety dummy equal to one if the student's believed safety ranks above the population median, and a high-academics dummy defined analogously using the average of believed value-added, peer quality, and college enrollment rate. Second, we replace the continuous racial composition variables R_s with the same set of majority-group dummies used in the vignette (majority White/Asian, majority Black, majority Hispanic). We then use the vignette estimates to pin down the relative preference weights on these five attributes, collapsing them into a single quality index $\tilde{X}_s$, while the administrative ranking data identify the remaining free parameters: the trade-off between this quality index and distance, and the application cost distribution.

Model Fit and Choice Gap Decomposition. Figure 6 shows the model fit and the decomposition of the channels driving racial differences in choices along the two dimensions highlighted in Section 2.2: the White and Asian share and the value added in applicants' top three high school choices.[39] The second bar in each panel compares the predicted choice gap (White and Asian choices minus Black and Hispanic choices) along these two margins with the gap observed in the data, reported in the first bar. The model fits both dimensions well, generating a gap of 0.3 in the share of White and Asian students, versus 0.28 in the data, and a gap of 0.1 in school value-added, matching the 0.1 observed in the data.

The remaining bars present the decomposition exercise. We simulate optimal rank-ordered lists under counterfactual scenarios that replace one component of the parameters for Black and Hispanic applicants with the corresponding parameters estimated for White and Asian applicants, then report the simulated choice gap under each scenario. For instance, the third column simulates the choice gap when Black and Hispanic applicants are assigned the admission-probability belief parameters estimated for White and Asian applicants. This procedure decomposes the racial choice gap into portions attributable to differences in beliefs about admission probabilities, application costs, awareness sets, beliefs about school characteristics, preferences over school racial composition R_s alone, and preferences over all school characteristics combined. The final column reports the simulated gap under the assumption that all model parameters are equalized across racial groups.

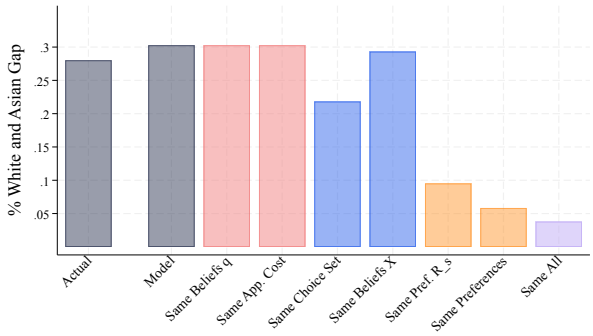
The results show that nearly the entire gap in the racial composition of applicants' top three high school choices can be explained by differences in preferences and awareness of school options. Equalizing awareness of schools across racial groups would close 28% of the gap, while equalizing preferences over school racial composition would close 68% of it. Differences in beliefs about admission probabilities or application costs, which could in principle generate racial differences in strategic reporting, are very small and do not play a role in explaining the choice gap, consistent with the descriptive evidence in the previous section. Similarly, the full choice gap in chosen value-added is explained by differences in information and preferences, and it even slightly reverses sign when all parameters are equalized. Appendix Figure B9 shows the same decomposition exercise for the version of the model that constrains relative preferences to match those observed in the vignette experiment, yielding a broadly similar takeaway. This version of the model assigns a smaller role to preferences, but in both models the majority of the choice gap can be closed by equalizing preferences and information. Importantly, this finding is not mechanical. The decomposition exercise

[39] Appendix Tables A3 and A4 include the full set of model estimates.

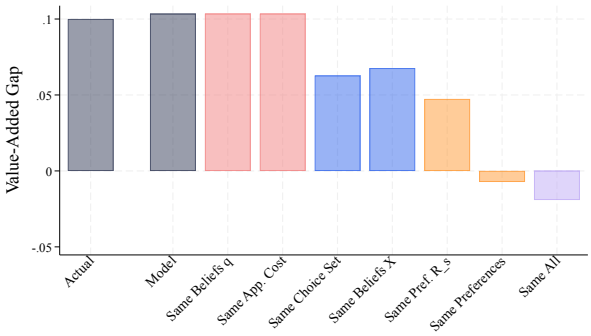
holds each student's residential location fixed, so differences in distance to schools across students could in principle contribute substantially to the choice gap, even if students had identical preferences and information. The fact that equalizing preferences and information nearly closes the gap tells us, once again, that residential segregation on its own does not explain school choice segregation in NYC.

Figure 6: Decomposition of the Choice Gap

(a) % White&Asian



(b) Value-Added



Notes: The first bar plots differences in the characteristics of top 3 school choices of White and Asian applicants relative to Black and Hispanic applicants. The second bar plots the estimated difference based on our model estimates. The remaining bars plot simulated differences in choice characteristics under alternative scenarios in which we replaced one component of the parameters for Black and Hispanic applicants with the corresponding parameters estimated for White and Asian applicants.

4 Exposure to Other-Race Peers Narrows School Choice Gaps

Establishing that racial choice gaps arise from differences in the set of schools families know about and from preferences over peer composition is important for designing policies to narrow these gaps. Prior research has focused primarily on addressing disparities in information, with mixed results for interventions that provide students with information about school performance and quality (Corcoran et al., 2018; Arteaga et al., 2022a; Borger et al., 2024; Corradini, 2024). Changing preferences over school characteristics may be more difficult, and it is less clear how a social planner could intervene along this margin, provided it seeks to do so at all.

We examine an alternative lever that could address both channels simultaneously: integration in earlier grades. Prior work shows that homophily may diminish through repeated exposure to peers of different racial and socio-economic backgrounds (Sacerdote, 2001, 2011;

Rao, 2019; Lowe, 2021; Carrell et al., 2019; Paluck et al., 2019) and that information spreads across school peers (Campos, 2024).[40] In NYC, middle schools are a key setting for information exchange as highlighted in Appendix Figure B7: 75% of respondents report discussing high school applications with other parents in their child's middle school at least once, 26% report doing so more than five times, and more than a quarter identify other parents as one of their most important information sources.[41] At the same time, middle schools expose students and parents to a diverse set of peers, which may shape preferences over the racial composition of future classmates. Earlier grades are also a feasible point of policy intervention: as documented in Section 2, residence-based priorities play a larger role in admissions to earlier grades than in high school, so changes to school admission rules or attendance zones could directly reduce segregation in elementary and middle schools.

The descriptive evidence in Figure 2 established that high school choices vary with the racial composition of students' middle school peers. We now leverage random assignment of students to middle schools to show that this relationship is causal: exposure to other-race peers in earlier grades shapes subsequent education choices by influencing both information and preferences. These results highlight how school diversity can affect factors that are otherwise difficult to change and provide new evidence on the downstream benefits of racially integrated schools.

4.1 Econometric set-up

We estimate the effect of exposure to other-race peers in middle school, captured by the parameter α in the following regression:

$$Y_i = \alpha C_i + X_i'\Gamma + u_i \tag{2}$$

C_i measures student i's exposure to other-race peers in the middle school grade they attend, X_i is a vector of controls (application year, student demographic characteristics including ELL status, gender, poverty status, district of residence, and 4th grade Math and ELA test scores), and u_i is a regression residual. Y_i measures an attribute of student i's high school

[40] Additionally, peers in earlier school grades may affect applicants' perceived admission probabilities by influencing grades and self-assessments of relative ability (Corno et al., 2022).

[41] Middle schools also provide institutional support for the high school application process through information sessions and guidance from school staff. Sixteen percent of families cite middle school sessions as one of their most important information sources, while 26% of respondents overall—and more than 30% of low-income, Black, and Hispanic families—identify middle school staff as a primary source of information about high schools.

choices, such as the average value-added of the schools ranked within their top three choices. 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.

We consider both continuous and discrete measures of the peer exposure variable C_i. The continuous measure is the leave-one-out share of other-race peers in the student's middle school. The discrete measure is an indicator for having a majority (more than 50%) of other-race peers. To construct these variables, we classify students into two race/ethnicity groups: White and Asian students, and Black and Hispanic students. Accordingly, "other-race peers" refers to Black and Hispanic peers for White or Asian students, and to White and Asian peers for Black or Hispanic students.

Estimating equation 2 directly via OLS is likely to yield biased estimates. Applicants may have persistent preferences for school attributes, meaning that students who enroll in more diverse middle schools may also prefer more diverse high schools simply due to underlying, stable tastes. To address this endogeneity concern, we employ an instrumental variables strategy that leverages the random variation in student assignments generated by tie-breaking in the middle school matching algorithm.

The construction of an instrument for the racial composition of middle-school peers adapts the method in Angrist et al. (2022). Specifically, we instrument the share of other-race peers in the middle school of enrollment with the share of other-race peers in the middle school a student is offered, while controlling for the peer composition a student should expect based on their specific application, priorities, and rankings across the middle schools they applied to. This control corresponds to the expected value of the instrument and accounts for systematic differences in potential outcomes between applicants who are offered schools with different racial compositions. Our framework extends Angrist et al. (2022) by recognizing that the racial makeup of the offered school depends not only on a student's own offer but also on the entire set of offers. To address this, we compute the potential racial composition of each school using students' offer distributions, rather than realized offers, when constructing the instrument. The remainder of this section provides a detailed description of this empirical strategy.

To build our offer-based instrument for middle school peers, we begin by computing the probability that each applicant i is offered each middle school s in the market. School offers depend on applicant preferences and priorities - collectively referred to as the applicant type t_i - as well as the tie-breaking variables: a common lottery number used by unscreened

schools and a set of non-lottery tie-breakers (such as test scores) used by screened schools. The assignment probability or propensity score, for each school can be written as:

$$\varphi_s(t_i, \tau_i(\delta_N)) = E[D_i(s) \mid t_i, \tau_i(\delta_N)],$$

where $D_i(s)$ indicates an offer at school s. This probability is a function of the applicant type t_i and indicators for proximity to admission cutoffs for non-lottery programs, denoted $\tau_i(\delta_N)$, which are determined using a data-driven bandwidth δ_N. Because t_i and $\tau_i(\delta_N)$ can be inferred directly from application data and from the known admission rules of each school, the propensity score $\varphi_s(t_i, \tau_i(\delta_N))$ can be readily computed for each student-school pair as described in Abdulkadiroglu et al. (2022).

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 the realization of tiebreakers:

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

where $O_i(j)$ is a dummy equal to 1 if student j is of a different race than student i. This measure incorporates the uncertainty in the assignment of all students in the match, as the expectation is taken with respect to the probability distribution of school offers for all applicants. Potential peer shares generally differ from the realized peer shares observed after enrollment due to randomness in the match process and factors such as imperfect compliance with offers, student attrition, and post-match enrollment of students who did not participate in the initial assignment.

Using the potential peer share at each school and each student's offer probabilities, we can construct instruments for peer exposure. Our instrument for the realized share of other-race peers in the middle school of enrollment is the potential share of other-race students in the middle school offered:

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

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:

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

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 t_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:

$$Y_i = \beta C_i + \kappa_2 \mu_i + \sum_s g_s(R_{is}) + X_i'\Gamma_2 + \epsilon_i \tag{3}$$

$$C_i = \psi Z_i + \kappa_1 \mu_i + \sum_s h_s(R_{is}) + X_i'\Gamma_1 + \nu_i. \tag{4}$$

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}.[42] These functions employ the bandwidth used to define $\tau_i(\delta_N)$.[43] First and second stage models also include a set of baseline covariates, denoted by X_i.[44]

We use the same empirical specification to estimate the effect of enrolling in a middle school where the majority of peers are of a different race, substituting the continuous instrument and control function with their discrete counterparts. Specifically, our instrument for enrolling in a middle school where the majority of peers are of a different race is an indicator for being offered a middle school whose potential other-race peer share exceeds 50%:

$$Z_i^D = \mathbb{I}\{\sum_s D_i(s) c_i^P(s) > 0.5\}.$$

The corresponding control function is obtained by taking the expectation of this instrument:

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

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

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

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

[43] 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.

[44] 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.

In all the regressions, 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 A5 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 Table A6 reports a set of results supporting the validity of the research design. Panel A checks whether differential attrition may lead to selection bias. It shows that the likelihood of observing high school choice outcomes is unrelated to the majority other-race offer and offered other-race share instruments. 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 our main samples, both the discrete instrument and the continuous instrument exhibit good balance across observable characteristics. In all cases, our 2SLS specifications include the baseline covariates listed in the tables as controls.

4.2 Estimates of Peer Effects on School Choices

Table 4 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.[45] 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 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.

[45]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.

Table 4: 2SLS Estimates of Peer Effects on Black & Hispanic Applicants' HS Choices

	% Black (1)	% Hispanic (2)	% White&Asian (3)	Peer Quality (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 2022-2023 and then successively applied to high school for enrollment in 2018-2019 to 2025-2026. 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 quality, and High School Math VA are defined in the notes of Table 1. Robust standard errors in parenthesis.

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.[46] 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

[46] Appendix Table A7 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.

Hispanic students to choose higher-quality 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, suggesting that exposure to White and Asian peers primarily influences students' top-ranked options. Additionally, because the total number of ranked schools is unchanged, these effects reflect substitution rather than the addition of new schools to students' lists.

The effects on Black and Hispanic students' high school choices are substantial, especially when compared with average racial choice gaps conditional on achievement and district of residence. Attending majority-white schools reduces the raw racial choice gap in school value-added by 36% and reduces the portion of the gap unexplained by differences in achievement or residence by roughly two-thirds. Similarly, it closes about one quarter of the raw gap, and about three-quarters of the unexplained gap, in the demographic composition of chosen schools.

Peer diversity also influences the high school choices of White and Asian students, though the effects are smaller than those observed for Black and Hispanic applicants. As shown in Table 5, middle school demographics do not significantly impact the selectivity or quality of high school choices for these students. However, exposure to other-race peers still affects the demographic composition of their preferred schools, though the effects are less than half as large as those observed for Black and Hispanic applicants. For example, attending majority–Black-and-Hispanic middle schools increases the share of Black or Hispanic students in White and Asian applicants' top three choices by only 2.9 percentage points. The effects are somewhat larger when considering all ranked schools, in contrast to the patterns found for Black and Hispanic students, suggesting that White and Asian applicants primarily adjust their lower-ranked choices in response to greater exposure to other-race peers in middle-school.

A comparison of the 2SLS estimates with the OLS estimates in Table A8 shows that OLS estimates of peer effects are not severely biased, particularly for Black and Hispanic students. The key OLS coefficients have the same sign as their 2SLS counterparts, and their magnitudes differ by roughly one-half. Specifically, the 2SLS estimates of exposure to White and Asian peers are less than 40% smaller than the OLS estimates for the racial composition of school choices, and the two sets of estimates are statistically indistinguishable for peer achievement and school value-added.

Finally, Table 6 shows that attending middle schools with more White and Asian peers

Table 5: 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 4. Robust standard errors in parenthesis.

affects not only school choices but also school offers for Black and Hispanic students. Black and Hispanic applicants who attend 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 offers are somewhat smaller than those on choices, they indicate that attending more diverse middle schools does not lead these students to rank schools beyond their reach. However, the increases in high school value-added, peer diversity, and peer achievement are accompanied by a 5–percentage point rise in the likelihood of being unmatched, consistent with the lower match rates observed 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, mirroring the smaller effects seen in their top choices.

Table 6: 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 4. Robust standard errors in parenthesis.

4.3 Effects of Other Middle School Characteristics

Attending a middle school with a high share of White and Asian students exposes students to a bundle of characteristics, as peer composition is correlated with many other school factors. Column 4 of Appendix Table A9 shows that majority-White middle schools have, on average, higher quality, as measured by both causal estimates of school value-added and the average of the seven dimensions included in the NYC School Quality Report. They also employ teachers with roughly one additional year of experience, are 1 percentage point more likely to have a certified school counselor on staff, and have higher peer quality, reflected in average fifth-grade Math test scores that are 0.86 σ higher. Because these schools are in high demand, however, they have a lower average teacher-pupil ratio.

The previous section attributed the effects of attending majority-white middle schools to exposure to other-race peers. In this section, we assess whether these effects could instead reflect other middle school characteristics by estimating 2SLS regressions with multiple endogenous treatments. We find that peer composition remains the factor most strongly associated with students' subsequent educational choices.

