Blueprint Labs

Discussion Paper #2024.05

Overcoming Racial Gaps in School Preferences: The Role of Peer Diversity in School Choice

Viola Corradini
Clémence Idoux

June 26, 2024

MIT Department of Economics
77 Massachusetts Avenue, Bldg. E53-390
Cambridge, MA 02139

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

Overcoming Racial Gaps in School Preferences: The Role of Peer Diversity in School Choice *

Viola Corradini† Clemence Idoux‡

This version: June 26, 2024

Abstract

Differences in school choice by race contribute to school segregation and unequal access to effective schools. Conditional on test score and district of residence, Black and Hispanic families consistently choose schools with fewer white and Asian students, lower average achievement, and lower value-added. This paper combines unique survey data and administrative data from New York City to identify the determinants of racial disparities in school choice and shows that attending a more diverse middle school can mitigate racial choice gaps. A post-application survey of guardians of high school applicants reveals that information gaps and homophily in school preferences explain cross-race differences in choice. In turn, instrumental variable estimates show that middle school students exposed to more diverse peers apply to and enroll in high schools that are also more diverse. These effects are consistent across racial groups, particularly benefiting Black and Hispanic students who enroll in higher value-added high schools. Notably, changes in application patterns due to exposure to diverse middle school peers appear driven by changes in the set of known school options and an increased preference for peer diversity.

*We are thankful to the New York City Department of Education's Enrollment Research and Policy office for graciously sharing data, and to participants of the MIT behavioral and labor lunch seminars for helpful comments. Thanks to Eryn Heying, Talia Gerstle, and Jim Shen for dependable administrative support, and to Erika Trevino and Amanda Schmidt for excellent research assistance. This paper reports on research conducted under data-use agreements between MIT, the project's principal investigators, and the New York City Department of Education. This research was made possible by grants from the George and Obie Shultz Fund at MIT and the Wharton Behavioral Lab at the University of Pennsylvania.

†Massachusetts Institute of Technology. Email: vcor@mit.edu

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

1 Introduction

In large urban school districts, school choice is often offered as a pathway to accessing better educational opportunities without the need for physical relocation. Breaking the connection between residential segregation and schooling could yield similar results to relocating to low-poverty neighborhoods during childhood, which research has shown to have long-term positive effects (Chetty and Hendren, 2018; Chetty et al., 2016b,a). Indeed, attending schools with more affluent peers is potentially one of the key drivers behind the effects of early exposure to a wealthier social environment on social mobility. Besides differences in instructional quality across neighborhoods, social interactions with school peers may play a pivotal role in learning, engagement with information, and subsequent decision-making (Conley and Udry, 2010; Cai et al., 2015; Campos, 2023; Golub and Sadler, 2016; Sacerdote, 2011, 2001; Epple and Romano, 2011).

Recent research on urban school integration, however, suggests that opting out of neighborhood assignments to attend more integrated schools does not increase academic achievement among disadvantaged students (Angrist et al., 2022). Despite limited direct academic benefits, in-school interactions with more affluent peers might influence long-term outcomes by shaping future behaviors and decision-making. This paper focuses on this particular avenue and examines how early-grade school diversity impacts subsequent educational choices. This research question gains greater significance in light of observed disparities in education choices across socio-economic status: disadvantaged families often choose lower-quality schools than their more privileged counterparts, even when afforded equal options (Hoxby and Turner, 2015; Chetty et al., 2020, 2023; Carlana et al., 2022). These differences in school choices might contribute to enduring achievement gaps and school segregation (Cohen, 2021; Laverde, 2020; Idoux, 2021).

In this paper, we combine novel survey data with administrative records from New York City (NYC) to examine how exposure to diverse peers in early grades influences subsequent school choices and reduces racial disparities in these decisions. NYC provides an ideal context for investigating the significance of peer effects in school choices: the city offers a wide array of school choices, but minority families apply to lower-quality high schools, as measured by value-added. We find that middle school diversity plays a pivotal role in shaping high school choices. Black and Hispanic students attending predominantly white and Asian middle schools select high schools that closely resemble the schools their white and Asian peers choose and have higher value-added on average. To understand the underlying mechanisms

of these peer effects, we conduct a comprehensive post-application survey of parents and guardians of high school applicants that explores the determinants of their school choices. By linking these survey responses to administrative records and applications to both middle and high schools, we show that exposure to diverse middle school peers alters high school choices by changing the set of known school options and increasing the preference for peer diversity.

We first document the determinants of racial gaps in choices by surveying 3,000 parents and guardians of high-school applicants during the 2022-2023 application cycle.[1] This survey delves into various factors that could influence high school choice: sources of information, significance of different school attributes, awareness of school offerings, perceptions of academic performance and admission probabilities to competitive schools, aspirations for higher education, and perceptions about discrimination. The survey also includes a vignette study to separately estimate preferences for various school characteristics, such as high-achieving peers, school safety, and racial composition of the student body.

Survey findings indicate that across all racial and socio-economic groups, families prioritize similar school attributes and are equally misinformed about these attributes. The most valued characteristics are school safety, academic progress, and college enrollment and graduation rates. Nonetheless, questions testing parents' accuracy of information about school attributes reveal that everyone is similarly biased in their assessment of these school features. We do not find racial differences in higher education aspirations, nor in perceptions about relative academic performance or beliefs about admission chances at competitive programs.

On the other hand, survey responses uncover significant racial disparities in the set of schools families are aware of and a preference for schools with certain racial compositions, all else being equal. Controlling for district of residence and student baseline achievement, Black and Hispanic households know on average fewer schools and are less likely to know about majority white and Asian schools and high value-added schools than their white and Asian counterparts. In parallel, results from the vignette study reveal a marked preference for majority white and Asian schools among white and Asian respondents and a small preference for racially-balanced schools among Black and Hispanic respondents. These preferences over school demographic make-up persist among respondents who observe a precise signal of student academic performance at the school, suggesting that these preferences are not entirely due to statistical discrimination.

The survey results indicate that racial disparities in school choice primarily arise from

[1] The survey is conducted after the submission of applications and prior to the release of high school offers.

differences in the sets of schools considered and preferences for schools' racial compositions. This suggests that a more diverse middle school experience might reduce these choice disparities. Firstly, having access to a diverse network of parents could lessen information imbalances, as survey participants emphasized the significant role of interactions with other parents and middle school staff in informing their high school choices. Secondly, prior engagement with diverse peers might change preferences for interactions with different demographic groups (Rao, 2019; Lowe, 2021; Carrell et al., 2019). In the second part of the paper, we examine the extent to which middle school diversity might reduce racial gaps in high school choice by reducing both information disparity and homophily in peer preferences.

Using longitudinal administrative records that follow students and their school choices over their entire school career, we estimate the causal impact of attending a more diverse middle school on subsequent high school choices. To tackle the problem of selection bias, we leverage the randomness embedded in the NYC school assignment mechanism. Conditional on an applicant's preferences and school priorities, the NYC choice algorithm randomizes seat assignments, thereby manipulating the middle school peer racial make-up independently of potential outcomes. The estimation strategy that exploits this variation builds on the propensity score and instrumental variables methods developed in Abdulkadiroğlu et al. (2017b), Abdulkadiroglu et al. (2022), and Angrist et al. (2022). Specifically, we extend the method of Angrist et al. (2022) to the case where the endogenous variable is a function of the schools of enrollment and observed covariates of all the students.

The instrumental variable (IV) estimates show that attending a more diverse middle school significantly affects high school choices and enrollment. Based on our IV estimates, Black and Hispanic students who attend majority white and Asian middle schools choose high schools which enroll 6.6 p.p. more white and Asian students and have, on average, 0.12 SD higher value-added. This corresponds to a reduction of the racial gaps in racial make-up and value-added of preferred school choices of approximately 60% and 70% respectively. High school choices of white and Asian students are less affected by middle school peer diversity: enrollment in a majority Black and Hispanic middle school increases the Black and Hispanic peer share of the average high school choice by 2.9 p.p and has no effect on chosen value-added. These shifts in high school choices significantly influence the schools that applicants are offered in the centralized match. Attending a majority white and Asian middle school leads Black and Hispanic students to enroll in high schools with 5.7 pp. more white and Asian students and a 0.09 higher value-added.

Furthermore, our estimates suggest that attending a more diverse middle school narrows

racial gaps in school choice by addressing the two key factors that contribute to these gaps: information frictions and homophily. Black and Hispanic families whose child attends a middle school that has a higher share of white and Asian students become aware of a broader array of high schools, particularly those with high achievement levels and high value-added. Additionally, attending a more diverse middle school attenuates homophily across all demographic groups. In contrast, IV estimates of middle school peer effects on achievement are not statistically significant, which suggests that the observed changes in application patterns are not due to an increased probability of admission to the more selective and coveted schools.