To separate the effect of peer diversity from other school factors, we modify equation (3) by adding a term on the right-hand side that captures exposure to other middle school characteristics. The resulting model has multiple endogenous variables and can be estimated using the following 2SLS system of equations:

$$Y_i = \beta_1 C_i + \beta_2 Other_i + \kappa_2 \mu_i + \phi_2 \mu_i^{Other} + \sum_s g_s(R_{is}) + X_i' \Gamma_2 + \epsilon_i \tag{5}$$

$$C_i = \psi_1 Z_i + \rho_1 Z_i^{Other} + \kappa_{11} \mu_i + \phi_{11} \mu_i^{Other} + \sum_s h_s^1(R_{is}) + X_i' \Gamma_{11} + \nu_i \tag{6}$$

$$Other_i = \psi_2 Z_i + \rho_2 Z_i^{Other} + \kappa_{12} \mu_i + \phi_{12} \mu_i^{Other} + \sum_s h_s^2(R_{is}) + X_i' \Gamma_{12} + u_i \tag{7}$$

.

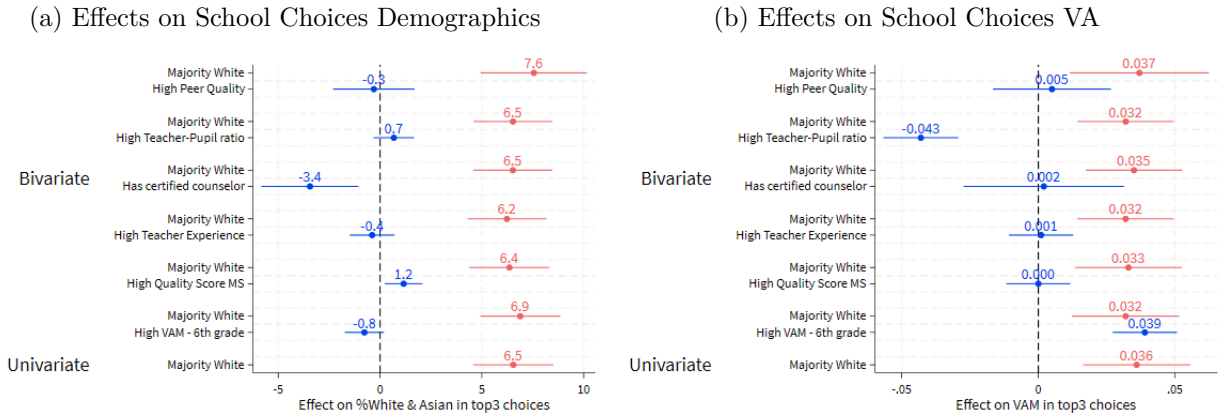
Depending on the model, the term $Other_i$ may represent middle-school quality, teacher experience, school resources, or peer achievement. This characteristic of the enrolled school is instrumented by Z_i^{Other}, the corresponding characteristic of the offered school. The term μ_i^{Other} denotes the expected value of the instrument Z_i^{Other}, which reflects the middle-school characteristic that student i can expect to be offered based on her preferences and priorities at the schools she ranked.

Figure 7 plots 2SLS estimates of β_1 and β_2 in equation (5) on the sample of Black and Hispanic applicants and reports as a reference point the estimate of β in the univariate model of equation (3). The endogenous variables are binary indicators for enrolling in a majority-White and Asian middle school and enrolling in a middle school in the top 25% of the distribution for the other characteristic considered in the model.[47] Each pair of coefficients comes from a separate regression. The dependent variables are the mean share of White and Asian students (Panel a) and the mean school value-added (Panel b) among an applicant's top three high school choices.

Two key findings emerge from Figure 7. First, the estimated effect of middle-school peer race remains stable across specifications and is indistinguishable from the estimate in the univariate model. Second, the share of White and Asian peers is the only middle-school characteristic that increases the share of White and Asian students in the high school choices of Black and Hispanic students, and is the only middle school characteristic, besides value-added, that improves the value-added of high school choices. Although attending a

[47] We define binary indicators using the 75^{th} percentile as the cutoff so they are comparable to the indicator for attending a majority-White and Asian middle school, which captures roughly the top 25% of middle schools by White and Asian share. Appendix Figure B10 reports estimates for models using continuous endogenous variables, rescaled in across-school standard deviations.

Figure 7: Peer Race Effects vs. Effects of Middle School Characteristics



Notes: This figure plots estimates of the coefficients β_1 and β_2 in equation (5) on the mean share of White and Asian students (Panel a) and the mean school value-added (Panel b) among an applicant's top three high school choices. The sample is restricted to Black and Hispanic applicants. Each pair of coefficients in the bivariate model rows comes from a separate regression of choice characteristics on two binary variables indicating enrollment in a majority-White and Asian middle school and in a middle school in the top 25% of the distribution of school quality/teacher experience/school resources/peer achievement, depending on the model. The last row reports as reference the estimate of β in the univariate model of equation (3).

high–value-added middle school similarly raises the quality of students' high school choices, it does not explain the effect of other-race peers, nor it affects the demographic composition of chosen schools in a way that would reduce racial segregation.

The effects of peer race are large, not only relative to the other observable middle schools characteristics in our data, but also compared to the impact of information interventions designed to improve school choices documented in the literature. Appendix Table A10 highlights this contrast, showing that attending a majority-White middle school increases the value added of chosen schools more than direct interventions that inform families about school quality (Panel A). Additionally, its effects on the future peer composition of schools students attend or choose (Panels B and C) exceed those of various information interventions.

4.4 Peer Effects Mechanisms

Our findings suggest that racial gaps in school choices are primarily due to differences in awareness of school options and preferences for schools with demographically similar 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' mid-

dle school applications and offers, we explore whether interactions with peers from different racial backgrounds reduce homophily, shape information, and affect test scores.

Preferences Recent studies find that positive interactions with individuals of different ethnic backgrounds help reduce inter-group or inter-racial prejudice (Rao, 2019; Carrell et al., 2019; Lowe, 2021; Corno et al., 2022; 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 were offered a majority-White and Asian middle school (Z_i^D), controlling for the control function μ_i^D. The estimates are presented in Table 7.[48] To increase precision, we pool respondents who saw precise academic-performance information with those who saw imprecise information. Preference estimates are therefore a weighted average across the two information treatments.

Across racial groups, being offered a majority–White and Asian middle school increases respondents' preferences for majority–White and Asian hypothetical high schools relative to racially balanced ones. The interaction coefficients are large but imprecisely estimated, reaching significance only at the 5 and 10 percent levels. Among White and Asian respondents, this reinforces an existing preference for majority–White and Asian schools; among Black and Hispanic respondents, it offsets an initial distaste. The interaction coefficients are null for the other dimensions of preferences: exposure to other-race peers shifts preferences over the racial composition of peers, but not the perceived importance of school academic performance or safety.

An 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 A11 presents OLS and IV

[48] These are reduced-form effects, based on quasi-random variation in the racial composition of middle school offers (which is random conditional on μ_i^D) rather than in the school of actual enrollment. Because compliance with match offers is high in this setting (∼80%), the reduced-form effects should closely approximate the causal effects of actual peer exposure.

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	3.01***	-0.92	2.49***	-0.15
	(0.08)	(0.58)	(0.12)	(0.37)
Imprecise Info on Academics	0.73***	0.74	1.07***	-0.33
	(0.10)	(0.68)	(0.15)	(0.50)
High-Academics	1.12***	0.28	1.26***	-0.22
	(0.07)	(0.47)	(0.10)	(0.31)
High-Safety	0.70***	0.48	0.73***	-0.02
	(0.06)	(0.42)	(0.09)	(0.27)
Majority Black	-0.26***	0.78	-0.42***	-0.30
	(0.08)	(0.65)	(0.12)	(0.37)
Majority Hispanic	-0.22***	-0.09	-0.32***	-0.60
	(0.08)	(0.57)	(0.12)	(0.37)
Majority White&Asian	-0.21**	1.45**	0.18	0.62*
	(0.09)	(0.67)	(0.12)	(0.37)
N Respondents	1,709		2,157	

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.

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.[49] Results show mostly null 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 3.2.2. Table 8 presents OLS and IV estimates of the effect of attending a majority-White and Asian middle school on the likelihood that a school is in the respondent's awareness set. The dependent variables and school types are consistent with those in Panel (a) of Figure 4.

[49]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.

Table 8: OLS and 2SLS Estimates of Peer Effects on Awareness 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)
(Majority White&Asian MS) × (Black&Hispanic)	0.06***	0.16	0.08***	-0.05	0.08***	0.20	0.06**	0.22*	0.09***	0.26**
	(0.02)	(0.11)	(0.03)	(0.17)	(0.02)	(0.13)	(0.02)	(0.13)	(0.02)	(0.13)
Majority White&Asian MS	-0.02	-0.08	0.01	0.07	-0.02	-0.08	-0.01	-0.14	-0.02	-0.11
	(0.01)	(0.07)	(0.02)	(0.10)	(0.02)	(0.08)	(0.02)	(0.09)	(0.02)	(0.08)
Black&Hispanic	-0.00	-0.00	-0.04*	-0.04*	-0.04**	-0.04**	-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). 2SLS regressions include the set of controls in equation 3. All regressions also control for school fixed effects, 4th grade test scores and commuting time to the school as in the specification of Figure 4. Clustered standard errors at the student-level in parenthesis.

The OLS estimates show that Black and Hispanic applicants attending majority–White and Asian middle schools are aware of a different set of high school options than other minority applicants with the same test scores living in the same neighborhood: they are more likely to recognize schools that are more popular, higher performing, higher quality, and that enroll larger shares of White and Asian students. In contrast, differences in school knowledge across middle schools are much smaller for White and Asian students. These OLS estimates may partly reflect selection of families into different middle schools, but the IV estimates, though noisier, reinforce the causal interpretation: attending a majority–White and Asian middle school 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 26 percentage points, respectively. These results suggest that awareness of school options not only contributes to racial gaps in school choices but also accounts for a meaningful share of peer effects on choice.

Test Scores and Beliefs of Admission Chances Lastly, we examine the effects of attending schools where most peers are of a different race on middle school achievement. As shown in Figure 2, respondents with higher test scores are more likely to apply to higher value-added high schools, as well as to high schools enrolling more White and Asian students, which are also more likely to screen applicants based on academic performance. Higher grades may increase students' chances of admission to selective programs and thus help explain their

greater propensity to apply.

We estimate the impact of middle-school diversity on student achievement using the same 2SLS framework applied 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 in both middle and high school.[50]

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 4. Robust standard errors in parenthesis.

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 0.14σ by 8^{th} 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.

[50]There is minimal concern that differential test-taking drives these results. Appendix Table A12 shows that middle-school peers have no significant effect on the likelihood of taking standardized tests.

Attending a majority White and Asian middle school increases Black and Hispanic students' math Regents scores by 0.19σ.

However, this increase in 7^{th}-grade test scores is not large enough to explain why middle-school peers influence Black and Hispanic students' application behavior. While Figure 2 and Appendix Figure B3 show that higher-achieving middle-school students are more likely to choose higher–value-added high schools and schools that enroll higher-achieving and Whiter peers, the estimated test-score gains are too small to account for the observed peer effects on high school choices.

The slope relating applicants' baseline achievement to the peer quality of their top high school choices implies that a 0.086σ increase in 7th-grade test scores, the estimated effect of attending majority-White middle schools, would raise the baseline peer math achievement in Black and Hispanic applicants' top three choices by only 0.014σ. This is just 11% of the corresponding middle-school peer effect on choices. Similarly, this test-score gain corresponds to only a 0.44 percentage point increase in the share of White and Asian students and a 0.005σ increase in the value-added of Black and Hispanic applicants' top three choices, representing just 7% and 14% of the estimated peer effects on choices, respectively.

These limited impacts of test-score gains on high-school choices are consistent with the idea that small test-score improvements do not meaningfully change admission chances or beliefs about admission to popular high-school programs. Appendix Table A13 presents IV estimates of peer effects on achievement and admission probabilities, as well as applicants' beliefs about these outcomes. Within this smaller survey sample, we again find that White and Asian school peers raise test scores and citywide academic rankings for Black and Hispanic students. Yet, these improvements are not large enough to enhance their relative standing in competitive programs or to increase actual admission probabilities for high-demand or "dream school" options. Likewise, although higher test scores lead students to hold more positive beliefs about their academic standing citywide, they do not change beliefs about admission chances to high-demand programs, consistent with the minimal effects on actual admission probabilities.

Summary In summary, we find that middle school peer composition influences students' high school choices primarily through changes in information and preferences over peer composition, rather than through test-score gains. Although attending majority–White and Asian middle schools significantly raises Black and Hispanic students' middle school test scores, these gains are not large enough to meaningfully affect admission probabilities or

beliefs about admission chances. Nor are they large enough to shift school preferences meaningfully: the estimated test-score gains map into only a small fraction of the observed peer effects on choices.

For Black and Hispanic students, attending middle schools enrolling more White and Asian peers shapes their future school choices by both (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. For White and Asian students, exposure to more Black and Hispanic peers also affects high school choices, though to a smaller extent, primarily by reducing their strong preference for majority–White and Asian schools and increasing their willingness to consider racially balanced options. We find no evidence that middle school demographics affect the set of schools White and Asian families are aware of, consistent with these families relying more on other sources of information or different social networks when gathering information about high schools.

5 Conclusions

Large urban school districts may adopt centralized school choice systems in a hope to equalize access to high-quality schools by weakening the link between residential segregation and schooling. In principle, these systems should also promote integration across income and race by allowing students to apply beyond the boundaries of their potentially segregated neighborhoods. This paper studies school choice in New York City, a setting where students can apply to any high school in the city, the student population is highly diverse, and public transportation makes most schools geographically accessible. Yet, despite these features, New York City Public Schools remains one of the most segregated school systems in the United States.

We document that students' application choices play an important role in sustaining racial segregation across schools and in amplifying achievement gaps. Even among otherwise similar students, living in the same neighborhoods and with comparable prior achievement, there are large racial differences in high school choices. On average, Black and Hispanic applicants apply to schools with lower academic quality and with smaller shares of White and Asian students. These patterns imply that the choices families make within a centralized assignment system can substantially shape both segregation and access to educational opportunities.

These findings also have implications for supply-side responses to school choice. A com-

mon argument for centralized choice systems is that they put competitive pressure on schools to improve quality. This mechanism, however, relies on the assumption that families' choices respond to school quality. We show that Black and Hispanic families are systematically less likely to apply to higher-value-added schools, weakening the demand-side incentives that schools serving these families face to invest in quality-improving inputs. Addressing the drivers of racial choice gaps is therefore important not only for ensuring equitable access in the short run, but also for strengthening the demand-side incentives that centralized choice systems are intended to generate.

Using administrative data combined with novel survey evidence, we provide new evidence on the sources of these choice gaps. We show that racial differences in school choices arise from two main mechanisms: preferences over peer composition and disparities in information about school options. White and Asian applicants are less likely to apply to schools where they would be part of the racial minority, contributing to patterns of segregation across schools. At the same time, Black and Hispanic students are less likely to be aware of high-quality schools, primarily because these schools enroll larger shares of White and Asian students and may therefore fall outside their social networks or information environments. We also document substantial information frictions in the form of inaccurate beliefs about school attributes and admission probabilities to competitive programs. However, these biases are often not systematically different across racial groups, suggesting that interventions focused solely on correcting biased beliefs may do little to reduce racial disparities in school choices. Instead, increasing awareness of high-quality school options appears to be a first-order margin for improving access.

We further provide causal evidence that exposure to diverse peers earlier in students' educational trajectories can help reduce these disparities. Students who attend middle schools where most peers are of a different race subsequently make different high school choices, becoming more willing to apply to schools with more diverse student bodies. These effects are particularly pronounced for Black and Hispanic students exposed to more White and Asian peers, who subsequently choose higher-quality high schools. Our results indicate that social interactions with better-informed or more diverse peers expand the set of schools students consider and attenuate strong preferences for same-race peer environments.

These findings suggest that efforts to promote integration earlier in students' schooling may generate downstream effects on later education choices. While policies aimed at reducing segregation are often motivated by equity or social cohesion, our results highlight an additional benefit: exposure to diverse peers shapes the information sets and prefer-

ences that guide future decisions. In this sense, integration in earlier grades may act as a lever that fosters more integrated schooling in later stages through changes in students' own choices. More broadly, our results underscore the role of social interactions in shaping the frontier of opportunities that young people consider when making consequential decisions, with implications that extend beyond high school choice.

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

Abdulkadiroğlu, A., Pathak, P. A., Schellenberg, J., and Walters, C. R. (2020). Do Parents Value School Effectiveness? *American Economic Review*, 110(5):1502–1539.

Ainsworth, R., Dehejia, R., Pop-Eleches, C., and Urquiola, M. (2023). Why do households leave school value added on the table? the roles of information and preferences. *American Economic Review*, 113(4):1049–82.

Allende, C., Gallego, F., and Neilson, C. A. (2019). Approximating the equilibrium effects of informed school choice. Working Paper.

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. A., and Walters, C. (2024). Credible school value-added with undersubscribed school lotteries. *The Review of Economics and Statistics*, 106(1):1–19.

Anstreicher, G., Fletcher, J., and Thompson, O. (2022). The long run impacts of court-ordered desegregation. Working Paper 29926, National Bureau of Economic Research.

Arteaga, F., Elacqua, G., Krussig, T., Méndez, C., and Neilson, C. A. (2022a). Can information on school attributes and placement probabilities direct search and choice? evidence from choice platforms in ecuador and peru. Idb working paper, Inter-American Development Bank.

Arteaga, F., Kapor, A. J., Neilson, C. A., and Zimmerman, S. D. (2022b). Smart matching platforms and heterogeneous beliefs in centralized school choice. *The Quarterly Journal of Economics*, 137(3):1791–1848.

Barrios-Fernández, A., Neilson, C., and Zimmerman, S. (2024). Elite universities and the intergenerational transmission of human and social capital. Technical report, IZA - Institute of Labor Economics.

Baum-Snow, N. and Lutz, B. F. (2011). School desegregation, school choice, and changes in residential location patterns by race. *American Economic Review*, 101(7):3019–46.

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.

Billingham, C. M. and Hunt, M. O. (2016). School racial composition and parental choice: New evidence on the preferences of white parents in the united states. *Sociology of Education*, 89(2):99–117.

Billings, S. B., Deming, D. J., and Rockoff, J. (2014). School Segregation, Educational Attainment, and Crime: Evidence from the End of Busing in Charlotte-Mecklenburg*. *The Quarterly Journal of Economics*, 129(1):435–476.

Bjerre-Nielsen, A. and Gandil, M. H. (2024). Attendance boundary policies and the limits to combating school segregation. *American Economic Journal: Economic Policy*, 16(1):190–227.

Black, S. E., Cortes, K. E., and Lincove, J. A. (2015). Academic undermatching of high-achieving minority students: Evidence from race-neutral and holistic admissions policies. *American Economic Review*, 105(5):604–10.

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.

Borger, M., Elacqua, G., Jacas, I., Neilson, C. A., and Westh Olsen, A. S. (2024). Report cards: Parental preferences, information and school choice in haiti. *Economics of Education Review*. October.

Böhlmark, A. and Willén, A. (2020). Tipping and the effects of segregation. *American Economic Journal: Applied Economics*, 12(1):318–47.

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

Campbell, S., Macmillan, L., Murphy, R., and Wyness, G. (2022). Matching in the dark? inequalities in student to degree match. *Journal of Labor Economics*, 40(4):807–850.

Campos, C. (2024). Social interactions, information, and preferences for schools: Experimental evidence from los angeles. Working Paper 33010, National Bureau of Economic Research.

Card, D., Mas, A., and Rothstein, J. (2008). Tipping and the dynamics of segregation*. *The Quarterly Journal of Economics*, 123(1):177–218.

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.

Cattan, S., Salvanes, K. G., and Tominey, E. (2025). First-generation elite: The role of school social networks. *American Economic Review*, 115(12):4369–4403.

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. 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.

Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., Gong, S., Gonzalez, F., Grondin, A., Jacob, M., Johnston, D., Koenen, M., Laguna-Muggenburg, E., Mudekereza, F., Rutter, T., Thor, N., Townsend, W., Zhang, R., Bailey, M., Barberá, P., Bhole, M., and Wernerfelt, N. (2022). Social capital I: Measurement and associations with economic mobility. *Nature*, 608(7921):108–121.

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

Cohodes, S. R., Corcoran, S. P., Jennings, J. L., and Sattin-Bajaj, C. (2025). When do informational interventions work? experimental evidence from new york city high school choice. *Educational Evaluation and Policy Analysis*, 47(1):208–236.

Corcoran, S. P., Jennings, J. L., Cohodes, S. R., and Sattin-Bajaj, C. (2018). Leveling the playing field for high school choice: Results from a field experiment of informational interventions. Technical report, National Bureau of Economic Research.

Corno, L., La Ferrara, E., and Burns, J. (2022). Interaction, stereotypes, and performance: Evidence from south africa. *American Economic Review*, 112(12):3848–75.

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

Damon Clark, Paco Martorell, and Matt Wiswall (2026). Why do parents care so much about peers? social wellbeing, perceived peer effects and school choices. Working paper.

Dillon, E. W. and Smith, J. A. (2017). Determinants of the match between student ability and college quality. *Journal of Labor Economics*, 35(1):45–66.

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: America's segregated schools 65 years after brown. page 45.

Guryan, J. (2004). Desegregation and black dropout rates. *American Economic Review*, 94(4):919–943.

Haderlein, S. A. K. (2022). How do parents evaluate and select schools? evidence from a survey experiment. *American Educational Research Journal*, 59(2):381–414.

Hahm, D. W. and Park, M. (2022). A dynamic framework of school choice: Effects of middle schools on high school choice. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 292–293.

Hailey, C. A. (2022). Racial Preferences for Schools: Evidence from an Experiment with White, Black, Latinx, and Asian Parents and Students. *Sociology of Education*, 95(2):110–132. _eprint: https://doi.org/10.1177/00380407211065179.

Hastings, J. S., Kane, T. J., and Staiger, D. O. (2008). Heterogeneous preferences and the efficacy of public school choice. Working paper.

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. M. and Avery, C. (2012). The missing one-offs: The hidden supply of high-achieving, low income students. Working Paper 18586, National Bureau of Economic Research.

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. Working paper.

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. (2026). Distance to schools and equal access in school choice systems. Working Paper.

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.

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.

Reber, S. J. (2010). School desegregation and educational attainment for blacks. *The Journal of Human Resources*, 45(4):893–914.

Rothstein, J. M. (2006). Good principals or good peers? parental valuation of school characteristics, tiebout equilibrium, and the incentive effects of competition among jurisdictions. *American Economic Review*, 96(4):1333–1350.

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.

Schachner, J. N. (2021). Racial stratification and school segregation in the suburbs: Evidence from los angeles county. *Social Forces*, 101(1):309–340.

Setren, E. (2024). Busing to opportunity? the impacts of the metco voluntary school desegregation program on urban students of color. Working Paper 32864, National Bureau of Economic Research.

Tedin, K. L. and Weiher, G. R. (2004). Racial/ethnic diversity and academic quality as components of school choice. *The Journal of Politics*, 66(4):1109–1133.

Online Appendix for "Choosing Segregation? The Determinants of Racial Gaps in School Choice and the Integrating Effects of Peer Exposure"

A. Appendix Tables

B. Appendix Figures

C. Survey Appendix

A Appendix Tables

Table A1: Summary Statistics

	2023 Sample			2025 Sample		
	All	Black&Hispanic	White&Asian	All	Black&Hispanic	White&Asian
	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Demographics and Baseline Scores						
Black	0.14	0.28	0.00	0.15	0.32	0.00
Hispanic	0.33	0.66	0.00	0.32	0.65	0.01
White	0.21	0.00	0.42	0.28	0.00	0.54
Asian	0.29	0.00	0.58	0.24	0.00	0.46
FRPL	0.64	0.75	0.53	0.60	0.75	0.47
Female	0.52	0.52	0.51	0.48	0.49	0.47
Ell	0.06	0.08	0.04	0.10	0.15	0.06
4th gr. Math	0.51	0.15	0.87	0.44	-0.03	0.82
4th gr. Ela	0.46	0.18	0.74	0.37	-0.01	0.67
Panel B: Middle School of Enrollment						
% Black&Hispanic peer in MS	54	71	37	52	71	38
Enrolled in MS with >50% white&Asian	0.47	0.21	0.73	0.49	0.19	0.72
Panel C: High School Choices						
Number of HS choices	8.9	8.9	8.9	10.1	9.6	10.6
% Black&Hispanic in top 3 choices	54	65	43	55	67	45
Mean baseline peer Math in top 3 choices	0.34	0.19	0.50	0.31	0.13	0.47
Mean popularity in top 3 choices	5.91	4.43	7.39	5.15	3.45	6.69
Lists a screened program among top 3 choices	0.78	0.68	0.88	0.53	0.41	0.64
Mean Math VA in top 3 choices	0.17	0.11	0.22	0.08	0.06	0.10
Panel D: High School Offers						
Assigned to 1st choice	0.42	0.50	0.34	0.53	0.62	0.44
Assigned to top 3 choices	0.72	0.79	0.65	0.79	0.86	0.72
N	3,628	1,817	1,811	2,510	1,194	1,316

Notes: This table reports descriptive statistics for the sample of respondents to the 2023 and 2025 surveys. The variables reported are the same as in Table 1.

Table A2: Different School Characteristics for the Vignette Experiment

School characteristic	Description	Percentage			
		Asian	Black	Hispanic	White
Demographics	Racially-balanced	15%	29%	38%	16%
	Majority Black	7%	68%	16%	8%
	Majority Hispanic	5%	13%	73%	7%
	Majority White and Asian	17%	15%	21%	45%
		Low		High	
Safety	Percentage of students who feel safe at school	77%		93%	
Treatment 1: Precise information about school academic performance					
		Low		High	
Academics	Percentage of students who graduate in 4 years	75%		93%	
	Percentage of students who enroll in College/career programs	51%		79%	
Treatment 2: Imprecise information about school academic performance					
Academics	Percentage of students who earned enough credits in ninth grade to be on track for graduation	83%			

Notes: This table reports the characteristics of the school cards presented to respondents in the vignette experiments (questions Q17 and Q18).

Table A3: Model Estimates - Information

Panel A: Beliefs - School Characteristics	Peer Quality			VAM			College Enr.			Safety		
	Black	Hispanic	White/Asian	Black	Hispanic	White/Asian	Black	Hispanic	White/Asian	Black	Hispanic	White/Asian
Constant	1.553	2.149	1.915	3.482	2.060	2.418	2.588	2.029	1.697	1.752	1.866	1.148
	(1.376)	(1.127)	(0.597)	(0.910)	(0.812)	(0.500)	(0.892)	(0.752)	(0.443)	(1.035)	(0.828)	(0.526)
White/Asian share	0.272	0.301	0.318	-0.045	0.115	-0.007	-0.274	-0.167	0.167	0.232	0.278	1.242
	(0.672)	(0.600)	(0.307)	(0.436)	(0.404)	(0.241)	(0.450)	(0.407)	(0.231)	(0.567)	(0.512)	(0.300)
Hispanic share	0.119	-0.233	0.194	-0.501	0.064	0.002	-0.403	-0.519	-0.343	0.016	0.267	0.452
	(0.595)	(0.520)	(0.284)	(0.436)	(0.405)	(0.244)	(0.447)	(0.408)	(0.244)	(0.565)	(0.511)	(0.311)
Peer quality	0.382	0.371	0.784	0.161	0.261	0.589	0.238	0.284	0.356	0.090	0.271	0.433
	(0.360)	(0.295)	(0.148)	(0.225)	(0.180)	(0.107)	(0.235)	(0.183)	(0.105)	(0.329)	(0.258)	(0.146)
VAM	0.375	-0.285	0.203	0.566	0.034	0.017	0.471	0.038	0.300	0.337	0.339	0.055
	(0.486)	(0.378)	(0.217)	(0.286)	(0.241)	(0.151)	(0.323)	(0.266)	(0.161)	(0.465)	(0.373)	(0.224)
College enr.	0.002	0.002	0.002	-0.004	-0.006	-0.001	0.004	-0.002	0.005	-0.003	-0.006	-0.001
	(0.008)	(0.007)	(0.004)	(0.006)	(0.005)	(0.003)	(0.007)	(0.005)	(0.003)	(0.008)	(0.007)	(0.004)
Safety	1.072	0.527	0.168	-0.024	1.299	0.473	0.133	1.190	0.759	1.328	1.054	1.034
	(1.433)	(1.172)	(0.620)	(1.013)	(0.912)	(0.551)	(0.981)	(0.806)	(0.466)	(1.118)	(0.848)	(0.550)

Panel B: Awareness Set	Black	Hispanic	White/Asian
Constant	1.126	0.775	0.262
	(0.362)	(0.295)	(0.144)
Distance	-0.012	-0.014	-0.013
	(0.001)	(0.001)	(0.001)
White/Asian share	-0.039	0.622	1.392
	(0.178)	(0.142)	(0.072)
Hispanic share	-0.890	0.307	0.439
	(0.152)	(0.128)	(0.064)
Peer quality	0.409	0.421	0.393
	(0.108)	(0.085)	(0.045)
VAM	0.119	0.255	0.036
	(0.106)	(0.085)	(0.044)
College enr.	-0.004	-0.002	0.007
	(0.003)	(0.002)	(0.001)
Safety	-0.202	-0.773	-1.265
	(0.405)	(0.325)	(0.161)

Panel C: Beliefs - p(admission)	Black/Hispanic	White/Asian
Constant	0.515	0.389
	(0.041)	(0.026)
Probability	0.449	0.426
	(0.164)	(0.090)
Probability sq.	-0.301	-0.128
	(0.165)	(0.093)
High demand	0.027	0.093
	(0.044)	(0.029)
Under-subscribed	-0.037	-0.136
	(0.067)	(0.043)

Notes: This table reports parameter estimates from the school choice model, capturing how students form beliefs about school attributes, construct their awareness sets, and assess their chances of admission.

Table A4: Model Estimates - Preferences

Panel A: Main Specification

	Black	Hispanic	White/Asian
Distance	-1.000	-1.000	-1.000
	(0.000)	(0.000)	(0.000)
White/Asian share pctile	0.297	0.540	0.451
	(0.016)	(0.022)	(0.021)
Hispanic share pctile	-0.156	0.296	-0.401
	(0.030)	(0.019)	(0.013)
Peer quality pctile	-0.079	0.506	-0.702
	(0.090)	(0.049)	(0.040)
VAM pctile	0.640	-0.866	0.651
	(0.137)	(0.063)	(0.049)
College enr. pctile	0.487	-0.010	0.521
	(0.137)	(0.073)	(0.050)
Safety pctile	-0.349	0.720	0.138
	(0.102)	(0.065)	(0.034)
$\bar{c}$	6.330	6.856	2.222
	(0.617)	(0.375)	(0.129)
σ_c	9.703	9.248	4.903
	(0.680)	(0.411)	(0.140)

Panel B: Using Vignettes Experiment Relative Preferences

	Black	Hispanic	White/Asian
Distance	-1.000	-1.000	-1.000
	(0.000)	(0.000)	(0.000)
Majority White and Asian School	-2.394	-1.583	4.890
	(0.183)	(0.098)	(0.109)
Majority Hispanic School	-2.993	-1.979	-4.890
	(0.229)	(0.122)	(0.109)
Majority Black School	-5.586	-3.695	-9.306
	(0.428)	(0.228)	(0.207)
High Academics	21.546	14.251	22.398
	(1.650)	(0.880)	(0.498)
High Safety	9.177	6.070	10.253
	(0.703)	(0.375)	(0.228)
$\bar{c}$	6.389	5.215	1.769
	(0.600)	(0.302)	(0.119)
σ_c	11.063	7.928	4.672
	(0.668)	(0.339)	(0.121)

Notes: This table reports preference and application cost parameter estimates from the school choice model, normalized by the coefficient on commuting distance so that all parameters are expressed in units of willingness to travel (minutes of commuting).

Table A5: Sample Construction

	N
General education 6th grade applicants	336,083
Enrolls in 6th grade in NYC public schools, with demographics	303,299
Applies to 9th grade NYC public schools	268,446
Has non-degenerate risk of middle school assignment	145,101
Are offered a 6th grade seat	127,104
Have baseline (4th grade) test scores	118,078

Notes: This table illustrates the construction of the experimental sample. The sample starts from students applying to enroll in 6th grade general education seats through the middle school match in the school years from 2015-2016 to 2022-2023. We restrict this sample to students who later enroll in NYC public middle schools, and later apply to the 9th grade NYC public high school match. We exclude students with missing demographic information (race, ell status, gender, poverty status, district of residence), who did not have risk of assignment in the middle school match, who were not matched in the first round, and who did not take 4th grade state test scores.