This paper builds on research that considers how interactions with peers of different backgrounds may impact social attitudes and beliefs (Corno et al., 2019; Boisjoly et al., 2006; Carrell et al., 2019; Rao, 2019), showing how school integration affects preferences for contact with other races in a real, high-stakes setting. It also contributes to the literature on frictions in school choice (Kapor et al., 2020; Arteaga et al., 2021; Ainsworth et al., 2022) and their unequal impact by socio-economic status (Hastings and Weinstein, 2008; Hoxby and Turner, 2013, 2015; Allende et al., 2019; Pathak and Sönmez, 2008). Our study provides novel insights by directly documenting the existence of information frictions and their role in shaping inequality in access to high-quality schools. Thanks to a unique combination of survey data and administrative records, we also show how peer networks reduce these frictions. Moreover, our findings suggest that biases in beliefs and sophistication may play a lesser role in explaining racial disparities in school applications.

Finally, our findings highlight the importance of path-dependence in school choice. As such, we speak to the literature on the impact of school integration reforms (Idoux, 2021; Laverde, 2020; Bjerre-Nielsen and Gandil, 2020), uncovering a potential dynamic effect mostly overlooked so far.[2] If exposure to more diverse peers is important in shaping student preferences, reducing school segregation in earlier school grades could lower school segregation in later grades through changes in demand for schools.

[2]Hahm and Park (2022) considers dynamic effects of integration reforms through changes in test scores that affect the probability of admissions.

2 Institutional setting

2.1 NYC school assignment system and school segregation

Enrollment in NYC public schools is determined by a centralized school assignment system at the entry grade of each school level. To enroll in pre-kindergarten, kindergarten, sixth grade and ninth grade, students and their families must submit applications through a centralized admission system run by NYC Public Schools (NYCPS). The assignment process unfolds similarly for each entry grade. Applicants are asked to rank academic programs by order of preference.[3] Academic programs also rank applicants, based on priority rules announced before families submit their school preferences. Finally, the centralized admission system combines the information and makes a single school offer to each applicant using the deferred acceptance (DA) algorithm.

To support families in the application process, NYCPS provides a physical admission guide and access to a personalized website. Each personalized website only includes schools to which the applicant is eligible. The website is comprised of an information page about each school, which includes a list of offered programs, courses, and extracurricular activities; the performance of enrolled students on standardized tests; admission priorities and selection criteria for each of its programs; the number of applicants per seat and the priority of the last admitted applicant in the prior year. NYCPS also issues annual school reports that list enrolled student demographics, teacher characteristics, and statistics about student performance and school environment. During the application cycle for enrollment in 2023-2024, applicants had access to their random lottery number on their application profile for the first time.

Each academic program ranks applicants using a set of eligibility and admission criteria based on residential location and academic achievement. Geographic eligibility and admission criteria are more stringent at lower grade levels. At the elementary level, 85% of schools only admit students in their school zone, and the remaining 15% non-zoned schools still give priority to students in their zone. NYC middle schools are intended to serve students residing in their local district, with 83% of middle school programs having zone or district eligibility requirements across the 32 districts.[4] On the contrary, high schools are open to all students in the city, with only approximately 39% of schools giving priority to students

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

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

residing in their borough or zone. Finally, high school and middle school programs rely on academic admission criteria to the same extent: approximately a third of these programs rank individual students based on prior grades, auditions, essays, and behavioral measures, in addition to the eight highly selective specialized high schools.[5]

In line with the higher importance of geographic priorities in earlier school grades, racial segregation across schools is also higher in elementary and middle schools. Appendix Figure B1 compares overexposure to Black and Hispanic peers for students of different races and grades. Across all grade levels, Black and Hispanic students attend schools which enroll disproportionately more Black and Hispanic students than their representation in the city's student population. For example, on average, Black and Hispanic students attend high schools where the proportion of Black and Hispanic students is 11 p.p. higher than the city's average of 68%. In contrast, white and Asian students typically go to high schools with 23 p.p. fewer Black and Hispanic peers than the city average. Moreover, the overexposure of Black and Hispanic students to Black and Hispanic peers is more marked in elementary school than in middle school and high school. Black and Hispanic students attend elementary schools with 28 p.p. more Black and Hispanic students than the elementary school population, middle schools with 25 p.p. more Black and Hispanic students than the middle school population, and high schools with 23 p.p. more Black and Hispanic students than the high school population.

2.2 Race differences in school choice

While school-side factors may contribute to segregation through admission and eligibility criteria, demand-side factors are as important. Past research has documented that school choices differ along several attributes by socio-economic status and ethnicity, with students from poorer families typically applying to schools with lower outcomes in terms of test scores and lower inputs in terms of quality and overall resources (Carlana et al., 2022; Laverde, 2020; Allende et al., 2019; Abdulkadiroğlu et al., 2017a). In part because of these differences in school choice, segregation by race and income remains high (Laverde, 2020; Idoux, 2021), and access to school quality and resources often differs by socio-economic group (Allende et al., 2019).

In Figure 1, we document that preferences for school attribute vary by race even for students with similar baseline achievement. The figure compares mean characteristics of the

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

high school listed first by applicants of different races as a function of their 7th grade test scores.[6] Across all racial groups, higher test score applicants prefer schools that enroll more white and Asian students and students with higher baseline achievement. These schools are also more likely to screen applicants on academic achievement and have higher math value-added on average.[7] Nonetheless, for any given level of baseline achievement, white and Asian applicants favor these characteristics more than Black and Hispanic applicants. Panel A of the figure shows that white and Asian students' first-choice schools have on average 20 p.p. more white and Asian students and enroll students with 0.25 standard deviations higher 8th grade test scores than Black and Hispanic students' first choices. Similarly, Panel B shows that white and Asian students are 10 p.p more likely than Black and Hispanic students to rank as their first choice a school that screens on academic performance, and they favor schools with 0.30 standard deviations higher math value-added on average.

These racial gaps in preferences for school attributes are approximately constant throughout the test score distribution and are not explained by differences in residential locations: controlling for district of residence explains half of the gap in the share of white and Asian students (conditional on test scores), but only 20% of the gap in value-added, 25% of the differences in peer baseline math achievement, and 0% of the difference in probability of applying to screened programs. As a result, the median Black or Hispanic applicant in the baseline test score distribution is as likely to apply to a screened high school program as their first choice as a white or Asian applicant whose test scores is worse by 0.47 standard deviations. This suggests that race plays a different role than test scores and geographic residence in school choice.

3 Conceptual framework

To shed light on the potential underlying causes of racial differences in school choice, we introduce a simple school choice framework where applicants choose which high schools to apply to solving a portfolio choice problem as in Chade and Smith (2006). Specifically, each applicant, indexed by i, chooses a ranked-ordered list (ROL) of schools $R \in \mathcal{R}_i$, where $\mathcal{R}_i$ comprises the sets resulting from all the k-permutations of A_i, the set of schools the

[6] To reflect the information that 8th graders had access to at the time of their high school application, school characteristics are computed on the 9th grade cohort enrolled in each school at the time of application.

[7] Schools' math value-added are estimated using an OLS regression of student Regents Math scores on school fixed effects controlling for demographics, baseline test scores, and student assignment risk to the school as in **?**.

applicant is aware of, which includes her outside option outside the traditional public sector denoted by school 0.[8] Each ROL can be mapped to a lottery over high schools whose weights depend both on the ordering of schools in the list and on applicant beliefs about admission probabilities.[9] Hence, applicants choose their ROL to maximize their expected utility, which depends on their expected utility for enrolling in any given school, the lottery over schools induced by the ROL and their beliefs about admission probabilities, and the cost of submitting the ROL:

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

The expected utility that student i gets from attending school s, $E[u_{is}(\theta_i, X_{is})|\mathcal{I}_i]$, depends on the student's preferences for the school attributes X_{is}, which also include distance from the school, parameterized by the vector of preferences θ_i. Students may hold imperfect knowledge about school attributes and form expectations about u_{is} according to their (potentially inaccurate) beliefs. I_i denotes the information set about X_{is} available to student i at the time of application. In addition, each applicant has a utility of u_{i0} for her outside option outside the traditional public sector.

The subjective probability of assignment to school s, $p_{is}(R, \widehat{q}_i)$, depends separately on the choice of R and on the subjective belief of the likelihood of admission at every school, $\widehat{q}_i$. A property of deferred acceptance is that applicants' admission probabilities at programs are independent of their rank-order lists. Assuming that applicants understand this property, applicant subjective beliefs about admission probability depend only on the student's assessment of her relative ranking in the pool of applicants at school s. This, in turn, depends on the student knowledge of school admission rules, of demand for the program among other students, and of her relative ranking in terms of priority and test scores.

Finally, applicant i incurs a cost $C_i(R)$ when forming her ROL. $C_i(R)$ can be interpreted as capturing any psychological or monetary cost that a student might face when forming her list, given her information set, outside option and beliefs. For instance, listing highly-selective programs may induce a psychological cost when students anticipate being disappointed if they are not granted admission. This cost is likely to be small but rationalizes applicants not including in their lists programs for which their admission chances are slim and submitting short lists if they are almost certain of being granted admission to one of their top choices.

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

[9] For simplicity, we assume that every applicant is guaranteed admission at her outside option, $q_{i0} = 1$

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

1. *Differences in preferences* - $u_{is}(\theta_i, X_{is})$. Applicants may put different weight on different school features, even when these attributes are perfectly observed.[10] For instance, experimental evidence from Hailey (2022) reveals that parents tend to prefer schools enrolling students of similar races or ethnicity.