Table A6: Attrition and Covariate Balance

	Offered majority white&Asian MS			Offered % white&Asian (10 p.p.)		
Dependent variable	All (1)	White/Asian (2)	Black/Hispanic (3)	All (4)	White/Asian (5)	Black/Hispanic (6)
Panel A: Attrition						
Has 9th grade application	0.004	-0.008	0.016	0.003*	0.002	0.003
	(0.008)	(0.012)	(0.011)	(0.002)	(0.003)	(0.002)
mean	0.89	0.90	0.88	0.89	0.90	0.88
N	132,592	50,686	79,391	132,592	50,686	79,391
Panel B: Covariates balance						
Black	0.006		0.007	-0.001		-0.001
	(0.010)		(0.016)	(0.002)		(0.003)
mean	0.19		0.31	0.19		0.31
Hispanic	0.016		-0.007	0.000		0.001
	(0.012)		(0.016)	(0.002)		(0.003)
mean	0.41		0.69	0.41		0.69
White&Asian	-0.027**			0.000		
	(0.013)			(0.002)		
mean	0.39			0.39		
Female	-0.001	-0.021	0.013	0.003	-0.001	0.005
	(0.013)	(0.019)	(0.018)	(0.003)	(0.005)	(0.004)
mean	0.51	0.50	0.53	0.51	0.50	0.53
English Language Learner	-0.007	-0.007	-0.009	-0.001	-0.003	0.000
	(0.006)	(0.008)	(0.008)	(0.002)	(0.002)	(0.002)
mean	0.09	0.07	0.10	0.09	0.07	0.10
Low-income	0.004	-0.004	0.008	-0.003	-0.004	-0.002
	(0.012)	(0.019)	(0.016)	(0.002)	(0.004)	(0.003)
mean	0.72	0.58	0.82	0.72	0.58	0.82
Baseline English	0.035*	0.043	0.036	0.007	0.013*	0.001
	(0.020)	(0.029)	(0.028)	(0.004)	(0.007)	(0.006)
mean	0.19	0.54	-0.05	0.19	0.54	-0.05
Baseline Math	0.008	0.040	-0.008	-0.001	0.008	-0.007
	(0.020)	(0.027)	(0.028)	(0.004)	(0.007)	(0.005)
mean	0.19	0.66	-0.12	0.19	0.66	-0.12
N	118,078	45,649	70,219	118,078	45,649	70,219

Notes: This table reports coefficients from regressions of the variables listed at left on an indicator for being offered a seat at a majority White and Asian middle school (columns 1–3) or on the continuous instrument for peer exposure measuring the potential share of White and Asian peers in the offered middle school (columns 4–6). In columns 4–6, coefficients are rescaled to express the effect of a 10 percentage point increase in the share of White and Asian peers. Column headings indicate the estimation sample. The sample is restricted to applicants with non-degenerate risk of middle school assignment.

Table A7: 2SLS Estimates of Peer Effects on Top 3 High School Choices - Multiple Treatment

Model:		% Black (1)	% Hispanic (2)	% White&Asian (3)	Peer Math (4)	Popularity (5)	Screened (6)	Math VA (7)	Length of rol (8)
				Panel A: Black applicants					
Univariate	Share white & Asian (10pp)	-1.553***	0.026	1.528***	0.019**	0.145**	0.009	0.006*	-0.014
		(0.351)	(0.231)	(0.354)	(0.007)	(0.053)	(0.009)	(0.003)	(0.073)
	Share white & Asian (10pp)	-1.648***	0.143	1.509***	0.018**	0.147**	0.006	0.006*	-0.008
Bivariate		(0.353)	(0.224)	(0.357)	(0.007)	(0.053)	(0.009)	(0.003)	(0.073)
	Share Hispanic (10pp)	-1.354***	1.675***	-0.255	-0.009	0.037	-0.029**	-0.006	0.087
		(0.409)	(0.309)	(0.390)	(0.009)	(0.077)	(0.011)	(0.005)	(0.094)
	Mean	33.75	37.19	26.96	0.02	3.14	0.62	0.37	8.43
	N	21,984	21,984	21,984	21,984	21,888	21,984	21,980	21,984
				Panel B: Hispanic applicants					
Univariate	Share white & Asian (10pp)	-0.574***	-0.650**	1.199***	0.023***	0.103**	0.011	0.009***	0.014
		(0.150)	(0.209)	(0.230)	(0.005)	(0.045)	(0.007)	(0.003)	(0.053)
	Share white & Asian (10pp)	0.155	-1.109***	0.921***	0.014**	-0.031	0.023***	0.002	0.141**
Bivariate		(0.118)	(0.172)	(0.187)	(0.004)	(0.038)	(0.006)	(0.002)	(0.058)
	Share Black (10pp)	1.984***	-2.297***	0.253	-0.006	-0.282***	0.058***	-0.022***	0.474***
		(0.223)	(0.245)	(0.261)	(0.007)	(0.068)	(0.009)	(0.003)	(0.088)
	Mean	18.46	49.85	29.95	0.07	3.70	0.59	0.48	8.12
	N	48,353	48,353	48,353	48,353	48,062	48,353	48,346	48,353
				Panel C: White & Asian applicants					
	Share Black (10pp)	0.724**	0.049	-0.864**	-0.013	0.048	0.010	0.005	-0.279**
Bivariate		(0.275)	(0.268)	(0.371)	(0.010)	(0.079)	(0.010)	(0.004)	(0.095)
	Share Hispanic (10pp)	-0.242	0.705**	-0.487*	-0.014**	0.036	0.002	-0.007**	0.036
		(0.154)	(0.217)	(0.261)	(0.007)	(0.058)	(0.007)	(0.003)	(0.069)
	Mean	12.82	27.67	57.40	0.45	5.16	0.81	1.13	7.51
	N	45,739	45,739	45,739	45,739	45,623	45,737	45,716	45,739

Notes: This table reports 2SLS estimates of middle school demographic composition effects on the characteristics of high school choices for models with two endogenous regressors (bivariate), one for each race share different from own, or for models with one endogenous regressor (univariate). Panel A focuses on Black applicants, panel B on Hispanic applicants, while panel C on White and Asian applicants. The sample and controls are defined in the notes of Table 4. Robust standard errors in parenthesis.

Table A8: OLS Estimates of Peer Effects on Students' Top 3 High School Choices

	% Black (1)	% Hispanic (2)	% White&Asian (3)	Peer Math (4)	Popularity (5)	Screened (6)	Math VA (7)		Length of rol (8)
Panel A: Black & Hispanic Applicants									
Majority white & Asian MS	-6.355***	-4.519***	10.899***	0.141***	0.502***	-0.024***	0.034***		-0.300***
	(0.161)	(0.168)	(0.217)	(0.004)	(0.038)	(0.006)	(0.002)		(0.050)
Share white & Asian (10 pp)	-1.728***	-0.995***	2.688***	0.034***	0.112***	0.006***	0.007***		-0.131***
	(0.031)	(0.033)	(0.039)	(0.001)	(0.007)	(0.001)	(0.000)		(0.009)
Mean	23.24	45.89	29.02	0.05	3.53	0.60	0.08		8.22
N	70,337	70,337	70,337	70,337	69,939	70,337	70,324		70,337
Panel B: White & Asian Applicants									
Majority Black & Hispanic MS	4.114***	4.067***	-8.357***	-0.096***	-0.007	0.048***	-0.035***		0.557***
	(0.120)	(0.153)	(0.203)	(0.005)	(0.037)	(0.005)	(0.002)		(0.047)
Share Black & Hispanic (10pp)	1.354***	1.389***	-2.772***	-0.037***	-0.021**	0.008***	-0.013***		0.181***
	(0.029)	(0.034)	(0.046)	(0.001)	(0.008)	(0.001)	(0.000)		(0.011)
Mean	12.82	27.67	57.40	0.45	5.16	0.81	0.19		7.51
N	45,739	45,739	45,739	45,739	45,623	45,737	45,716		45,739
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	

Notes: This table reports OLS estimates of middle school demographic composition effects on the characteristics of top 3 high school choices. Panel A restricts the sample to Black and Hispanic applicants, while panel B to White and Asian applicants. The sample and controls are as defined in the notes of Table 4. Robust standard errors in parenthesis.

Table A9: Summary Statistics of Middle Schools

	All MS		Majority-minority MS	Majority-white MS
	Mean	SD	Mean	Difference
	(1)	(2)	(3)	(4)
Peer Math	-0.263	0.661	-0.433	0.855
				[0]
Teacher-pupil ratio	0.079	0.019	0.082	-0.016
				[0]
Has certified counselor	0.476	0.500	0.476	0.011
				[0.116]
Teachers' years of experience	9.624	3.009	9.443	0.800
				[0]
Quality score SD	0.002	0.834	-0.044	0.215
				[0]
Value-Added (6th grade test scores)	-0.025	0.210	-0.040	0.078
				[0]
N school-year	3521		2801	
N school per year (mean)	587		467	

Notes: This table presents summary statistics of middle school characteristics. Columns 1 and 2 report the mean and variance of school characteristics across all school years. Column 3 shows the mean for majority-Black and Hispanic schools, while column 4 reports the difference in characteristics between majority-Black and Hispanic schools and majority-White and Asian schools, with p-values in brackets.

Table A10: Comparison of Peer Diversity Effects with Treatment Effects in the Literature

Paper (1)	**Treatment** (2)	**Outcome** (3)	**Treatment effect** (4)	**Majority White MS effect** (5)
Panel A: Effects on Value Added of schools of choice or assignment				
Allende et al. (2019)	Information on returns to school quality	VA of school chosen	0.0274 (0.0273)	0.036*** (0.010)
Patrick Agte et al. (2024)	Information on neighborhood VA and prices Information on school VA and prices	VA of 1st choice	0.007 (0.017) 0.001 (0.017)	0.030** (0.015)
Patrick Agte et al. (2024)	Information on neighborhood VA and prices Information on school VA and prices	VA of enrolled school	0.001 (0.019) -0.013 (0.019)	0.034** (0.014)
Arteaga et al. (2022b)	Warning about nonplacement risk	VA of enrolled school	0.022 (0.011)	0.034** (0.014)
Ainsworth et al. (2023)	Information on school VA	VA of assigned school (school SD)	0.048* (0.025)	0.115** (0.046)
Campos (2024)	Information on school VA Information on school peer quality	VA percentile of 1st choice	0.205 (3.416) - 7.816** (3.717) 0.871 (3.410) - 6.307 (4.156)	3.324** (1.365)
Panel B: Effects on Achievement Levels of schools of choice or assignment				
Hastings and Weinstein (2008)	Information about school test scores - natural experiment Information about school test scores - RCT	Mean test scores SD of 1st choice	0.047*** (0.004) 0.043 (0.024)	0.081** (0.030)
Allende et al. (2019)	Information on returns to school quality	Math test scores SD of chosen school	0.0147 (0.0293)	0.128*** (0.021)
Cohodes et al. (2025)	Customized list of schools with grad rates >75% (paper) Customized list of schools with grad rates >75% (digital) Personalized school recommendation list with high grad rates School Finder	Graduation rate pp of top3 choices	0.837** (0.419) 0.391 (0.578) 0.521 (0.383) 0.172 (0.458)	-0.370 (0.030)
Corcoran et al. (2018)	Customized list of schools with grad rates >75% (paper) Customized list of schools with grad rates >75% (digital) Personalized school recommendation list with high grad rates School Finder	Graduation rate pp of top3 choices	0.326 (0.48) -0.367 (0.665) -1.019* (0.553) -0.339 (0.447)	-0.370 (0.030)
Panel C: Effects on Composition of Students in schools of choice or assignment				
Ainsworth et al. (2023)	Information on school VA : sample eligible for at least one of two top choice Information on school VA : sample ineligible for two top choices	Peer SES of assigned school	0.017 (0.013) -0.009 (0.052)	0.095*** (0.023)
Arteaga et al. (2022b)	Warning about nonplacement risk	% vulnerable students in enrolled school	-0.006 (0.003)	-2.575*** (0.780)
Campos (2024)	Information on school VA Information on school peer quality	Peer Quality pctile in 1st choice	-5.317** (2.373) - 0.079 (2.480) -5.294* (2.921) - 0.563 (2.231)	4.424*** (1.024)

Notes: This table compares the 2SLS effects of attending a majority-White and Asian middle school for Black and Hispanic applicants (column 5) with the effects of information interventions found in the literature (column 4). The dependent variables, described in column 3, include characteristics of school choices and enrollment. When the outcomes or their units differ from those used in the main tables, we re-estimate our baseline 2SLS model from equation (3) using the outcome metrics of the cited studies. For comparability, in Panel C we measure peer SES as in Ainsworth et al. (2023): the mean baseline test score (in standard deviation units) in the middle schools attended by students in a HS program.

Table A11: Peer Effects on Perceived Discrimination

	Peer discrimination		Teacher discrimination		Act on fear discrimination		Fit well other races	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Asian × (High Other-Race MS)	-0.05	-0.22	-0.03	-0.35**	-0.15***	-0.29**	0.10*	-0.11
	(0.05)	(0.18)	(0.04)	(0.15)	(0.05)	(0.13)	(0.05)	(0.14)
Black × (High Other-Race MS)	0.03	0.00	0.08	0.20	0.03	0.32**	0.08	0.19
	(0.07)	(0.15)	(0.08)	(0.14)	(0.07)	(0.16)	(0.08)	(0.15)
Hispanic × (High Other-Race MS)	-0.00	-0.38**	0.05	-0.11	0.04	0.03	-0.00	0.08
	(0.04)	(0.17)	(0.04)	(0.29)	(0.04)	(0.27)	(0.05)	(0.21)
White × (High Other-Race MS)	-0.00	-0.13	-0.00	-0.07	-0.05	-0.07	0.07	0.01
	(0.04)	(0.08)	(0.03)	(0.06)	(0.04)	(0.08)	(0.05)	(0.11)
Asian	0.12***	0.09**	0.09***	0.10***	0.22***	0.22***	-0.05	-0.06
	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)
Black	0.06	0.04	0.11***	0.10**	0.07*	0.05	0.03	0.03
	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.06)
Hispanic	0.04	0.04	0.04	0.06	0.01	-0.01	0.06	0.08
	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)	(0.06)
White mean	0.12	0.12	0.05	0.05	0.12	0.12	0.65	0.65
N	1,871	1,871	1,874	1,874	1,874	1,874	1,869	1,869

Notes: This table reports OLS and 2SLS estimates of the effect of enrolling in middle schools where the majority of peers are from a different race, relative to attending schools enrolling a majority of same-race peers, on measures of perceived discrimination. The effect of exposure to other-race peers is allowed to vary across student race, as capture by the interaction of the exposure dummy "High other race MS" with a dummy indicating respondent race. Column headings summarize the survey measures of perceived discrimination, which use survey questions Q19c,Q20a - Q20c. From left to right, they indicate agreement with the following statements: "My student is likely to be treated negatively by their classmates based on their race" (columns 1-2) , "My student is likely to be treated negatively by their teachers based on their race" (columns 3-4), "My student would feel like they belong in a school where the majority of peers are from a different race" (columns 7-8). The remaining outcome indicates responding positively to the question "Did the fear of negative treatment based on race influence the schools you listed on your student's application?" (columns 5-6).

Table A12: 2SLS Estimates of Peer Effects on Test Score Taking

	Student took:					
	6th grade test	7th grade test	8th grade test	Regents Math in MS	Regents Math in HS	SAT
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Black & Hispanic students					
Majority white&Asian MS	-0.006	0.018	-0.017	-0.037	-0.031	-0.025
	(0.010)	(0.017)	(0.026)	(0.038)	(0.044)	(0.040)
Share white&Asian (10pp)	-0.002	0.004	-0.007	-0.022**	0.011	0.003
	(0.002)	(0.003)	(0.004)	(0.007)	(0.009)	(0.008)
Mean	0.99	0.96	0.91	0.56	0.52	0.62
N	51,532	48,222	45,727	28,727	35,357	39,148
	Panel B: White & Asian Students					
Majority Black&Hispanic MS	-0.011	-0.019	0.009	0.002	-0.039	0.036
	(0.014)	(0.019)	(0.030)	(0.038)	(0.060)	(0.050)
Share Black&Hispanic (10pp)	0.000	-0.004	-0.001	0.004	-0.010	-0.006
	(0.003)	(0.004)	(0.006)	(0.008)	(0.011)	(0.012)
Mean	0.99	0.98	0.93	0.79	0.34	0.75
N	32,857	31,363	29,010	22,037	22,699	24,169

Notes: This table reports 2SLS estimates of middle school demographic composition effects on binary outcomes indicating taking 6th, 7th, and 8th grade state tests, taking the Regents Math exam in middle school or taking it in high school, and taking the SAT. 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 2017 and 2018, column 5 and 6 the 2016, 2017 and 2020-2022 cohorts. The initial experimental sample and control variables are as those defined in the notes of Table 4. Robust standard errors in parenthesis.