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

   (a) *Differences in awareness sets* - A_i. Due to search costs and cognitive overload, applicants can only know about a subset of the 400+ high schools in the city.

   (b) *Differences in beliefs about school attributes* due to differences in information accuracy or information processing - $E[.|I_i]$. In line with the existence of these information frictions, a few existing studies look at school choice responses to information disclosure about school attributes (Ainsworth et al., 2022; Bergman et al., 2020; Andrabi et al., 2017; Allende et al., 2019; Campos, 2023).

   Awareness sets and the extent and nature of information frictions about attributes of any school may vary across demographic groups.

3. *Differences in perceived probabilities of admissions* - $\widehat{q}_i$. Applicants may differ in their probability of admission at each school, as well as their subjective belief about this probability. Admission probabilities differ across students due to differences in priorities and test scores. Nonetheless, holding fixed these attributes, subjective belief about admission probability may still differ if applicants do not hold rational expectations but hold biased beliefs instead.[11] Applicants from different socio-economic and racial background might hold different subjective beliefs about admission chances because they differ in their degree of optimism and confidence in their relative ability.[12]

[10] Applicants might also differ in the utility they would derive from enrolling in their outside option. Nonetheless, differences in outside options are unlikely to affect applicants' first choices.

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

[12] A large literature documents confidence gaps across gender and socio-economic status. In the context of racial differences, Corno et al. (2019) find that Black students assigned to racially mixed rooms were less likely to over-estimate the GPA of their white peers compared to Black students assigned to classrooms with a majority of black peers, a finding that they attribute to improvements in Black students' self-image and a reduction in stereotype threat. Similar differences in confidence and mis-perceptions of one's relative ability might also bias perceived chances of admission in competitive programs.

4 Data

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

4.1 Administrative data

We use administrative data to measure key school and student attributes, to document differences in application patterns by race, and to estimate the causal effect of middle school peer diversity on high school choices. This data covers all students who either enrolled in or applied to a NYC public middle school or high school through the centralized school matches. Our sample focuses on applicants seeking 6th grade seats in traditional public middle schools for enrollment in 2015 to 2020 and who three years later (2018-2023) apply for a 9th grade seat in traditional public high school within NYC. Applicants who only apply to NYC specialized (exam) and charter schools are omitted from the applicant file.

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

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.

From the administrative data, we are able to measure key middle and high schools characteristics:

- High school quality: we measure it using value-added models (VAM) which capture the contribution of schools to student achievement. Our main measure of achievement

is Regents math scores. In particular, we adopt a recent methodological improvement in the school VAM literature introduced by Angrist et al. (2021) and referred to as *Risk-controlled value-added (RC VA)*. [13]

- Measures of school student-body composition: using enrollment and demographic data, we measure the share of enrolled students of each race or ethnicity and average enrolled student baseline achievement using the average 7th grade Math standardized test score of enrolled students. These school statistics vary by year.

- Measures of high school selectivity and popularity: we construct a dummy indicating high schools that have at least one program screening students on the basis of test scores, audition, or other ability assessments. We also construct a measure of popularity using the ratio of rejected to accepted applicants at a school. [14]

On average, students rank 8 high school programs on their list. Top choices tend to be of higher quality, more selective, and more popular than the average school in the city. On average, applicants' top 3 program choices have a Regents math VA which is 0.8 standard deviations higher than the average school in the city. In addition, 70% of applicants rank at least one screened program among their top three choices. 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.

4.2 Survey data

We conducted a post-application survey of guardians of 9th grade applicants in partnership with the NYC Department of Education. The survey was conducted from February 17th to March 6th, 2023, after applicants had submitted their high school applications but before they had received their match offers. The survey was sent electronically to the email addresses provided during the high school application. Respondents could answer in English, Spanish, or Chinese. Upon completion of the survey, participants that had answered at least one

[13] The main difference with respect to standard methods is the inclusion of additional controls for a richer set of student covariates coming from student applications and priority status assigned by schools at the time the student applied to high school.

[14] For schools with more than one program, we construct a school-level measure of popularity by taking a weighted average of program-level ratios of rejected to accepted applicants, with weights proportional to the seat capacity of the programs.

survey question were sent a 10-dollar amazon gift card. Only parents and guardians were permitted to respond to the survey.[15]

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

The survey examined various dimensions of the choice process for families, including sources of information, essential school characteristics, knowledge of school options and their features, perceptions of admission probabilities and their influence on choice, perceptions of discrimination and its impact on decision-making, and educational aspirations. Additionally, the survey conducted a vignette experiment, described in more details in section 5.1.2, which aimed at disentangling families' relative preferences for different school characteristics and uncovering potential statistical discrimination. This comprehensive approach enables us to systematically document differences in the three main potential drivers of choice differences outlined in our conceptual framework: differences in preferences, information about schools, and beliefs about admissions probabilities among different racial and socio-economic groups. The complete survey, as presented to participants, together with detailed information on the construction and randomization of the questions, is available in Online Appendix C.

5 Family survey results

In this section, we use survey answers to comprehensively investigate the reasons behind observed racial gaps in school choice. Motivated by the framework presented in section 3, we consider three main channels: differences in preference for school characteristics in section 5.1, differences in information about school existence and attributes in section 5.2, and differences in beliefs about likelihood of admissions in section 5.3.

[15] 75% of the survey respondents reported that parents and guardians played an essential role in their student's high school selection.

5.1 Racial gaps in preferences for school characteristics and peers

The racial disparity in school choices may simply be attributed to different preferences over school attributes across applicant race. We start by examining respondents' stated preference for school attributes and subsequently explore *ceteris paribus* differences in revealed preferences for school attributes using results from a vignette experiment.

5.1.1 Stated preferences

As suggested by Figure 3, respondents from all racial groups prioritize the same school features when selecting schools. Most respondents report the same six school features as most important when selecting a school: safety, academic progress of students at the school, college and graduation rates, commuting time, the number of AP classes offered, and whether their students would feel they belong. Each of these features was mentioned by at least 20% of respondents as one of the three most important attributes of a school. The race differences in the proportion of respondents mentioning each school feature as most important are not statistically significant for the top three school features (safety, academic progress, and college and graduation rates). This is supported by Appendix Figures B3, wherein a majority of families from various racial backgrounds express that their child would thrive in an academically rigorous school environment, yet may not fit as well in school with significant disparity in peer achievement levels.

The emphasis on academic performance of schools by respondents across all racial groups may reflect their high aspirations for their children's future education. As depicted in Figure 5 panel (a), more than half of respondents agree that attending college is crucial for achieving success in life, and 87% express a desire for their children to obtain at least a four-year college degree. Educational aspirations among respondents from different racial groups are similar. When controlling for district of residence and baseline test scores, Black and Hispanic respondents are equally likely compared to white and Asian respondents to desire that their children attend college for at least four years. However, they are 7 percentage points less likely to consider college important for achieving success in life, potentially due to differences in personal experiences or trajectories. This evidence suggests that, overall, families across demographic groups share the same school selection criteria and have comparable aspirations for their children's education.

5.1.2 Revealed preferences: Vignette experiment

To explore the causal influence of different school features on family choice, we conducted a vignette experiment as part of the survey. The experiment consisted of two parts. In the first part, respondents' cardinal preferences for hypothetical schools were elicited by asking them about the likelihood on a scale from 1 to 6 of including two hypothetical schools in their application list. In the second part, respondents' ordinal preferences for hypothetical schools were elicited by asking them to rank two sets of three hypothetical schools.

In both parts of the experiment, all hypothetical schools were described as identical except for their safety rating, academic performance rating, and racial composition. Respondents were also told that their student would have high admission chances at any of these schools. As for safety, hypothetical schools had either high-safety or low-safety ratings. In terms of student demographics, hypothetical schools had either a balanced racial composition representative of the school district, a majority of Black students, a majority of Hispanic students, or a majority of white or Asian students.

The academic performance information provided to participants varied based on the treatment arm they were assigned to: 60% of participants received precise information about the schools' academic performances, while 40% received imprecise information. The precise academic information consisted of the 4-year graduation rate and the college and career program enrollment rate. The imprecise information consisted of the share of students that earned enough credits in 9^{th} grade to be on track for graduation. Participants who received precise information were presented with either a high-performing or low-performing school, whereas those who received imprecise information were always presented with a school with median academic performance. Table C4 outlines the information presented to participants in both the precise and imprecise academic information treatment arms.[16]

The experiment employed a factorial design to randomly combine these characteristics, resulting in 16 unique combinations for the precise-information treatment arm and 8 unique combinations for the imprecise-information treatment arm. The two schools for the first part of the experiment were randomly selected without replacement from these unique combinations. For the second part of the experiment, two sets of three distinct schools were randomly chosen without replacement.[17] Figure 4 shows an example of the vignettes as seen by survey participants.

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

[17]Therefore, respondents may have encountered the same school in at most three instances.

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

$$u_{is} = \alpha_r Z_r(i) + \beta_r X_s \times Z_r(i) + e_{is}$$

.

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

We combine the absolute preferences for schools and relative rankings of schools provided by respondents to estimate their respective weights for different school characteristics.[18] The scale and location of the utility is thus normalized by respondents' likelihood of listing the schools. A respondent indicating a likelihood of a to list a school implies that $u_{is} \in [a - 0.5, a + 0.5]$. The full parameter vector $\theta = (\beta, \sigma)$ is estimated using a Gibbs sampler to maximize the likelihood of observing the responses to both questions.