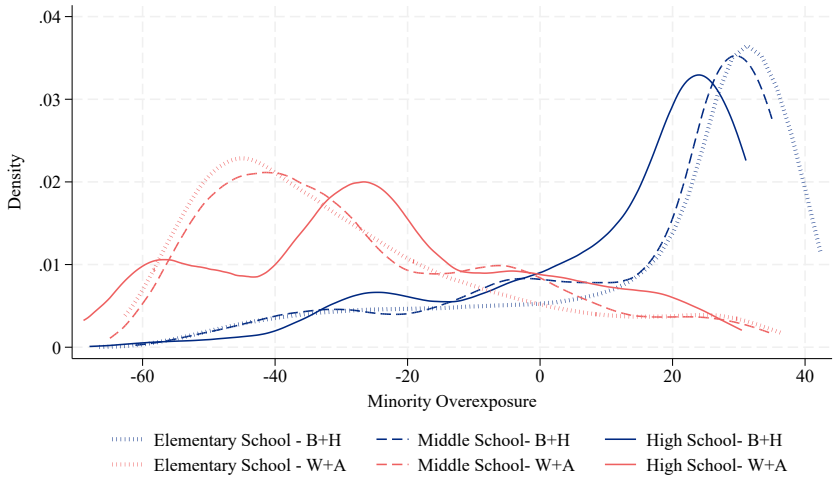
Table A13: 2SLS Estimates of Peer Effects on Test Score Ranking and Beliefs

	Actual Performance and Chances				*Beliefs about Perfomance and Chances*		
	7th grade Math (1)	Tercile in City (2)	Tercile in High-Demand schools (3)	Admission Probability (4)	Tercile in City (5)	Tercile in High-Demand schools (6)	Admission Belief (7)
				Panel A: Black & Hispanic Respondents			
Majority white&Asian MS	0.77**	0.62**	0.22	0.04	1.13*	0.88	-0.06
	(0.31)	(0.30)	(0.24)	(0.05)	(0.62)	(0.63)	(0.19)
Share white&Asian (10pp)	0.08	0.10*	0.05	0.00	0.11*	0.10*	0.01
	(0.06)	(0.05)	(0.04)	(0.01)	(0.06)	(0.06)	(0.03)
Mean	0.09	1.92	1.27	0.51	2.40	2.33	0.60
N	1,518	1,699	1,699	2,608	582	410	1,310
				Panel B: White & Asian Respondents			
Majority Black%&Hispanic MS	-0.09	0.01	0.06	0.04	-0.32	-0.34	-0.08
	(0.22)	(0.21)	(0.25)	(0.04)	(0.24)	(0.31)	(0.10)
Share Black&Hispanic (10pp)	-0.03	0.02	-0.02	0.00	-0.07	-0.14	-0.03
	(0.04)	(0.04)	(0.06)	(0.01)	(0.07)	(0.09)	(0.03)
Mean	0.96	2.62	1.85	0.59	2.63	2.43	0.53
N	1,723	1,823	1,823	3,058	721	595	1,822

Notes: This table reports 2SLS estimates of middle school demographic composition effects on students' test scores, relative academic performance and chances of admission to high-demand or dream programs (columns 1-4) and respondents' beliefs about these (columns 5-7). Panel A restricts the survey sample to Black and Hispanic respondents, while Panel B restricts the sample to White and Asian respondents. The sample and control variables are as defined in the notes of Table 4. Standard errors are clustered at the student level in columns 4 and 7 and robust in the remaining columns.

B Appendix Figures

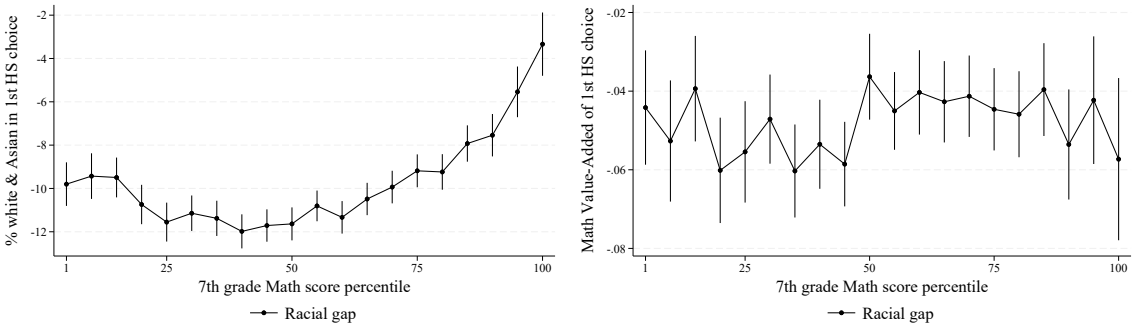
Figure B1: Differences in Exposure to Black & Hispanic Peers by Race and School-Level



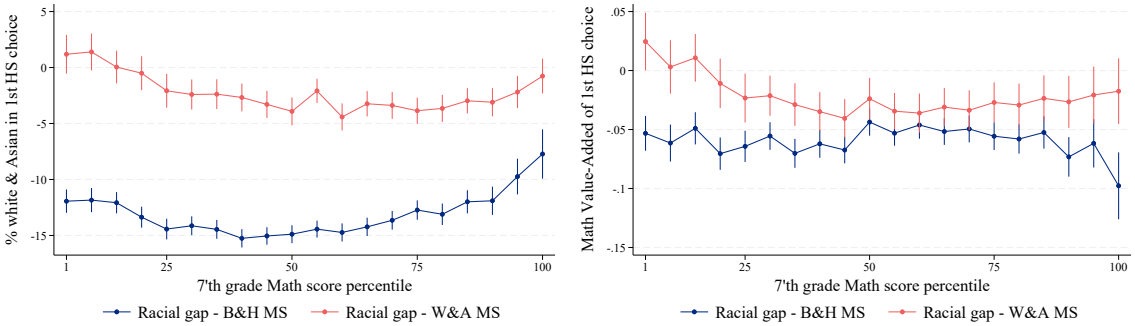
Notes: The figure shows the distribution of minority overexposure, a measure of school segregation, for Black and Hispanic students (blue lines) and White and Asian students (orange lines) in different grade levels (elementary, middle and high school).

Figure B2: Differences in High School Choices by Race, Controlling for District

(a) Conditional on 7th Grade Achievement



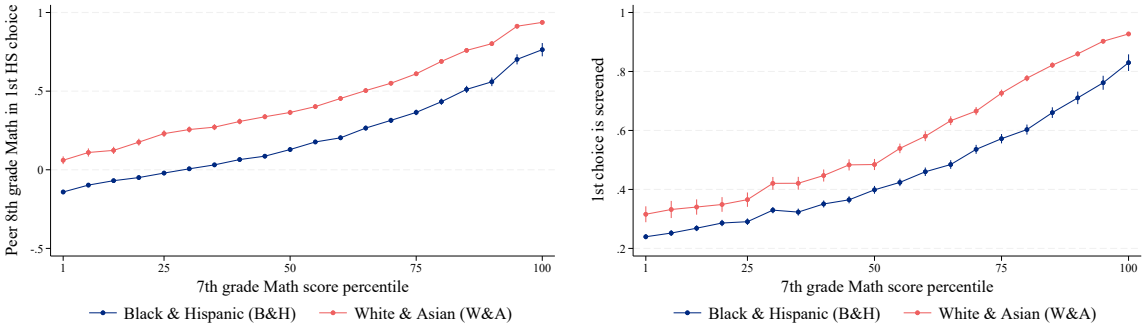
(b) Conditional on 7th Grade Achievement, and Middle School Demographics



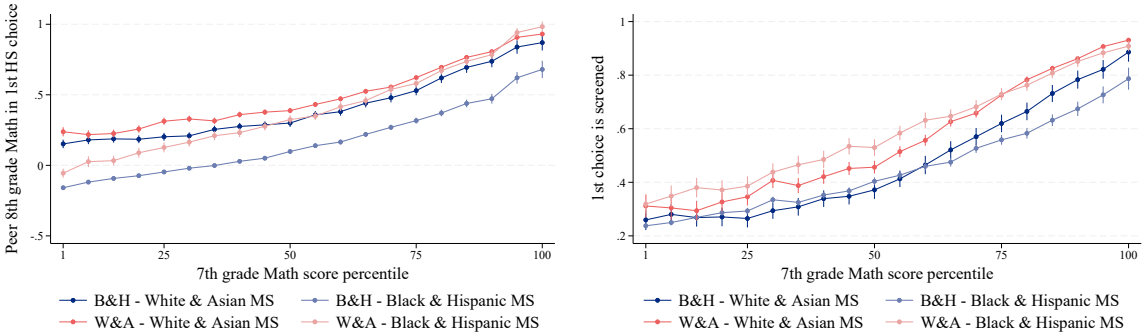
Notes: This figure shows how the racial gap unexplained by differences in residential location varies across student baseline achievement for all students (left), and separately for students enrolled in majority Black and Hispanic middle schools and majority White and Asian middle schools (right). Panel (a) uses 7th grade Math test scores as measures of baseline achievement, while Panel (b) uses 4th grade test scores. The graphs plot the coefficient on a dummy indicating Black and Hispanic applicants in regressions of school value-added in first choices on race, controlling for district of residence. Each dot corresponds to a separate regression restricting to applicants with baseline achievement in a different vingtile. Vertical bars indicate 95% confidence intervals.

Figure B3: Differences in High School Choices by Race: Peer Quality and Selectivity

(a) Conditional on 7th Grade Achievement

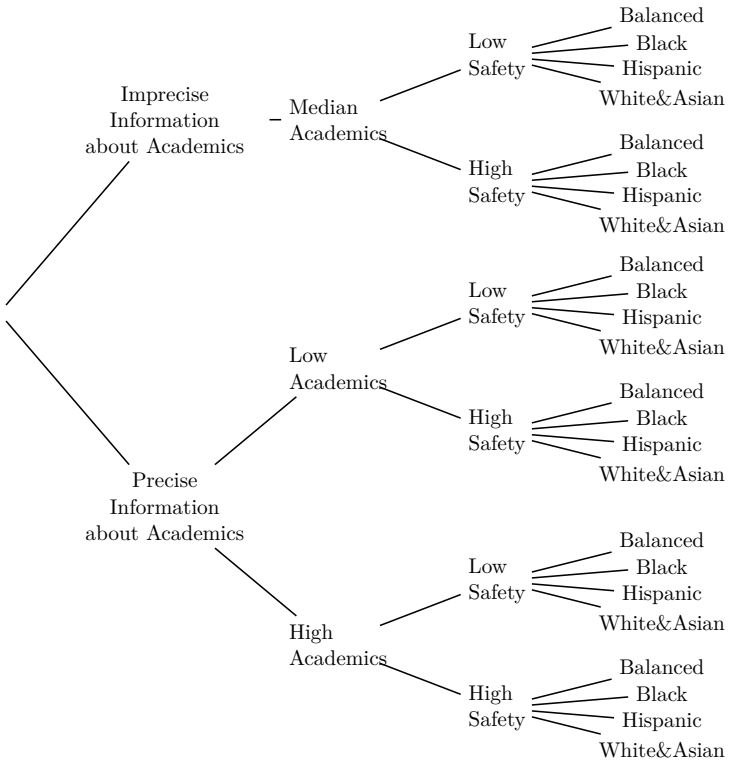


(b) Conditional on 7th Grade Achievement and Middle School Demographics



Notes: This figures plots average peer quality (left) and selectivity (right) of the high school ranked first by applicants' race and middle school test score percentile. Panel (b) distinguishes students by 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 is restricted to applicants applying to high schools in years 2018, 2019, 2020, 2023, 2024, and 2025. The 2021 and 2022 cohorts are not included because 7th grade test scores are not available due to COVID.

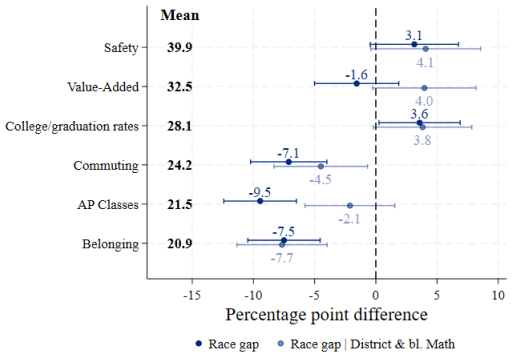
Figure B4: Treatment Arms for the Vignette Experiment



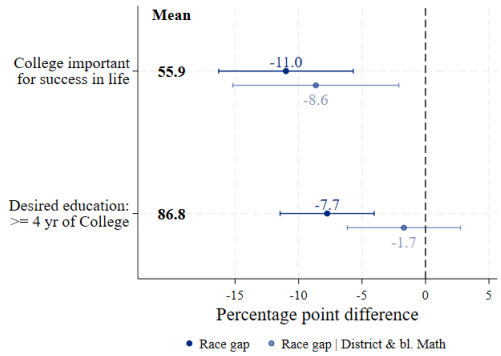
Notes: This figure describes the treatment arms in the school card vignette experiments. 60% of target survey participants were assigned to receive precise information about school academics, while the remaining 40% received imprecise information. Within the Precise Information treatment arm, school cards were equally likely to display either low or high academic ratings. Safety ratings (high or low) and school racial composition (balanced, majority Black, majority Hispanic, or majority White&Asian) were additionally cross-randomized with equal probabilities.

Figure B5: Differences in Stated Preferences and Aspirations

(a) Stated Preferences



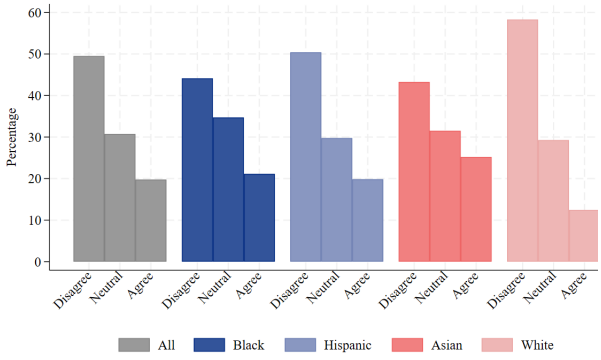
(b) Aspirations



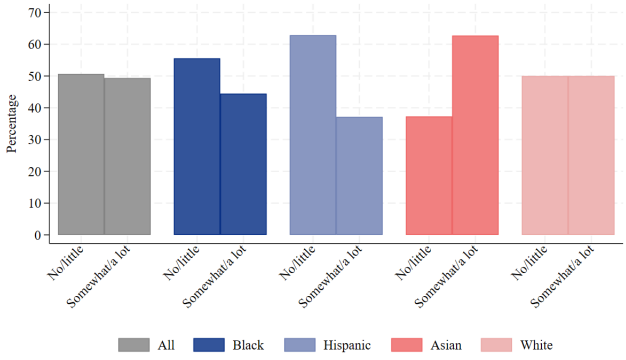
Notes: This figure summarizes racial differences in respondents' stated preferences for school features and aspirations for their children's education. Panel (a) shows the share of respondents who identified each attribute among their top three factors when selecting schools (numbers on the left), alongside racial differences in selection rates (coefficients on the right). Panel (b) displays the share who believe college is important for future success and the share who hope their child will complete at least four years of college, again paired with racial differences. For each outcome, the darker-shade marker reflects raw percentage-point differences across racial groups (Black and Hispanic relative to White and Asian), while the lighter-shade marker conditions on district of residence and middle school math test scores. Error bars represent 95% confidence intervals. This figure uses data from survey questions Q8, Q14 and Q15.