Table 3 presents the estimates in likelihood units with respect to a racially-balanced, low-safety, and low-performing school. Column (1) shows that school academic and safety ratings are the primary factors that influence families' school choices for all respondents. Holding all else constant, a high academic rating increases utility by 1.4 points, while a high safety rating increases it by 0.7 points. The magnitude of these effects is consistent across racial groups. Nonetheless, contrary to Black and Hispanic respondents, white and Asian respondents also hold some preference over schools' demographic make-up.[19] White and Asian respondents are 0.27 points more likely to list a majority white and Asian school, and respectively 0.28 and 0.44 points less likely to list a majority Hispanic or Black school compared to a racially-balanced school. Conversely, there is no evidence to suggest that Black

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

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

and Hispanic respondents take into account the demographics of schools, as the coefficients on the racial make-up of schools are not statistically significant.

This preference over the school's racial composition is partly due to statistical discrimination, whereby respondents infer the academic performance of a school based on its demographic make-up. The comparison between columns (1) and (2) of Table 3 suggests that respondents who receive less precise information about schools' academic performance are more influenced by the schools' racial composition in their rankings.

This effect is observed across all racial groups. In the case of Black and Hispanic respondents, the coefficients pertaining to the school's demographic make-up are only significant when they receive imprecise information about the schools' academic performances. In such instances, they are 0.22 points more likely to list racially-balanced schools than other types of schools. Similarly, white and Asian respondents exhibit stronger preferences for the racial composition of their peers when they receive imprecise information about the schools' academic performance. They demonstrate a heightened preference for majority white and Asian schools, as indicated by the corresponding coefficient increasing from 0.27 to 0.36. Additionally, they show a stronger aversion towards majority Hispanic or Black schools, with the corresponding coefficients decreasing from -0.22 to -0.59 and from -0.44 to -0.70, respectively.

Overall, the vignette experiment suggests that the racial gaps in choice are unlikely to be due to differences in preferences for peer achievement or safety, in line with the evidence shown in the previous subsection, but may be due to homophily. White and Asian respondents show a consistent preference for schools that enroll more white and Asian students, while Black and Hispanic respondents are less likely to choose majority white and Asian schools when they have more imprecise information about schools' academic performances. As suggested by Appendix Figure B4, this racial preference for peers may stem from concerns regarding racial discrimination by students' teachers and peers, which are more common among Asian, Black, and Hispanic families compared to white families. Indeed, among the 23% of respondents who reported that their students might face discrimination from their teachers or peers, 70% mentioned that these concerns influenced their high school choices.

5.2 Racial gaps in information about schools

Disparities in information about schools could also result in differences in application behavior. We evaluate two distinct aspects of families' information about schools: schools they are aware of and accuracy of their information about specific school features.

5.2.1 Awareness sets

Because New York City has more than 400 high schools, it is unlikely that families have heard about all of them. We call the set of schools that a family is aware of "awareness sets." Racial differences in choices may simply stem from differences in awareness sets. To explore this hypothesis, we asked respondents to indicate which schools they had heard of from a list of ten schools. These schools were selected to be relatively close to the respondent's home, popular, and diverse in characteristics. The specific schools shown to each respondent were randomized based on their district of residence.

Panel (a) of Figure 6 presents the share of schools respondents were aware of among the schools presented to them, categorized by different types of schools. Panel (b) shows racial differences controlling for district of residence and baseline test scores. On average, respondents from all racial backgrounds appeared to be familiar with approximately one third of the schools presented to them. While there were no notable differences in the total number of schools respondents were aware of, Figure 6 reveals significant racial disparities in the types of schools respondents were familiar with. Compared to their white and Asian counterparts, Black and Hispanic families appear to be less aware of majority white and Asian schools, high-quality and high-performing schools, as indicated by having high VAM or high college enrollment and graduation rates. Black and Hispanic respondents are aware of 4.3 percentage points fewer majority white and Asian schools, 4.1 percentage points fewer high-VAM schools and of 3.0 percentage points fewer schools with high college enrollment and graduation rates. In contrast, Black and Hispanic respondents are aware of 4.3 percentage points more high Black and Hispanic schools, 4.8 percentage points more low-VAM schools.

A potential reason why Black and Hispanic families are less likely to know of high-quality school options is that they rely on different sources to gather information about schools. In Figure B6 we document the difference in the probability that a given information source is listed among the most used by respondents' race. Black and Hispanic families are 19 percentage points less likely to rely on networks of family and friends to collect information about schools compared to white respondents, and 5.6 percentage points less likely to rely on parent networks within their student's middle school. They also appear less likely to use information sources that are costlier and more time consuming, such as attending individual high school information sessions and browsing on websites different from the official NYCPS website. On the contrary, they are more likely to rely on institutional resources provided by their student's middle school, such as guidance counselors, other school staff, and middle school information sessions, and equally likely to use NYCPS online resources.

5.2.2 Accuracy of information

Even if families have heard about a school, they may hold inaccurate beliefs about it. To measure inaccuracy of information about school characteristics, and how misinformation varies across demographic groups, we ask two sets of questions. In the first, we ask survey respondents to compare two schools on a specific aspect, such as which school has higher graduation rates. In the second set of questions, we ask them to compare a school to all other schools in the borough, for instance by asking which quartile of the distribution of graduation rates that school belongs to.

Figure **??** suggests that families have somewhat accurate information about school features important to them when selecting a school. The left chart in panel (a) shows that respondents are more likely, on average, to correctly rank the two schools they are presented based on school safety, value-added college and graduation rates, commuting distance, and peer preparedness than if they had guessed randomly. Conversely, applicants are not well-informed about which schools offer more AP classes. Respondents' most precise information is about commuting time, with respondents being 18.4 percentage points more likely to accurately guess the school with the shortest commuting time compared to guessing randomly. Information about value-added is more limited, with respondents only answering correctly 9.1 percentage points more often than if they had guessed randomly.

The left chart in panel (b) provides additional evidence that families possess some knowledge about these school features: respondents' quartile rankings of schools are positively correlated with actual quartile rankings of schools. Families' information appears to be the most precise about school location, followed by the college and graduation rates. In summary, this evidence is consistent with the view that families are better informed about aspects that are more easily observed, such as where the school is located, but are less well informed about value-added, which is arguably harder to observe.

The schools selected for these questions are well-known in the applicant's neighborhood, even though they may not necessarily belong to the respondent's awareness set.[20] Even when restricting the sample to the schools that are definitely in the respondent's awareness set, information accuracy improves only for easily-observed attributes, such as commuting time and peer achievement levels. Information accuracy does not improve for school value-added. We show this in appendix Figure B5. Panel (a) reports the raw average probability

[20]The reason why the survey does not restrict to schools in the respondent's awareness set is that we do not want to condition on an outcome but rather we want to capture how families form beliefs based on cues such as the school name, borough, and district, which is basic information that is easily accessible from browsing the school directory.

of answering correctly in the pairwise comparison when using all questions (first bar), when restricting to questions in which one school is certainly known (second bar), and when restricting to questions in which both schools are certainly known (third bar).[21] Panel (b) reports the rank-rank correlation coefficient, restricting the sample to schools we are certain the applicant knows in the second bar.

While families appear substantially misinformed about school characteristics such as value-added, differences in information accuracy across race are unlikely to drive inequalities in application behavior. The right chart in panel (a) shows that the race difference in the probability of guessing correctly in pairwise school comparisons is never significantly different from zero. The right chart in panel (b) instead shows that Black and Hispanic applicants' beliefs about peer achievement and college rates are less correlated with the actual school rankings, and this is driven by their lower propensity to select extreme answers. However, there is no significant difference in belief accuracy about school value-added across respondent race. In appendix Table A3 we pool the answers to the two sets of questions by regressing an indicator for answers that are approximately correct on respondent race, baseline test score, and district of residence, finding no significant difference in information accuracy by race.[22]

5.3 Racial gaps in admission beliefs

Finally, differences in beliefs about admission probabilities at competitive schools might contribute to racial gaps in application. When deciding where to apply, applicants may exclude programs from their lists if they perceive their chances of admission as being too low.[23] The survey provides direct evidence supporting this possibility. According to column (1) of Table 4, 16% of survey respondents indicated that they did not apply to their "dream program", i.e. their preferred program if guaranteed admission.[24] Nevertheless, we do not find evidence that racial differences in pessimism about admission chances to most preferred

[21] We can say that a respondent certainly knows a school if the school appears in her awareness set question and she selects it or if the school was ranked in her high school application.

[22] The indicator takes a value of 1 if the answer is exactly correct, which in the ranking question means correctly guessing the position of the school in terms of quartiles of the within-borough distribution of that attribute. It also takes the value of 1 if the answer is approximately correct, meaning the difference in the two school attributes is low, or, for the ranking question, if the respondent ranks the school in the quartile next to the correct answer and the school real position in the distribution is in the quartile half closer to the respondent's answer. More details are provided in the Appendix.

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

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

programs is a channel underlying racial gaps in applications.

Panel A of Table 4 shows that there is no significant difference in the likelihood of Black and Hispanic applicants applying to their favorite programs compared to white and Asian applicants. Controlling for the respondent's "dream program", the difference in application rates to the dream school, as reported in column 1, is less than 1.7 percentage points and lacks statistical significance. Additionally, there appears to be no disparity in pessimism about admission chances, as applicants from all racial groups are equally inclined to apply to their favorite program when faced with the same admission probability. Column (2) further indicates that Black and Hispanic applicants are on average as optimistic about their admission chances as white and Asian applicants. Nonetheless, Black and Hispanic applicants' actual admission probabilities are less associated with their beliefs about admission chances than white and Asian applicants. A 1 percentage point increase in admission probability corresponds to a 0.235 percentage point increase in reported admission belief for white and Asian applicants, while it only corresponds to a 0.16 percentage point increase for Black and Hispanic applicants.

As suggested in Panel B of Table 4, the similarity in beliefs about admission chances across racial groups partly arises from similar beliefs about their students' relative performance compared to students attending sought-after schools. While at the bottom of the performance distribution, Black and Hispanic respondents are more optimistic about their kids' relative academic performance. Their higher optimism fades at higher levels of student achievement. For each increase in actual performance by one tercile, Black and Hispanic beliefs about their students' relative performance increase by 0.121 points less than those of white and Asian respondents. Moreover, Column 2 shows no racial differences in beliefs about relative academic performance conditional on student achievement when respondents are asked how their students would compare to other students enrolling in high-demand schools, who are typically higher-achieving. These results suggest that Black and Hispanic families are not less likely to apply to popular programs enrolling high-achieving students due to under-confidence about their admission chances or their student academic ability.

6 Middle school effects on the racial choice gap

Our results so far indicate that families prefer schools enrolling students with similar demographics and that Black and Hispanic families choose lower value-added schools despite caring equally about school quality. Most of the Black and Hispanic shortfall in preferences

over quality is explained by racial gaps in awareness of higher-quality schools, with the remainder possibly explained by differences in preferences over the demographic composition of schools.

Middle schools may contribute to leveling the playing field for families across income and race by offering more equitable access to information and providing a setting where families from diverse backgrounds can interact. First, middle schools serve as a place for parents to share information: 75% of respondents reported discussing high school applications with other parents at their student's middle school at least once, 26% engaged in such discussions more than five times, and over a quarter indicated other parents as one of their most important information sources. Second, middle schools are an institutional source of information about high school applications: they organize information sessions, and school staff may provide guidance to families during the application process.[25] Finally, diversity within earlier grades may attenuate preferences for more homogeneous peers in high schools, which drive part of the differences in application behavior across racial and socio-economic groups.

This discussion motivates us to study the effects of middle school demographics on high school choice. Using the variation arising from the NYC middle school match, we show that Black and Hispanic families randomly assigned to middle schools enrolling more white and Asian students choose whiter and higher-quality high schools as a result. At the end of the section, we use our survey to ask why middle school demographics affect school choices. In addition to information-sharing within peer networks, we also explore whether interaction reduces inter-group prejudice and its effects on confidence and beliefs.

6.1 Correlating peer exposure and school choices

We are interested in understanding how exposure to other-race peers in middle school affects high school choices, as measured by the parameter α in the following regression:

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

C_i is a measure of contact with other-race peers in the middle school where i enrolls, X_i is a vector of controls, and u_i is a regression residual.

In most analyses, we bundle student races in two categories: white and Asian students

[25] 16% of families cite middle school sessions as one of the most important sources of information, while 26% of respondents overall and over 30% of low-income, Black, and Hispanic families rely on middle school staff as one of the main sources of information about high schools.

and Black and Hispanic students. C_i is a measure of contact with other-race peers in the same middle-school grade as student i, that is, with minority peers if i is white or Asian, and white and Asian peers if i is Black or Hispanic. In some specifications, C_i indicates the leave-one-out share of other-race peers in students' middle school class, while in others, it indicates having a majority (above 50%) of other-race middle school peers.

To gain some insight, Figure 2 compares the attributes of applicants' top choices by whether they come from mostly white or mostly minority middle schools. The figure reveals three interesting facts.

First, minority applicants attending majority white and Asian middle schools choose high schools with similar peer baseline achievement and value-added as white and Asian applicants. On the contrary, white and Asian students' preferences for these attributes do not vary much depending on the racial composition of their middle school. Second, the racial composition of first choices varies depending on the race of middle school peers more than other choice attributes. Minority students attending majority white and Asian middle schools choose high schools enrolling 25 p.p. more white and Asian students than other minority students. Similarly, white and Asian students from predominantly minority middle schools tend to select high schools with 15 p.p. fewer white and Asian peers. Third, minority students are consistently less likely to apply to screened programs than white and Asian students, even when they attend a majority white and Asian middle school. Middle school diversity appears to reduce the race gap only for minority students in the top third of the test score distribution. White and Asian students apply to screen programs at the same rate, regardless of the racial mix of their school of origin.

Appendix Table A6 shows the correlation between middle school peer diversity and additional attributes of high school top choices. It reports OLS estimates of α in equation (2), both when C_i indicates having a majority of other-race peers (top rows) and when it indicates the share of other-race peers in middle school (bottom rows). Here we consider the correlation with the average attributes of applicants' top 3 choices, rather than first choices as in Figure 2. In addition to the patterns highlighted in the figures, the table also shows that, for minority students, having more white and Asian middle school peers is associated to choosing more popular schools. On the contrary, white and Asian students with more minority peers choose schools with lower-achieving peers and are marginally less likely to choose less popular schools. These patterns may reflect selection bias due to students with different high school preferences sorting in middle schools with different racial composition. In the next section, we present our instrumental variable approach to identify the causal

effects of attending middle schools with a more diverse set of peers.

6.2 Instrumental variables framework

Our econometric framework identifies the causal effect of exposure to other-race middle school classmates for students for whom classmate diversity is determined in part by random assignment. The tie-breaking in the middle school assignment algorithm in fact generates a research design that identifies causal effects. School offers are a function of applicant preferences and priorities, which we refer to as applicant type θ_i, and the set of tie-breaking variables. Tie-breakers include a common lottery number used by unscreened schools and a set of non-lottery tie-breakers (such as test scores) used by screened schools. This means that school assignment differences for students with the same value of θ_i and proximity to non-lottery cutoffs are due solely to the tie-breaking embedded in the match.

Angrist et al. (2022) shows that the causal effect of any ordered school characteristic, such as peer racial make-up, can be estimated via a 2SLS regression that instruments the enrolled school characteristic with the offered school characteristic and controls for the expected value of the instrument. We adopt a similar method. We instrument the share of other-race peers in the middle school of enrollment with the other-race peer share in the offered school, controlling for the expected offered other-race peer share. The instrument's expected value controls for systematic differences in potential outcomes between applicants who are offered schools with different racial compositions.

However, we adapt this framework to take into account that peer racial make-up is dependent on all students' enrollment, and thus our instrument not only depends on each student's individual offer but also on the full set of offers. To circumvent this issue, we compute the potential school racial make-up which uses students' offer distributions instead of realized offers in our construction of the instrument. The remainder of this section describes the empirical strategy in more detail.

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

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

where $D_i(s)$ indicates an offer at school s. This probability is a function of the applicant type θ_i and indicators for proximity to cutoffs for non-lottery programs, denoted by $\tau_i(\delta_N)$ and determined by a data-driven bandwidth, δ_N. In the large-market theoretical framework

outlined in Abdulkadiroglu et al. (2022), the propensity score $\varphi_s(\theta_i, \tau_i(\delta_N))$ depends only on a few match-determined parameters and is easily tabulated from data on the match.

Next, we define the *potential* leave-one-out share of other-race peers in school s as the share of other-race peers in school s that we should expect before any uncertainty over tiebreakers is resolved:

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

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

Potential other-race peer shares will typically differ from realized other-race peer shares, computed using the set of enrollment decisions $\{E_j(s)\}$,

$$c_i(s) = \frac{\sum_{j \neq i} O_i(j) \cdot E_j(s)}{\sum_{j \neq i} E_j(s)}.$$

The discrepancy originates both from the uncertainty in the match and from imperfect offer compliance, drop-outs, and late-enrollment of students who did not participate in the match.

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

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

.

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

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

As shown in Angrist et al. (2022), conditioning on μ_i ensures instrument validity as:

$$\epsilon_i \perp Z_i | \mu_i$$

. Intuitively, μ_i controls for any variation in offered peer race that is due to applicant type θ_i. Hence, once controlled for μ_i, any remaining variation in offered peer race 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. We also control for local-linear functions of non-lottery-school tie-breakers; these functions employ the bandwidth used to define $\tau_i(\delta_N)$.[26]

The causal effect of interest is an estimate of coefficient β in the 2SLS system:

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

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

. Because β might differ by race, we estimate this system of equations separately by race. First and second stage models control for linear control functions $g_s(\cdot)$, and $h_s(\cdot)$ are linear control functions of the running variables R_{is} at non-lottery programs.[27] Both stage models also include a set of baseline covariates, denoted X_i. [28]

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

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

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

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

.

[26]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.

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

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

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

[28]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.