Figure B6: Perceived racial discrimination and its role in high school choices

(a) Discrimination by peers because of race is likely

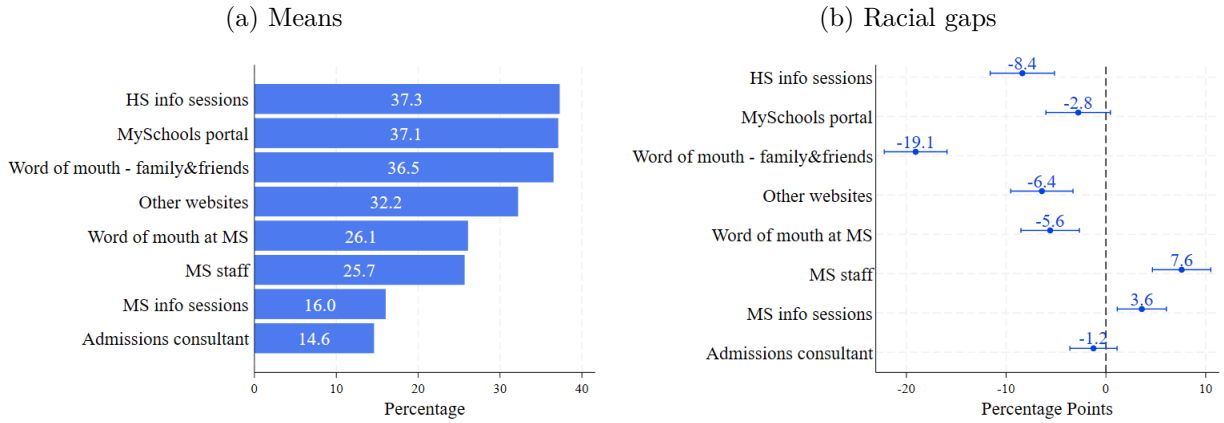


(b) Fear of discrimination affected HS choices



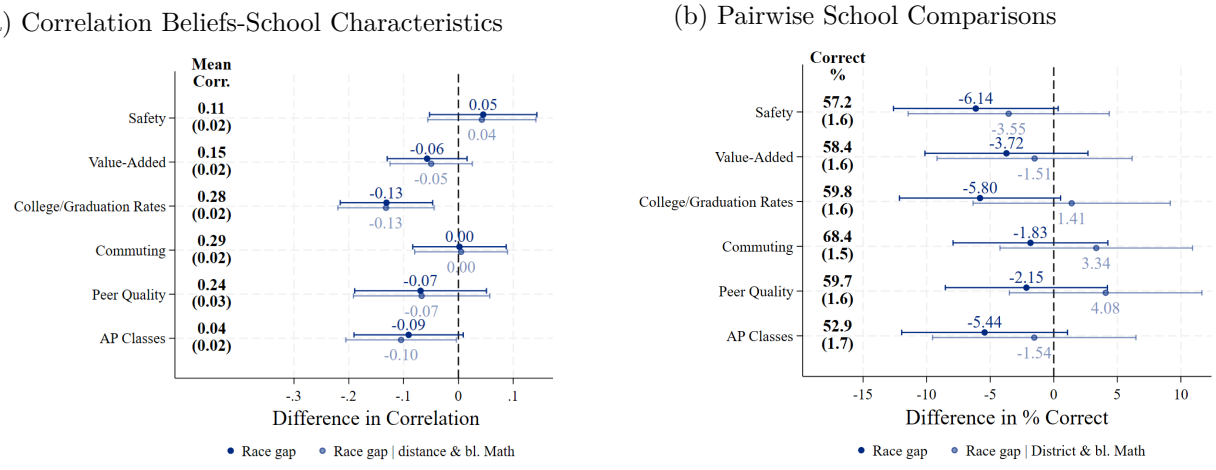
Notes: This figure summarizes perceptions of discrimination and their impact on high school choices. Panel (a) shows the share of respondents who agree to varying extents with the statement, "My student is likely to be treated negatively by their classmates because of their race." Among respondents who reported that their children were likely to face discrimination by teachers or classmates, Panel (b) reports the extent to which fear of negative treatment based on race influenced the schools they listed in their application. This figure uses data from survey questions Q20a and Q20c.

Figure B7: Information Sources



Notes: This figure illustrates how families obtain information about high schools. Panel (a) shows the share of respondents who list each source among their three most important when making high school decisions. Panel (b) shows regression estimates of racial differences in the likelihood of mentioning each source, where each point represents the coefficient on an indicator for Black and Hispanic respondents from a separate regression controlling for district of residence and baseline achievement.

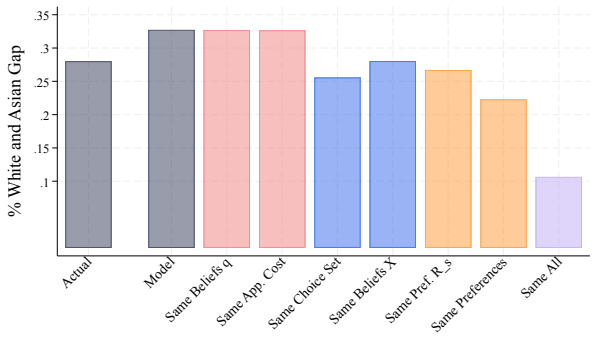
Figure B8: Accuracy of Information



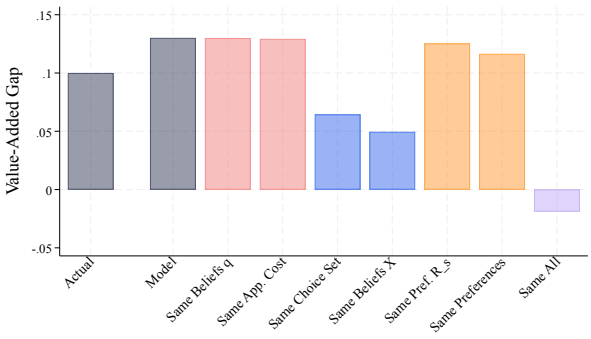
Notes: This figure compares our preferred measure of accuracy of information used in the main text (panel (a)) to an alternative way of measuring belief accuracy: the proportion of respondents who correctly rank a pair of schools (panel (b)). For each measure, the numbers on the left show overall mean outcomes and the bars on the right show racial differences, reported both raw and conditional on district of residence and students' baseline Math achievement. Error bars represent 95% confidence intervals. This figure uses data from survey questions Q10a-g.

Figure B9: Decomposition of the Choice Gap - Using Vignettes Parameters

(a) % White&Asian



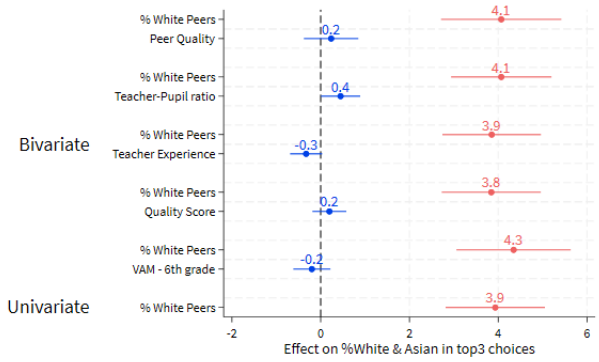
(b) Value-Added



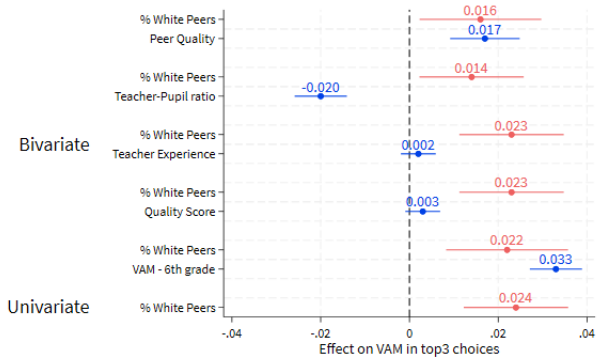
Notes: The first bar plots differences in the characteristics of top 3 school choices of White and Asian applicants relative to Black and Hispanic applicants. The second bar plots the estimated difference based on our model estimates, for a version of the model that constraints relative preferences over school attributes to match those in the vignettes experiment. The remaining bars plot simulated differences in choice characteristics under alternative scenarios in which we replaced one component of the parameters for Black and Hispanic applicants with the corresponding parameters estimated for White and Asian applicants.

Figure B10: Peer Race Effects vs. Effects of Middle School Characteristics

(a) Effects on School Choices Demographics



(b) Effects on School Choices VA



Notes: This figure plots estimates of the coefficients β_1 and β_2 in equation (5) on the mean share of White and Asian students (Panel a) and the mean school value-added (Panel b) among an applicant's top three high school choices. The sample is restricted to Black and Hispanic applicants. Each pair of coefficients in the bivariate rows comes from a separate regression of choice characteristics on two continuous variables measuring across-school standard deviations of the share of White and Asian middle school peers or school quality/teacher experience/school resources/peer quality, depending on the model. The last row reports as reference the estimate of β in the univariate model of equation (3).

C Survey appendix

This appendix provides comprehensive details on the content and implementation of the post-application survey conducted in partnership with the New York City Department of Education during the 2023-2024 high school admission cycle. This appendix is organized as follows. Section C.1 explains the survey logistics, including timeline, the emails sent to participants, and the Qualtrics design. Section C.2 describes the selection of potential participants. Section C.3 describes the survey blocks and the randomization of participants to survey versions. Section C.4 explains the selection of randomized schools in the survey.

C.1 Survey logistics

Survey timeline High school applications in NYC closed on December 5th, 2022. The survey was designed to be sent after families applied to high school but before the offers were sent out on March 9th, 2023. The timeline allowed parents to have at least two weeks to complete it, and the survey had no time constraints beyond the March 6th deadline. Incomplete surveys were automatically submitted by the deadline. Participants who answered at least one question by the deadline received a $10 Amazon gift card.

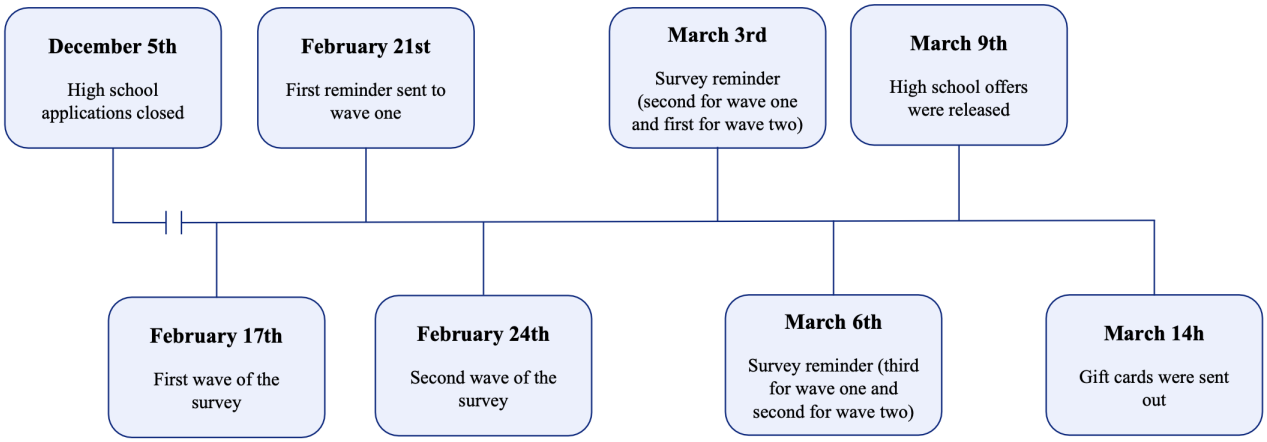
The survey was sent electronically using the email addresses of families used in the high school application process. It was conducted in two waves between February 17th and March 6th, 2023. The first wave, including 11,415 families, was sent on Friday, February 17th. A week later, on Friday, February 24th, the second wave was sent out to 9,986 families. Wave one participants received three reminders: one on February 21st, the second on March 3rd, and the last on March 6th, 2023 (the last day to respond to the survey). Wave two participants received two reminders: one on March 3rd, and the second on March 6th. All gift cards were sent out on March 14th. Figure C1 illustrates the survey timeline.

Survey emails The survey was sent by email in the top three most spoken languages in NYC[51]: English, Spanish, and Simplified Chinese. Figure C2 shows the first email sent to the potential participants.

Qualtrics design The survey was designed on Qualtrics and it was available in English, Spanish, and Simplified Chinese. Participants could change the survey language at any point

[51] Among all of the students enrolled in any NYC public school between SY 2012-13 and 2016-17, the top three home languages were divided as follows: 59.13% English speakers, 22.88% Spanish speakers, and 4.36% Chinese speakers.

Figure C1: Survey Timeline



Notes: This figure illustrates the timeline from the distribution of the first survey wave to the delivery of gift cards, all within the year 2023.

Figure C2: Invitation to Participate. English Version

MIT Economics

Dear parent/guardian,

You are receiving this email because you recently applied to NYC high schools.

We would like to invite you to **complete a 10-minute survey about the deciding factors in your high school choice.** This survey is part of a research study conducted by researchers at the Massachusetts Institute of Technology in partnership with the New York City Department of Education.

This study is independent from the high school choice placement process and the survey is **confidential** and **voluntary**. You must be **at least 18 years old** to participate in this study.

For more information about the study and to participate in the survey, please **Follow this link:**
Take the survey

Or copy and paste the URL below into your internet browser:
https://mit.co1.qualtrics.com/jfe/preview/previewId/5d4b996c-9d20-432c-a2b2-f3b5d98d7ff2/SV_0JoXfRC5AlOew0C?Q_CHL=preview&Q_lang=EN

The first 5,000 respondents will receive a **$10 Amazon gift card as compensation for their time**.

Your participation is very important to us! By taking this survey, you'll help us better understand how families choose high schools. Our goal is to use the insights gained to help future families make informed decisions about school choices.

Please note: the survey will close on March 6th before midnight Eastern Time and the gift cards will be sent out then.

Thank you very much!

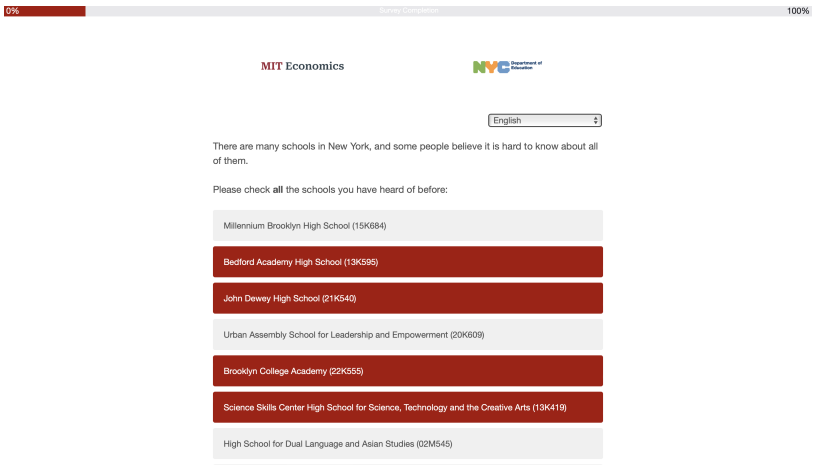
The research team

If you have any questions, you may contact the principal investigator, Clemence Idoux at cidoux@mit.edu

when filling it out. All questions were marked as optional, except the consent to participate one: to access the survey, participants had to check a box stating that they were over 18 years old. We personalized the survey to present participants with different sets of schools. For instance, question 9 displayed a distinct set of high-demand and popular schools based

on the participant's borough. See section C.4 for more details.

Figure C3: Question 9. Variation in Schools Presented to Respondents Based on Their Characteristics



Notes: This figure shows one of the possible school sets presented to respondents in Q9. The set of schools in this question was tailored to each respondent, as explained in detail in section C.4.

C.2 Selection of participants

The sample of all eligible participants consisted of parents or guardians of students applying to start 9th grade in fall 2023. We categorized participants into survey waves based on their informativeness by assigning each participant a priority, which sorted them into different waves. We only sent the survey to some of the waves. Eligible participants' priorities range from one to six and were determined as follows:

1. Survey priority is 1 if the student was enrolled in a DOE school in 6th grade, has all demographic information, has all baseline scores, and is at risk of middle school assignment.

2. Survey priority is 2 if the student was enrolled in a DOE school in 6th grade, has all demographic information, has all baseline scores, and is not at risk of middle school assignment.

3. Survey priority is 3 for all other students

The first wave of the survey included all families of students with a survey priority of 1. The second wave included the first 10,000 priority 2 students, sorted by their scrambled

ID. The additional wave of participants was created for potential expansion of the survey, although it was not used. The final sample of potential participants comprised 21,401 parents or guardians.