For all the 2SLS estimations, our 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 more than one other-race peer share value. Appendix Table A1 describes the restrictions applied to construct this experimental sample with greater detail. Columns (2)-(4) in Table 1 compare demographics, other-race shares in middle schools, and high school choices of students in the experimental samples to those of the universe of students observed applying to both middle school and high school in NYC in the study period. While Black and Hispanic students are slightly over-represented in the experimental sample, the sample appears to be quite similar to the population of applicants in column (1).

Appendix Tables A4 and A5 report a set of results meant to validate our research design. The first panel of both tables checks whether differential attrition may lead to selection bias. Virtually all the middle school applicants in our analysis sample are observed enrolling in 6th grade within the public school system, while only 89% of them subsequently apply to enroll in a public high school in NYC. Both tables show that the likelihood of observing these outcomes is unrelated to the majority other-race offer (Table A4) and offered other-race share instruments (Table A5).

A second set of diagnostics evaluates covariate balance. In both tables, Panel B reports coefficients on offer instruments from regressions of covariates on the instruments, with appropriate controls for estimated μ_i and for functions of non-lottery program tiebreakers. For the discrete instrument in Appendix Table A4, the estimates show no statistically significant relationships between the majority other-race offer instrument and baseline covariates. For the continuous instrument in Appendix Table A5, the estimates show small differences in baseline math test scores and subsidized lunch status. Nonetheless, the magnitudes of these differences seem unlikely to lead to substantial omitted variables bias. In any case, all 2SLS estimates are from models that include the baseline covariates listed in the table as controls.

6.3 2SLS estimates on school choices

Black and Hispanic students with higher shares of white and Asian middle school peers apply to high schools that enroll fewer Black students and more white and Asian students. This is documented in Table 5 which 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. The table shows estimates separately for the top three choices in Panel A and for

all the choices in an applicant's list in Panel B.[29] To account for a change in the number of choices, the table also reports the effect on the length of rank order lists.

The estimated effects indicate that attending a middle school with a higher proportion of white and Asian peers significantly influences the overall application profile of Black and Hispanic students. As a result of attending a majority-white and Asian middle school, the first three high school choices of Black and Hispanic applicants have on average 4.8 p.p. fewer Black students and 6.6 p.p. more white and Asian students. This corresponds to an increase in the chosen share of white and Asian students of more than 20%. Similarly, attending a middle school with 10 p.p. more white and Asian students decreases the share of Black students in top school choices by 1 p.p. and increases the share of white and Asian students by 1.3 p.p. The table also shows that attending white and Asian middle schools 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. The magnitude and significance of these effects are similar when considering all the high school choices, although results are larger when looking at the most preferred schools.

Middle school peer diversity also impacts other dimensions of choice. In particular, attending majority-white middle schools (middle schools with 10 p.p. more white students) increases average value-added in top 3 choices by 0.12 SD (0.03 SD). Nonetheless, other-race peers seem to have little to zero effect on the popularity of ranked high schools and the probability of applying to a program that uses screened admission methods. Finally, the list length is unaffected, suggesting that students are changing most of their choices within a fixed-length list rather than adding extra schools. The pattern and magnitude of the effects of exposure to white and Asian peers are similar when estimated independently for Black and Hispanic students in Tables A7 and A8.

The estimated peer effects on high school choice attributes for white and Asian students are reported in Table 6. White and Asian students' high school choices are affected by the racial make-up of their middle schools by a lesser degree. First, we observe much smaller effects on the racial composition of high school choices: attending majority-minority middle schools increases the share of Black students and of Hispanic students in the top 3 choices by 1.3 p.p. but only when focusing on all ranked schools. Similarly, the magnitude of the decrease in the chosen same-race share is about half of what we estimated for Black and

[29] 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.

Hispanic students. Effects are larger when considering all choices rather than the top 3 choices, contrary to what we observed for Black and Hispanic students. Since list length is, if anything, shortened, these estimates suggest that white and Asian students respond to a change in middle school peer diversity by mostly modifying their bottom high school choices, while contact with white and Asian students seems to affect Black and Hispanic students' top choices.

Attending middle schools with larger shares of Black and Hispanic students induces white and Asian students to rank schools with lower-achieving peers. This negative effect, while smaller in magnitude, is comparable to the increase in chosen peer achievement by Black and Hispanic students observed in Table 5. The effects on choice popularity and screening method are non-significant, while the effect on school value-added is negative but mostly insignificant and smaller in magnitude to the positive peer effect found for Black and Hispanic students.

Overall, attending a majority-white and Asian middle school closes respectively about half of and two-thirds of the racial gaps in the school value-added and in the racial composition of school choices, conditional on baseline achievement. A comparison of the 2SLS estimates with the OLS estimates in Table A6 reveals that OLS estimates of peer effects are not extremely biased, especially for Black and Hispanic students. 2SLS estimates of white and Asian peer effects are less than 40% smaller than OLS estimates for the racial composition of school choices and not statistically distinguishable for peer achievement and school value-added.

Table 7 shows that differences in high school choices induced by middle school peer diversity translate into differences in high school offers for all students. Estimated effects on the offered high school's racial make-up and peer achievement are generally larger for Black and Hispanic students than for white and Asian students. Black and Hispanic students who attend a majority-white and Asian middle school receive as a result high school offers with 5.6% more white and Asian students on average. Middle school diversity also has a positive significant effect on the offered high school's value-added of Black and Hispanic students, which increases by 0.09 SD. Nonetheless, this increase in high school peer diversity and value-added partly comes at the cost of a marginally significant increase in the probability of being unmatched by 2.5 percentage points. On the other, white and Asian students are offered schools with similar value-added and peer composition regardless of their middle school of enrollment.

In appendix Table A9 we estimate models that allow for a more granular definition of race and distinguish between the effects of Black and Hispanic peers. These models have two endogenous regressors, one for each race share different from own in the middle school of

enrollment (e.g. Black peer share and Hispanic peer share if the student is white or Asian). The two instruments are the two corresponding potential race shares in the offered middle school. Similar to what we observed for the binary other-race exposure treatment, attending middle schools with students from a particular ethnicity induces students to rank schools with a higher share of students from that ethnicity and a lower share of students from their own race.

In a nutshell, the 2SLS analysis suggests that exposure to more diverse peers in middle school significantly affects high school choices. The estimated effects vary in magnitude for applicants of different races, but all point to a reduction in the share of same-race peers in high school choices and, ultimately, in the high school offer received from the match. Increasing middle school diversity might then be a lever for high school desegregation, mediated by a change in choices. Moreover, attending majority-white and Asian middle schools causes Black and Hispanic students to choose higher-quality schools. The next section discusses potential channels through which middle school peer diversity may impact high school choice.

6.4 Peer effects mechanisms

Why does the race of middle school classmates affect high school choices? In this section we explore three main explanations: peer effects on education achievement, changes in preferences for interacting with other demographic groups, and information-sharing through social networks.

Test scores We first consider other-race peer effects on middle school achievement. Positive peer effects might explain why Black and Hispanic students attending majority-white high schools apply to more selective programs. For instance, students might prefer attending schools with students at their achievement level to avoid mismatch. Moreover, higher test scores increase chances of admission at screened programs.

We measure the effect of middle school diversity on students' achievement using the same 2SLS strategy used to study peer effects on high school choices. Table 8 reports middle school peer race effects on 6th and 7th grade Math and ELA standardized state test scores. We find mostly no effect of other-race peers on the test scores of any students, the only exception is a statistically significant increase in 7th grade math score by 0.09 SD for Black and Hispanic students attending a majority white and Asian middle school. These results are in line with several studies finding small to zero peer effects in achievement when using well-identified

empirical designs (Angrist, 2013).

Racialized preferences and perceptions of discrimination A recent literature finds that contact with individuals from a different ethnicity reduces inter-racial, or more generally, inter-group prejudice (Rao, 2019; Carrell et al., 2019; Lowe, 2021; Corno et al., 2019; Boisjoly et al., 2006). Our vignette experiment allows us to isolate the effects of attending middle schools with a higher share of students from a different race on preferences for the demographic composition of future classmates.

To investigate the effect of middle school diversity on racial preferences for high schools, we re-estimate respondents' preferences in the vignette experiment as a function of whether their student attended a majority white and Asian middle school. These estimates are reported in Table 9. Respondents whose students attended a majority-white and Asian middle school on average prefer hypothetical high schools that are majority white and Asian over hypothetical schools that are racially neutral. This is in contrast with other Black and Hispanic respondents, who tend to prefer racially-neutral schools and other white and Asian respondents who are indifferent between racially-neutral schools and majority white and Asian schools.

One way of interpreting these results is that interaction with other-race families in earlier grades reduces taste-based discrimination. An alternative hypothesis is that it reduces statistical discrimination, i.e., the extent to which household rely on race to make inferences about school academics. The second explanation is likely to play a larger role in this setting, given the smaller importance of pure taste-based discrimination found in the context of this experiment, as discussed in section 5.1.2.

An additional mechanism might be a reduction in perceived discrimination, which we investigate in Table A10. We only report the differential effect of having more middle school peers of different races on perceived discrimination by respondent race or ethnicity. We find mostly null effects, except for Asian students, the group with the highest levels of stated perceived discrimination. Attending majority Black and Hispanic middle schools makes Asian respondents more likely to agree with the statement: "*My student would fit well in a school where the majority of peers are from a different race*" and less likely to report that their high school application choices were influenced by fear of discrimination.

While we need more statistical power to make stronger causality claims, we think that our estimates provide suggestive evidence that past experience might modify preferences for inter-group interaction.

Information In section 5.2 we showed that Black and Hispanic students were significantly less likely to have heard of high-value-added schools and schools enrolling higher achievers and a high share of white and Asian students. Here we study whether attending schools with white and Asian peers might close some of these gaps. We measure the effect of attending majority-white schools on the probability of knowing of different types of schools by estimating the following regression:

$$Know_{is} = \alpha + Minority_i + \beta W_i + \gamma Minority_i \cdot W_i + X_i'\delta + \epsilon_{is} \tag{5}$$

. $Know_{is}$ takes value 1 if respondent i reports having heard of high school s and 0 otherwise, and W_i is a dummy for attending a majority-white and Asian middle school.

The OLS and IV estimates of β in Table 10 are non-significantly different from zero, while estimates of γ are positive and significant, and somewhat larger when using IV. They indicate that, while middle school peer demographics matter little for white and Asian families, attending majority white and Asian schools significantly expands the set of schools known by Black and Hispanic students. Black and Hispanic students are less likely to know schools with high value-added or higher-achieving students, but attending schools with more informed peers entirely closes these gaps. Moreover, the effect of attending a white and Asian middle school on the total number of schools known for Black and Hispanic students appears larger than the race gap, suggesting that white and Asian peers expand Black and Hispanic students' awareness sets rather than simply changing their composition.

Summary In summary, we conclude that the main reason why having more white and Asian middle school peers changes Black and Hispanic high school choices is by reducing information frictions in the form of limited awareness of school options. Attending majority white and Asian middle schools increases the probability of Black and Hispanic students having heard of high schools with higher shares of white and Asian students, higher-achieving peers, and high value-added. In addition, middle school peer racial composition has no effect on student test scores, while some evidence suggests attending a diverse middle school might increase white and Asian families' preferences for inter-group interactions. Importantly, we find that middle school demographics have no effect on white and Asian families' knowledge of schooling options, in line with the idea that they might have access to, or rely more on, other sources of information or different social networks to get information about schools.

7 Conclusions

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

Combining administrative data and novel survey evidence, we show that these differences are driven by a combination of preferences for the racial composition of schools and differences in information in the form of limited awareness of school options. Black and Hispanic students have heard of fewer schools, in particular fewer majority-white and high-quality schools. Attending majority white and Asian middle schools, however, expands their awareness sets and in turn affects their high school choices, which look more similar to those of their white peers.

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

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

References

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

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

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

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

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

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

Angrist, J. (2013). The perils of peer effects. Working Paper 19774, National Bureau of Economic Research.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

Hahm, D. W. and Park, M. (2022). A dynamic framework of school choice:effects of middle schools on high school choice. Technical report.

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

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

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

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

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

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

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

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

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

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

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

8 Tables

Table 1: Summary Statistics

	Administrative data sample				Survey respondents		
	MS applicants applying to HS	Experimental sample					
		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	0.66	0.77	0.55
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	-0.13	0.65	0.51	0.15	0.88
4th gr. Ela	0.23	0.19	-0.05	0.53	0.46	0.18	0.75
Panel B: Middle school characteristics							
% Black+Hispanic peer in MS	61	61	76	38	54	71	37
Enroll MS with >50% white+Asian	0.38	0.36	0.15	0.68	0.49	0.23	0.76
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 top3 choices	58	58	69	41	51	62	39
Mean baseline peer math in top3 choices	0.24	0.21	0.05	0.44	0.40	0.21	0.59
Mean popularity in top3 choices	4.18	4.06	3.43	4.98	5.45	4.19	6.71
Lists a screened program among top3 choices	0.69	0.67	0.59	0.79	0.78	0.68	0.88
Mean RC math VA in top3 choices	0.78	0.71	0.44	1.11	1.04	0.69	1.40
Panel D: High school offers							
Assigned to 1st choice	0.47	0.47	0.52	0.40	0.39	0.48	0.31
Assigned to top 3 choices	0.76	0.77	0.80	0.71	0.68	0.76	0.60
N	254,742	118,078	70,219	47,859	3,628	1,818	1,810

Notes: The administrative data sample in columns 1 to 4 includes students who applied to middle school for enrollment in 2015-2016 to 2019-2020 and then successively applied to high school for enrollment in 2018-2019 to 2022-2023. Columns 1 reports descriptive statistics for the sample of applicants who have demographic information. Columns 2 to 4 restrict the sample to the experimental sample which includes offered Middle school applicants who have (i) non-degenerate risk of school assignment, (ii) non-missing baseline test scores, and (iii) non-missing geographic information. The survey respondents in column 5 to 7 include any survey participants who answered at least one survey question. The baseline scores are 4th grade scores from the NY state standardized assessments. High school popularity corresponds to the number of applicants rejected by the program divided by the number of accepted applicants (city-mean is 1.37). Screened programs are programs that admit students based on their Middle school grades and/or

auditions and essays. The risk-controlled value-added computation (RC Regents math VA) follows that in Angrist et al. (2021).

Table 2: 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 on 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 3: Vignette Experiment Preference Estimates

Respondent race	Characteristic	Precise Info (1)	Imprecise Info (2)
White+Asian	Constant	2.17***	3.04***
		(0.08)	(0.09)
	High-academics	1.44***	
		(0.06)	
	High-safety	0.74***	1.08***
		(0.06)	(0.08)
	Majority Black	-0.44***	-0.70***
		(0.08)	(0.12)
	Majority Hispanic	-0.28***	-0.59**
		(0.08)	(0.11)
	Majority white+Asian	0.27***	0.36***
		(0.08)	(0.12)
Black+Hispanic	Constant	2.71***	3.41***
		(0.09)	(0.10)
	High-academics	1.28***	
		(0.07)	
	High-safety	0.66***	1.16***
		(0.07)	(0.09)
	Majority Black	-0.11	-0.28**
		(0.09)	(0.12)
	Majority Hispanic	0.00	-0.16
		(0.09)	(0.12)
	Majority white+Asian	-0.08	-0.22*
		(0.10)	(0.12)
N respondents		1,212	957

Notes: This table reports preference estimates for school cards for white and Asian respondents and Black and Hispanic respondents separately. The constant captures the absolute likelihood on a scale from 1 to 6 of listing the school. Preferences are estimated through Gibbs sampling using answers to survey questions Q17 and Q18.

Table 4: Beliefs About Admission Probability

Panel A: Beliefs about admission probabilities

	Applied to "dream school"	Admission beliefs
	(1)	(2)
Actual admission probability	0.113***	0.235***
	(0.040)	(0.029)
(Actual admission probability) × (Black+Hispanic)	0.049	-0.075**
	(0.053)	(0.036)
Black+Hispanic	-0.017	0.041
	(0.039)	(0.027)
Mean	0.84	0.425
N	2,408	3,354

Panel B: Beliefs about performance tercile

	Within the City	Within High-demand school
	(1)	(2)
Actual performance tercile	0.379***	0.171***
	(0.047)	(0.027)
(Actual performance tercile) × (Black+Hispanic)	-0.121**	0.066
	(0.055)	(0.044)
Black+Hispanic	0.268*	-0.102
	(0.149)	(0.092)
Mean	2.000	2.242
N	1,274	986

Notes: This table reports OLS estimates of the relationships between applicants' actual admission probabilities and relative performance and their beliefs about these. The student relative performance is measured as the tercile in the distribution of city's test score in panel A and in the distribution of students' test scores at a high-demand school. All models control for residential district fixed effects and 4th grade test score tercile. All columns except column 1 of Panel B control for school fixed effects. Column 2 of panel A also controls for applicants' random numbers, as actual admission probabilities estimates account for the uncertainty coming from the lottery. Robust standard errors are reported in parenthesis, clustered at the student level for column 2 of panel A. Panel A uses data from survey questions Q7a and Q7c, column 2 adds data from survey question Q13. panel B column 1 uses data from survey question Q11, panel B column 2 uses data from survey question Q12.