C.3 Survey questions, survey block design and randomization to survey version

C.3.1 Description of survey questions and survey blocks

The complete survey had a total of 47 questions, including the consent question and the end-of-survey comment box. The questions are described in Table C1. We grouped all the questions by type and created five different blocks (1a, 1b, 2, 3a, 3b), which are shown in Table C2. Block 1 consists of the information questions. Block 2 includes questions about student aspirations, beliefs about student academic performance, and knowledge of tiebreakers and how they affect application decisions. Block 3 is the vignette experiment described further in Section 3.2.1 of the paper. It consists of questions about school preferences and perceptions of discrimination and has two versions: 3a and 3b. The version 3a includes precise academic information, while 3b presents imprecise academic information.

Table C1: High School Post-Application Survey Questions and Answer Options

Question	Question text	Possible answers
Q0	Age verification / consent to participate in the study.	Consent checkbox certifying respondent is at least 18 years old
Q1	What is your relationship to the student?	Parent; Brother/Sister; Grandparent; Uncle/Aunt; Other: please specify
Q2	Who played the most important role in deciding how to fill out the high school application?	The student; The parents/guardians; Both the parents/guardians and the student; School staff; Other: please specify
Q3	In the past year, how many times did you talk to other parents/guardians from your student's middle school about which high schools to apply to?	Never; 1–5 times; More than 5 times
Q3b	What were the most important sources of information for deciding which schools to include in your student's application? Rank up to three (1 should be the most important to you).	Word-of-mouth from other parents/guardians at your student's middle school; Word-of-mouth from family and friends; School-admissions consultant; Attending information sessions at your student's middle school; Attending high school information sessions; Other websites (InsideSchools, GreatSchools etc.); MySchools' high school directory; Staff at your student's middle school
Q3b.other	If you think of an important source of information that is not mentioned, please add it here:	Open text response
Q4	How important is it to you that your student goes to the same high school as their friends from middle school?	Not important at all; Not very important; Somewhat important; Very important
Q5	Which school was ranked first on your student's application?	School; Program dropdown
Q6	Please select your "dream school" — the school that you would pick if your student could attend any high school in the city — from the drop down below. This might be a school that was on your student's application, but it doesn't have to be.	School; Program dropdown
Q7a	Did you list the dream school on your student's application?	Yes; No
Q7b	Why not?	We thought our student's chances to get in were too low; Our student was not eligible to apply to the program; We realized too late that the program required an admission test/audition/assessment; We knew the program required an admission test/audition/assessment, but our student did not want to complete it; Other: please specify
Q7c	How likely do you think your student would be to get into the dream school if you had listed it first on your application?	Impossible (0% chance); Almost impossible (1–10% chance); Somewhat unlikely (11–33% chance); Somewhat likely (34–66% chance); Very likely (67–89% chance); Almost certain (90–100% chance)
Q7c.2	How likely do you think your student is to get into the dream school?	Impossible (0% chance); Almost impossible (1–10% chance); Somewhat unlikely (11–33% chance); Somewhat likely (34–66% chance); Very likely (67–89% chance); Almost certain (90–100% chance)

Question	Question text	Possible answers
Q8a	What is most important to you when you choose a school? Rank up to three (1 should be the most important to you).	Ease of commute; Extracurricular activities offered; Advanced classes (for instance AP classes) offered; Other incoming students are academically-focused; Safe school environment; Graduation and/or college enrollment rate(s); Academic progress of students at the school; Whether I think my student would feel like they belong; School's neighborhood; School campus and facilities; School interest areas; Demographics of the school; Career and Technical education
Q8a.other	If you think something important that influenced your decision is not mentioned, please add it here:	Open text response
Q9	There are many schools in New York, and some people believe it is hard to know about all of them. Please check all the schools you have heard of before:	Embedded school list: school1_Q9 through school10_Q9
Q10	Now we would like to ask some questions about how you think different high schools compare. Please fill out based on what you already know about these schools; you do not need to do any additional research. If you're unsure, it is okay to take a guess.	Instruction text; no response options
Q10a	Which of these schools would take your student more time to get to by public transit?	Embedded school A; Embedded school B
Q10b	Which of these schools attracts more academically-focused students?	Embedded school A; Embedded school B
Q10c	Which of these schools has more students that enroll in college?	Embedded school A; Embedded school B
Q10d	Which of these schools better prepares students for their Regents exam?	Embedded school A; Embedded school B
Q10f	Which of these schools offers the safest environment for students?	Embedded school A; Embedded school B
Q10g	Which of these schools offers more AP courses?	Embedded school A; Embedded school B
Q10_v2	Now we would like to ask some questions about high schools in your borough. Please fill out based on what you already know about these schools; you do not need to do any additional research. If you're unsure, it is okay to take a guess.	Instruction text; no response options
Q10a_v2	How far is the selected school from your home compared to other schools in your borough, by public transit?	1 = Among the closest schools; 2; 3; 4 = Among the furthest schools
Q10b_v2	Does the selected school enroll fewer or more students in college than other schools in your borough?	1 = Fewer students that enroll in college compared to other schools in my borough; 2; 3; 4 = Many more students that enroll in college compared to other schools in my borough
Q10c_v2	Does the selected school have fewer or more academically-focused students than other schools in your borough?	1 = Fewer academically-focused students compared to other schools in my borough; 2; 3; 4 = Many more academically-focused students compared to other schools in my borough
Q10d_v2	How well does the selected school prepare students for their Regents exams compared to other schools in your borough?	1 = Among schools with the worst preparation for Regents; 2; 3; 4 = Among schools with the best preparation for Regents
Q10f_v2	Is the selected school safer than other schools in your borough?	1 = Among the least safe schools; 2; 3; 4 = Among the safest schools

Question	Question text	Possible answers
Q10g_v2	Does the selected school offer fewer or more AP classes than other schools in your borough?	1 = Among schools that offer the fewest AP classes; 2; 3; 4 = Among schools that offer the most AP classes
Q11	Now, we would like to ask you questions about your student's academic performance.	Instruction text; no response options
Q11a	How do you think your student's 7th grade final grades compare to other students in the city?	Worse than most students in the city (bottom third); About average compared to other students in the city (middle third); Better than most students in the city (top third)
Q12	If your student received an offer to the selected school, how well do you think they would perform compared to the other students in the school?	Worse than most students in this school (bottom third); About average compared to other students in this school (middle third); Better than most students in this school (top third)
Q13	How likely do you think your student would be to get into the selected school if you had listed it first on your application?	Impossible (0% chance); Almost Impossible (1–10% chance); Unlikely (11–33% chance); Somewhat likely (34–66% chance); Very likely (67–89% chance); Almost certain (90–100% chance)
Q14	Now we would like to ask you questions about your aspirations for your student.	Instruction text; no response options
Q14a	Do you believe that it is important to go to college in order to do well in life?	Strongly disagree; Somewhat disagree; Neither agree nor disagree; Somewhat agree; Strongly agree
Q15	What is the minimum level of education that you would like your student to complete?	It does not matter; Complete high school; 2 years of college; 4 years of college; Graduate school (master, PhD, law or medical school etc.)
Q16a	Did you know that you could see your student's random number in MySchools this year?	Yes; No
Q16b	Did knowing the random number impact which schools you included on your student's application?	Yes; No
Q16c	If yes, how?	It made us apply to more programs; It made us apply to fewer programs; It made us apply to more high-demand programs; It made us apply to fewer high-demand programs; Other: please specify
Q17	Imagine you set up a meeting with your student's school counselor to discuss high school options. The counselor would like to know how you feel about some high schools. All schools have common listed characteristics.	Instruction text; no response options
Q17a	How likely are you to list this school on your student's application?	1 = Very unlikely; 2; 3; 4; 5; 6 = Very likely
Q17b	How likely are you to list this school on your student's application?	1 = Very unlikely; 2; 3; 4; 5; 6 = Very likely
Q18a	Please rank these 3 schools from your most (1) to your least (3) preferred choice.	Rank School 1; Rank School 2; Rank School 3
Q18b	Please rank these 3 schools from your most (1) to your least (3) preferred choice.	Rank School 1; Rank School 2; Rank School 3
Q19	Many things can make a student feel comfortable or not comfortable at a school. We would like to ask how important some of these things are to you and whether they played a role when filling out your student's application. Do you agree with the following statements?	Instruction text; no response options

Question	Question text	Possible answers
Q19a	My student would fit in well at a school that places a strong emphasis on grades.	Strongly disagree; Somewhat disagree; Neither agree nor disagree; Somewhat agree; Strongly agree
Q19b	My student would fit in well at a school where most students have different grades than them.	Strongly disagree; Somewhat disagree; Neither agree nor disagree; Somewhat agree; Strongly agree
Q19c	My student would feel like they belong in a school even if most students are from a different race or ethnicity.	Strongly disagree; Somewhat disagree; Neither agree nor disagree; Somewhat agree; Strongly agree
Q20a	My student is likely to be treated negatively by their classmates based on their race.	Strongly disagree; Somewhat disagree; Neither agree nor disagree; Somewhat agree; Strongly agree
Q20b	My student is likely to be treated negatively by their teachers based on their race.	Strongly disagree; Somewhat disagree; Neither agree nor disagree; Somewhat agree; Strongly agree
Q20c	Did the fear of negative treatment based on race influence the schools you listed on your student's application?	Not at all; Little; Somewhat; A great deal
End of survey	Thank you for taking the time to complete this survey. Your feedback is greatly appreciated. If you have any additional comments, please share them below.	Open text response

C.3.2 Randomization of participants to survey version

To reduce the time it takes to complete the survey and increase participation, we devised eight different survey versions by creating different combinations of the five question blocks. Each survey version consisted of between 31-35 questions (see Table C3). All survey versions included the consent to participate, general questions, and the end-of-survey comment box. Potential participants were randomly assigned to one of the eight survey versions. The marginal probability for each block was 75% for block 1 (37.5% for 1a and 37.5% for 1b), 50% for block 2, and 75% for block 3 (37.5% for 3a and 37.5% for 3b). The detailed distribution of blocks to survey version is shown in Table C3).

Table C4 evaluates the covariate balance and attrition rates by survey version. Among all the balance regressions conducted, the majority show no statistically significant relationship between survey version assignment and the covariates. Similarly, in most of the response attrition regressions, the coefficients do not show statistical significance. These attrition findings remain consistent for all participants and when segmenting the sample by white and Asian as well as by Black and Hispanic (minority). The results confirm that the survey randomization successfully achieved the expected balance across the covariates. Regarding attrition, there is no statistically significant difference in the response rate observed by survey version among all potential participants. Two small differences are evident when dividing the sample by race. In survey version 8, white and Asian potential participants are slightly more likely to respond. Similarly, in survey version 3, Black and Hispanic potential participants show a slightly higher likelihood of responding.

Table C2: Survey Questions by Block and Type

Type	Description	Number of questions and type
	General Block	
Age verification	Question to ensure participant is old enough.	**1 checkbox question** (Q0)
General questions	Questions about the relationship with the student, who played the most important role in the application, sources of information, the importance of going to school with friends, attention check, dream school, important aspects when choosing a school.	**15 possible questions:** - 9 multiple choice (Q1, Q2, Q3, Q3b, Q4, Q5, Q6, Q8a, Q9) - 2 open-ended (Q8a.2, Q3b.2) - 1 Yes/No question (Q7a): If "Yes," 1 extra multiple choice question (Q7c); if "No," 2 extra multiple choice questions (Q7b, Q7c.2)
End of survey	Question to leave any comments.	**1 open-ended question** (Q21)
	Block 1a	
Information (version 1)	Questions comparing two high schools in terms of commuting time by public transportation, academically focused students, college enrollment, Regents preparation, safe environment, and AP courses.	**6 multiple choice questions with two options each** (Q10a, Q10b, Q10c, Q10d, Q10f, Q10g)
	Block 1b	
Information (version 2)	Questions comparing a high school to the ones in the borough of residence in terms of commuting time by public transportation, academically focused students, college enrollment, Regents preparation, safe environment, and AP courses.	**6 multiple choice questions following 1-4 Likert scale** (Q10a_v2, Q10b_v2, Q10c_v2, Q10d_v2, Q10f_v2, Q10g_v2)
	Block 2	
Beliefs on academic performance and admission probability	Questions about beliefs on student 7th grade grades compared to all students in the middle school and the city, and about likelihood to admission to a school.	**3 multiple choice questions** (Q11a, Q12, Q13)
Aspirations for the student	Questions about the importance of going to college, and aspirations for the highest level of education.	**2 multiple choice questions** (Q14a, Q15)
Tiebreaker knowledge	Questions about knowledge of the tiebreaker number and how that affected the application.	**3 possible multiple choice questions:** - 1 Yes/No question (Q16a): If "Yes," 1 extra Yes/No question (Q16b); if "No," 1 extra Yes/No question (Q16c)
	Block 3a	
Preferences for attributes (experiment, version 1)	Two types of questions, the first belongs to a vignette experiment with hypothetical schools that varied by safety rating, academic performance ratings, and racial composition (read more on Section 3.2.1). The second type of question is about perceived race-based discrimination.	**10 possible questions:** - 9 multiple choice (Q17a, Q17b, Q18a, Q18b, Q19a, Q19b, Q19c, Q20a, Q20b) - 1 extra multiple choice if the response to any of the race-related questions was neutral or some degree of agreement (Q20c).
	Block 3b	
Preferences for attributes (experiment, version 2)	Two types of questions, the first belongs to a vignette experiment with hypothetical schools that varied by safety rating, academic performance ratings, and racial composition (read more on Section 3.2.1). The second type of question is about perceived race-based discrimination.	**10 possible questions:** - 9 multiple choice (Q17a, Q17b, Q18a, Q18b, Q19a, Q19b, Q19c, Q20a, Q20b) - 1 extra multiple choice if the response to any of the race-related questions was neutral or some degree of agreement (Q20c).

Table C3: Eight Survey Versions and Their Respective Block Combinations

Survey version	Blocks included	Number of possible questions (from blocks + general)
1	1a, 2	14 + 17 = 31
2	1b, 2	14 + 17 = 31
3	1a, 3a	16 + 17 = 33
4	1a, 3b	16 + 17 = 33
5	1b, 3a	16 + 17 = 33
6	1b, 3b	16 + 17 = 33
7	2, 3a	18 + 17 = 35
8	2, 3b	18 + 17 = 35

Table C4: Survey Attrition and Covariate Balance

	Mean	Survey Version 1 Beliefs, aspirations, tiebreaker, info v1			Survey Version 2 Beliefs, aspirations, tiebreaker, info v2			Survey Version 3 Info v1, 2-factor cards			Survey Version 4 Info v1, 3-factor cards			Survey Version 5 Info v2, 2-factor cards			Survey Version 6 Info v2, 3-factor cards			Survey Version 7 Beliefs, aspirations, tiebreaker, 2-factor cards			Survey Version 8 Beliefs, aspirations, tiebreaker, 3-factor cards		
Dependent Variable	All	All	W&A	B&H	All	W&A	B&H	All	W&A	B&H	All	W&A	B&H	All	W&A	B&H	All	W&A	B&H	All	W&A	B&H	All	W&A	B&H
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)
Panel A: Attrition																									
Respondent	0.17	0.006	-0.005	0.014	-0.004	-0.01	-0.003	0.003	-0.013	0.019*	0.003	-0.001	0.001	-0.006	-0.009	-0.007	-0.002	0.004	-0.002	-0.003	0.009	-0.007	0.002	0.020*	-0.012
		(0.008)	(0.013)	(0.01)	(0.008)	(0.013)	(0.01)	(0.009)	(0.014)	(0.011)	(0.007)	(0.012)	(0.009)	(0.009)	(0.014)	(0.011)	(0.007)	(0.012)	(0.009)	(0.008)	(0.014)	(0.011)	(0.007)	(0.012)	(0.009)
N		21,401	9,064	11,778	21,401	9,064	11,778	21,401	9,064	11,778	21,401	9,064	11,778	21,401	9,064	11,778	21,401	9,064	11,778	21,401	9,064	11,778	21,401	9,064	11,778
Panel B: Covariates balance																									
White+Asian	0.42		0.006			0.013			0.002			0.007			-0.017			0.003			-0.007			-0.011	
			(0.01)			(0.01)			(0.011)			(0.009)			(0.011)			(0.01)			(0.011)			(0.009)	
Black+Hispanic	0.55		-0.005			-0.013			-0.004			-0.005			0.018			0.001			0.007			0.004	
			(0.01)			(0.01)			(0.011)			(0.009)			(0.011)			(0.01)			(0.011)			(0.01)	
ELL	0.09		0.005			0.004			-0.002			0.005			0			-0.002			-0.006			-0.006	
			(0.006)			(0.006)			(0.006)			(0.005)			(0.007)			(0.005)			(0.006)			(0.005)	
FRPL	0.73		0.014			-0.002			0.012			-0.005			-0.015			-0.002			0.011			-0.009	
			(0.009)			(0.009)			(0.01)			(0.008)			(0.01)			(0.009)			(0.01)			(0.008)	
Baseline English	0.28		-0.024			0.005			0.034*			0.011			0.003			-0.013			0.019			-0.022	
			(0.018)			(0.018)			(0.02)			(0.017)			(0.02)			(0.017)			(0.02)			(0.017)	
Baseline Math	0.3		-0.007			0.003			0.023			-0.003			-0.012			-0.021			0.036*			-0.006	
			(0.018)			(0.018)			(0.02)			(0.017)			(0.02)			(0.017)			(0.02)			(0.017)	
Borough K	0.31		-0.005			0.008			-0.017			-0.008			0.016			-0.007			0.019*			0	
			(0.01)			(0.01)			(0.01)			(0.009)			(0.011)			(0.009)			(0.01)			(0.009)	
Borough M	0.09		0.007			0.002			0.004			-0.001			-0.007			0.011**			-0.015**			-0.005	
			(0.006)			(0.006)			(0.006)			(0.005)			(0.006)			(0.005)			(0.006)			(0.005)	
Borough Q	0.35		-0.004			-0.015			-0.006			0.01			-0.01			-0.003			0.014			0.01	
			(0.01)			(0.01)			(0.011)			(0.009)			(0.011)			(0.009)			(0.011)			(0.009)	
Borough R	0.08		0.002			0.009			0.009			-0.006			0			-0.005			0			-0.005	
			(0.006)			(0.006)			(0.006)			(0.005)			(0.006)			(0.005)			(0.006)			(0.005)	
Borough X	0.18		-0.001			-0.004			0.01			0.005			0.001			0.004			-0.018**			0	
			(0.008)			(0.008)			(0.009)			(0.007)			(0.009)			(0.007)			(0.009)			(0.007)	
Participants in this version			2,688			2,650			2,153			3,233			2,135			3,161			2,179			3,202	
Percentage			12.56%			12.38%			10.06%			15.11%			9.98%			14.77%			10.18%			14.96%	
N			21,401			21,401			21,401			21,401			21,401			21,401			21,401			21,401	

Notes: This table reports the attrition and balance among all potential participants by each of the eight survey versions. Column 1 shows the sample means for each dependent variable. Panel A reports coefficients from regressions of being a respondent on a survey version dummy. Columns 2, 5, 8, 11, 14, 17, 20, and 23 report the coefficients of the sample of all potential participants. Columns 3, 6, 9, 12, 15, 18, 21, and 24 report the coefficients for the sample of white & Asian potential participants. Columns 4, 7, 10, 13, 16, 19, 22, and 25 report the coefficients for the sample of Black & Hispanic potential participants. Panel B reports coefficients from regressions of the variables listed in each row on a survey version dummy. Columns 3, 6, 9, 12, 15, 18, 21, and 24 report the coefficient for the sample of all participants. The "Participants in this version" row indicates the potential participants in the survey version. The percentage indicates the percentage of all the potential participants who received that survey version.

C.4 School Selection for Randomized Survey Questions

C.4.1 Definition of school attributes

We consider the following school-level characteristics to select the schools that populate the embedded data of the survey.

Attributes:

- **Demographics:** Ethnic/racial composition of students enrolled in school during the 2021-2022 school year, using all grades 9-12. In particular, we care about the share of white and Asian students (or the share of Black and Hispanic students) in the school.
- **Baseline scores:** Average (standardized) 7th grade test scores of the students enrolling in 9th grade in 2020-2021, by school. This means the test scores are typically measured in 2018-19 SY.
- **Popularity:** Popularity is the share of applicants rejected to applicants accepted for each program at each school in the 2022 admission cycle. We aggregate at the school level using a weighted average across programs at the school, with weights proportional to program capacities. The data used includes schools from any of the five NYC boroughs: Bronx (X), Brooklyn (K), Manhattan (M), Queens (Q), Staten Island (R).
- **Admission method:** We consider as screened schools those that in the 2022-2023 program crosswalk had at least one program that screened students on the basis of academics or both language and academics. These are the high school programs available for the 2023-2024 school year.
- **Language and AP stem classes:** Number of language classes and AP classes in STEM subjects offered by each school.
- **College attendance:** Share of students enrolling in college within 6 months of (on time) graduation per school for 2020-21 SY.
- **Safety:** Percent of students that felt safe in the hallways, bathrooms, locker rooms, and cafeteria by school during the 2019-20 SY.
- **Size:** Total enrollment count at school for grades 9-12 in the 2021-22 SY.

- **Applicants per seat:** Total number of applicants at the school (regardless of whether they got in a preferred school or not) per seat in 2022 admission cycle. This is a school-level measure.
- **Regents VA:** OLS VA on Algebra 1 and ELA Regents using test scores from years 2013 to 2017 cohorts (cohort = fall of 9th grade) and 7th grade baselines.
- **College VA:** OLS VA on a dummy for whether a student enrolls in any type of college using data from 2013 to 2016 cohorts and 7th grade baselines.

C.4.2 Districts' school choice set construction based on school characteristics

The set of high schools eligible for inclusion in certain survey questions was determined as follows:

1. Start from schools in the 2021-2022 high school directory and keep only those in the 2022-2023 program crosswalk.
2. Drop specialized schools, special districts (75 and 79), and home schools.
3. For each district, take a subset of schools that:
    - are in the same borough, or
    - are out of borough but to which at least 1% of students in the district applies in the 2022 cycle.

This returns, on average, 143 schools per district. The average share of students in the district applying to a school in this choice set is 5%.

C.4.3 Selection of high-demand high schools: Questions 12, 13, and part of 9

We selected a few high-demand schools per borough: seven for Manhattan, two for Staten Island, and six for Queens, the Bronx, and Brooklyn. The high-demand schools were determined using the following criteria:

- In the top 20 schools per popularity (share of applicants rejected to applicants accepted) among students residing in the borough.
- In the top 20 schools in terms of applicants per seat among students residing in the borough.

- In the top quintile of average baseline (7th grade) Math test scores across schools in the city.

We then ranked the selected schools based on popularity, applicants per set, and baseline Math. We chose the highest-ranked schools while ensuring some variation in the demographic composition of the schools selected per borough. Specifically, we ensure that at least one school selected per borough had a high share of white and Asian students (>50%) and at least two schools had at least 26% white and Asian students. If none of the top six highest-ranked schools had these characteristics, we replaced the lowest-ranked school among the top six with the highest-ranked school with enough demographic variation in the student body composition.

C.4.4 Selection of 10 "known" schools: Choice question (Q9)

We assign each student a list of 10 schools, based on their district of residence. We start with the district-specific choice set of schools (on average 105 schools) and we select 10 schools as follows:

- **Schools 1 and 2:** Randomly chosen among the high-demand schools of the district borough. Randomization at the student level.
- **School 3:** A school with a high share of white and Asian students. That is, a school with a share above 26% of white and Asian students, which corresponds to the top 25% of schools in the city-wide distribution. For each district, we randomly selected two such schools from the district choice set as follows: one with high baseline Math test scores and one with low baseline Math test scores. High baseline Math test scores are the top 25% of schools city-wide, while low baseline Math are the bottom 50% of schools city-wide. If the restrictions returned an empty set, we selected the school with the highest share of white and Asian students from high (low) Math baseline schools. If empty again, we selected the school with the highest (lowest) baselines among schools with a high share of white and Asian students. Finally, we randomized at the student level between these two white schools.
- **School 4:** A school with a high share of Black and Hispanic (minority) students. That is, a school with a share above 94% of minority students, which corresponds to the top 25% of schools in the city-wide distribution. For each district, we randomly selected two such schools from the district choice set as follows: one with high baseline Math

test scores and one with low baseline Math test scores. We followed the same procedure as for school 3. If the restrictions returned an empty set, we selected the school with the highest share of minority students, among high (low) Math baseline schools. If empty again, we selected the school with the highest (lowest) baselines among schools with a high share of minority students. Finally, we randomized at the student level between these two minority schools.

- **School 5:** A school with a high share of Black students. That is, a school with a share above 41% Black students, which corresponds to the top 25% of schools in the city-wide distribution. For each district, we randomly selected two such schools from the district choice set as follows: one with high baseline Math test scores, and one with low baseline Math test scores. We followed the same procedure as for schools 3 and 4. Finally, we randomized at the student level between these two Black schools.

- **School 6:** A school with high Regents Math VA. That is, above 0.35 standard deviation, which corresponds to the top 25% of schools in the city-wide distribution. For each district, we randomly selected two such schools from the district choice set as follows: one with a high share of white and Asian students, and one with a lower share of white and Asian students. A high share of white and Asian is above 26%, or top 25% of schools. The low share of white and Asian is below 26%. If the restrictions returned an empty set, we selected the school with the highest value-added among high-white and Asian (low-white and Asian) schools. If empty again, we selected the school with the highest (lowest) share of white and Asian students among high-VA schools. Finally, we randomized at the student level between these two high VA schools.

- **School 7:** A school with low Regents Math VA. Thar is, a school corresponding to the bottom 25% of schools in the city-wide distribution. We use the exact same procedure described for school 6, but for low-VA schools to select two schools per district. Then, we randomized at the student level between these two low VA schools.

- **School 8:** A school that screens students on the basis of academics. For each district, we randomly selected two such schools from the district choice set as follows: one with a high share of white and Asian students, and one with a lower share of white and Asian students. A high share of white and Asian students is above 26%, or top 25% of schools. The low share of white and Asian students is below 26%. If the restrictions returned an empty set, we selected the school with the highest (lowest) share of white

and Asian students among screened schools. Finally, we randomized at the student level between these two screened schools.

- **School 9:** A school that does not screen students on the basis of academics. For each district, we randomly selected two such schools from the district choice set as follows: one with a high share of white and Asian students, and one with a lower share of white and Asian students. If the restrictions returned an empty set, we selected the school with the highest (lowest) share of white and Asian students among unscreened schools. Finally, we randomized at the student level between these two unscreened schools.
- **School 10:** A large school. That is, a school with more than 622 students, which corresponds to the top 25% of schools in the city-wide distribution. For each district, we randomly selected two such schools from the district choice set as follows: one with a high share of white and Asian students, and one with a lower share of white and Asian students. If the restrictions returned an empty set, we selected the school with the largest size among high-white and Asian (low-white and Asian) schools. If empty again, we selected the school with the highest (lowest) share of white and Asian students among large-size schools. Finally, we randomized at the student level between these two large schools.

C.4.5 Selection of two schools to compare: Information question (Q10, version 1)

We measure information about schools by asking to compare two schools along the following school characteristics: commute time by public transit, baseline test scores, college enrollment rates, Regents VA, safety, and the number of AP classes. For each district and each school characteristic, we selected four pairs of schools:

1. Both are high-white and Asian-share
2. Both are non-high-white and Asian share
3. The first is high-white and Asian and the second is not
4. The second is high-white and Asian and the first is not

In each pair, the first school is the one with the highest value of the school characteristics of interest. With high-white and Asian we mean schools with a share of white and Asian

students above 26%, corresponding to the 25% of schools with the highest share of white and Asian students in the city.

We selected the schools among the ones in the district choice set, further restricting to schools ranked by at least 2% of students in the district. This limits the choice set for each district to 77 schools per district, on average. A school in this subset is ranked on average by 9% of students residing in the district.

For each school pair, we randomly selected the first school from the restricted choice set, conditional on the demographic constraint of the pair. Subsequently, we randomly selected a (different) second school from the same restricted set, ensuring it satisfies the demographic constraint of the pair and has a characteristic value that is "different enough" from the first school in the pair.

"Different enough" by school characteristic is defined as follows:

- **Baseline test scores:** Different by at least 0.33σ in the average baseline test score means of incoming students. We use an average of mean Math and mean ELA test scores for each school.
- **College enrollment rates:** Different by at least 5pp.
- **Regents VA:** Different by at least 0.3σ in the average Regents VA. We use an average of Regents Algebra VA and Regents ELA VA for each school.
- **Safety:** Different by at least 5 pp.
- **AP stem classes:** Discrete difference (at least 1 more/less class).

Sometimes these restrictions yield an empty set, so not all pairs have two schools, meaning not all pairs are valid. However, most district-questions have three or four valid pairs. To randomly assign each student a valid pair for each question, we use their district of residence. The randomization probability is uniform across valid pairs within each district-question.

C.4.6 Selection of schools to compare within borough: Information question (Q10, version 2)

We measure information about schools by asking to compare one school to the borough distribution of the following school characteristics: baseline test scores, college enrollment rates, Regents VA, college VA, language and AP STEM classes.

For each characteristic (question), we selected four schools per district to include all combinations of high and low white share schools that are above or below the median characteristic value. The median value is calculated based on the borough median.

We selected schools at random among the ones in the district choice set, further restricting to 1) schools in the same borough, and 2) schools ranked by at least 5% of students in the district. This reduces the choice set for each district to 55 schools per district, on average. A school in this subset is ranked, on average, by 11% of students residing in the district.

If the intersection of high-white and above (below) median characteristic returned an empty set, we selected the school with the highest share of white students, conditional on being above (below) the median characteristic. If this also resulted in an empty set, we chose the school with the highest (lowest) value of the characteristic, conditional on being a high white school.

Similarly, if the intersection of non-high-white and above (below) median characteristic returned an empty set, we selected the school with the lowest share of white students, conditional on being above (below) median characteristic. If this also returned an empty set, we selected the school with the highest (lowest) value of the characteristic, conditional on being a non-high white school.

While defining above or below the median for most school characteristics in the borough is trivial, further clarification is needed for how we determine above and below median baseline scores and Regents VA. We consider a school to be above (below) median baseline scores if it is above (below) the median for both average Math and average ELA 7th grade test scores. Similarly, we classify a school as above (below) the median Regents VA if it is above (below) the median for both Regents Algebra 1 VA and Regents ELA VA.

C.4.7 Selection of school characteristics for vignette experiment: Racial preferences question (Q17 and Q18)

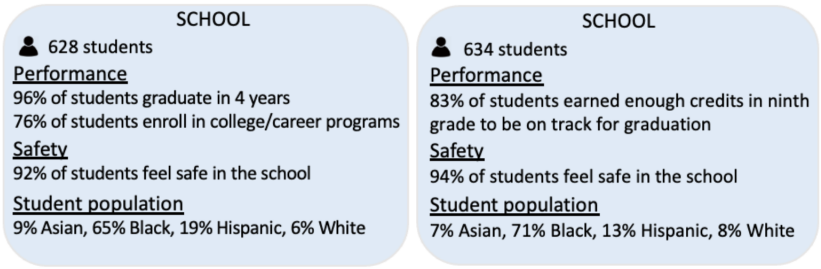
The vignette experiment is described in Section 3.2.1. The experiment includes 24 possible vignettes, also referred to as school cards. The first set consists of 16 school cards in which three school characteristics vary, which we refer to as the 3-factor list": academics (2 levels), safety (2 levels), and racial composition (4 levels). The second set consists of 8 school cards in which only two school characteristics vary, which we refer to as the 2-factor list": safety (2 levels) and racial composition (4 levels). Academics and safety each take either a high or low rating. Student demographics take one of four values: 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. Average school characteristics are reported in Table A2. Examples of the 2- and 3-factor cards are shown in Figure C4.

For Question 17, each student was randomly shown either two schools drawn from the 3-factor list or two schools drawn from the 2-factor list. For Question 18, which asks students to rank three schools on a relative scale, each student was randomly shown either two sets of three schools drawn from the 3-factor list or two sets of three schools drawn from the 2-factor list, with schools drawn without replacement within each set.

Finally, we randomized at the student level whether students received the 3-factor or 2-factor version of the vignette questions. Sixty percent of students were assigned to the 3-factor version.

Figure C4: School Cards for Vignette Experiment



Notes: This figure displays an example of two cards used in the vignette experiment. The left card displays precise academic information (Treatment 1, received by around 60% of the experiment participants). The right card shows imprecise academic information (Treatment 2, received by around 40% of the experiment participants).