Table 5: 2SLS Estimates of Peer Share Effects on Black & Hispanic Students' HS Choices

	% Black (1)	% Hispanic (2)	% White (3)	Peer Ela (4)	Peer Math (5)	Popularity (6)	Screened (7)	Regents math VA (8)	Length of rol (9)
					Top 3 choices				
Majority white+Asian MS	-4.790***	-1.557**	6.562***	0.073***	0.127***	0.341**	-0.004	0.120***	0.267
	(0.822)	(0.773)	(1.008)	(0.019)	(0.021)	(0.165)	(0.028)	(0.032)	(0.223)
Share white+Asian (10pp)	-1.014***	-0.328**	1.331***	0.015***	0.021***	0.111***	0.009*	0.025***	0.017
	(0.158)	(0.158)	(0.194)	(0.004)	(0.004)	(0.034)	(0.005)	(0.007)	(0.042)
mean	23.25	45.90	29.00	-0.01	0.05	3.43	0.59	0.28	8.22
					All choices				
Majority white+Asian MS	-3.881***	-1.358**	5.490***	0.058***	0.109***	0.257**	0.024	0.096***	
	(0.696)	(0.608)	(0.800)	(0.015)	(0.016)	(0.120)	(0.021)	(0.024)	
Share white+Asian (10pp)	-0.768***	-0.284**	1.047***	0.012***	0.017***	0.073**	0.009**	0.024***	
	(0.133)	(0.122)	(0.149)	(0.003)	(0.003)	(0.025)	(0.004)	(0.005)	
mean	24.04	46.78	27.37	-0.07	-0.01	3.05	0.78	0.21	
N	70,219	70,219	70,219	70,219	70,219	69,834	70,219	70,208	

Notes: This table reports 2SLS estimates of middle school demographic composition effects on Black and Hispanic high school choices. Panel A focuses on each applicant's 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 2019-2020 and then successively applied to high school for enrollment in 2018-2019 to 2022-2023. 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 and RC VA are defined in the notes of Table 1. Robust standard errors in parenthesis.

Table 6: 2SLS Estimates of Peer Share Effects on White & Asian Students' HS Choices

	% Black (1)	% Hispanic (2)	% White (3)	Peer Ela (4)	Peer Math (5)	Popularity (6)	Screened (7)	Regents math VA (8)	Length of rol (9)
					Top 3 choices				
Majority Black+Hispanic MS	1.271*	1.387*	-2.879**	-0.066**	-0.074**	0.189	0.012	-0.042	-0.758**
	(0.663)	(0.786)	(1.029)	(0.025)	(0.027)	(0.198)	(0.029)	(0.040)	(0.280)
Share Black+Hispanic (10pp)	0.135	0.461**	-0.644**	-0.012**	-0.014**	0.042	0.006	-0.014*	-0.091
	(0.133)	(0.166)	(0.211)	(0.005)	(0.005)	(0.042)	(0.006)	(0.008)	(0.057)
mean	12.82	27.67	57.41	0.35	0.45	5.00	0.79	0.59	7.51
					All choices				
Majority Black+Hispanic MS	1.622**	1.581**	-3.423***	-0.053**	-0.065**	0.124	-0.013	-0.027	
	(0.587)	(0.653)	(0.896)	(0.019)	(0.021)	(0.151)	(0.022)	(0.031)	
Share Black+Hispanic (10pp)	0.303**	0.432**	-0.776***	-0.009**	-0.012**	0.014	0.002	-0.012*	
	(0.120)	(0.137)	(0.182)	(0.004)	(0.004)	(0.033)	(0.004)	(0.006)	
mean	13.60	30.41	53.96	0.23	0.33	4.33	0.87	0.49	
N	45,649	45,649	45,649	45,649	45,649	45,595	45,649	45,626	

Notes: This table reports 2SLS estimates of middle school demographic composition effects on white and Asian high school choices. Panel A focuses on each applicant's top 3 choices, panel B includes all the choices. The sample, controls and endogeneous variables are as defined in the notes of Table 5. Robust standard errors in parenthesis.

Table 7: 2SLS Estimates of Peer Share Effects on Characteristics of Offered High School

	Matched in 1st round (1)	Offered rank (2)	% Black (3)	% Hispanic (4)	% White (5)	Peer Ela (6)	Peer Math (7)	Popularity (8)	Screened (9)	Regents math VA (10)
Panel A: Black & Hispanic students										
Majority white+Asian MS	-0.025*	0.091	-4.321***	-1.132	5.675***	0.042*	0.082***	-0.121	-0.023	0.094**
	(0.014)	(0.150)	(1.204)	(1.007)	(1.235)	(0.022)	(0.024)	(0.169)	(0.029)	(0.045)
Share white+Asian (10pp)	-0.003	0.004	-0.864***	-0.376*	1.215***	0.009**	0.012**	0.070**	0.006	0.028**
	(0.002)	(0.029)	(0.226)	(0.200)	(0.218)	(0.004)	(0.004)	(0.032)	(0.005)	(0.009)
mean	0.96	2.35	27.42	49.32	21.54	-0.19	-0.14	1.29	0.21	-0.00
N	70,195	66,726	67,539	67,539	67,539	67,508	67,508	67,555	67,577	67,401
Panel B: White & Asian students										
Majority Black+Hispanic MS	0.002	-0.873***	1.825	-0.255	-1.940	0.012	-0.029	0.837***	0.066*	-0.050
	(0.019)	(0.211)	(1.136)	(1.161)	(1.515)	(0.032)	(0.034)	(0.250)	(0.038)	(0.059)
Share Black+Hispanic (10pp)	-0.000	-0.107**	0.421*	0.418*	-0.894**	0.002	-0.006	0.167**	0.016**	-0.006
	(0.004)	(0.044)	(0.239)	(0.250)	(0.319)	(0.007)	(0.007)	(0.054)	(0.008)	(0.012)
mean	0.93	2.91	15.38	32.21	50.43	0.15	0.25	2.74	0.42	0.39
N	45,633	41,961	42,643	42,643	42,643	42,641	42,641	42,659	42,700	42,560

Notes: This table reports 2SLS estimates of middle school demographic composition effects on high school offers. Panel A focuses on Black and Hispanic applicants, while panel B focuses on white and Asian applicants. The sample, controls and endogeneous variables are as defined in the notes of Table 5. Robust standard errors in parenthesis.

Table 8: 2SLS Estimates of Peer Share Effects on Test Scores

	Middle school outcomes				***High school outcomes***			
	Has 6th grade test (1)	Has 7th grade test (2)	6th grade Math (3)	7th grade Math (4)	Took Regents Math (5)	Took SAT (6)	Regents Math (7)	SAT Math (8)
Panel A: Black & Hispanic students								
Majority white+Asian MS	-0.005	0.012	0.031	0.086**	-0.022	-0.010	0.140*	0.024
	(0.007)	(0.011)	(0.040)	(0.041)	(0.026)	(0.020)	(0.075)	(0.063)
Share white+Asian (10pp)	-0.001	0.002	-0.006	0.012	0.003	0.002	0.005	-0.003
	(0.001)	(0.002)	(0.008)	(0.008)	(0.005)	(0.004)	(0.014)	(0.013)
mean	0.72	0.66	-0.11	-0.09	0.34	0.34	-0.09	-0.29
N	70,219	70,219	50,435	45,205	70,219	70,219	23,624	24,193
Panel B: White & Asian students								
Majority Black+Hispanic MS	-0.011	-0.015	-0.024	0.020	0.014	0.020	0.002	-0.099
	(0.009)	(0.013)	(0.049)	(0.048)	(0.030)	(0.024)	(0.087)	(0.091)
Share Black+Hispanic (10pp)	0.001	-0.003	-0.005	-0.003	0.000	-0.002	0.003	0.002
	(0.002)	(0.002)	(0.011)	(0.011)	(0.006)	(0.005)	(0.019)	(0.020)
mean	0.71	0.67	0.73	0.75	0.29	0.40	0.65	0.75
N	45,649	45,649	32,245	30,200	45,649	45,649	13,388	18,308

Notes: This table reports 2SLS estimates of middle school demographic composition effects on 6th, 7th grade State standardized test scores and SAT and Regents math test scores. Panel A focuses on Black and Hispanic applicants, while panel B focuses on white and Asian applicants. The sample and controls variables are as defined in the notes of Table 5. Robust standard errors in parenthesis.

Table 9: Vignette Experiment Preference Estimates by MS

	Black & Hispanic (1)	White & Asian (2)
Constant	2.76***	2.24***
	(0.09)	(0.13)
Majority white+Asian MS	-0.66***	-0.09
	(0.20)	(0.15)
Imprecise info	0.81***	0.99***
	(0.09)	(0.12)
(Imprecise info) × (Majority white+Asian MS)	0.21	-0.13
	(0.20)	(0.14)
High-academics	1.26***	1.58***
	(0.08)	(0.11)
(High-academics) × (Majority white+Asian MS)	0.36**	-0.23*
	(0.18)	(0.13)
(High-safety)	0.86***	0.92***
	(0.06)	(0.09)
(High-safety) × (Majority white+Asian MS)	0.10	-0.13
	(0.14)	(0.10)
Majority Black	-0.22**	-0.55***
	(0.09)	(0.12)
(Majority Black) × (Majority white+Asian MS)	0.21	0.04
	(0.20)	(0.14)
Majority Hispanic	-0.12	-0.47***
	(0.09)	(0.12)
(Majority Hispanic) × (Majority white+Asian MS)	0.34*	0.12
	(0.20)	(0.14)
Majority white+Asian	-0.24***	0.04
	(0.09)	(0.12)
(Majority white+Asian) × (Majority white+Asian MS)	0.51***	0.34**
	(0.20)	(0.14)
N respondents	914	1,086

Notes: This table reports estimates of middle school demographic composition effects on preferences for school cards. Column 1 reports estimates for Black and Hispanic respondents separately, column 2 for white and Asian respondents. The constant captures the absolute likelihood on a scale from 1 to 6 of listing the school. Preferences are estimated through Gibbs sampling using answers to survey questions Q17 and Q18.

Table 10: Peer Effects on Consideration Sets

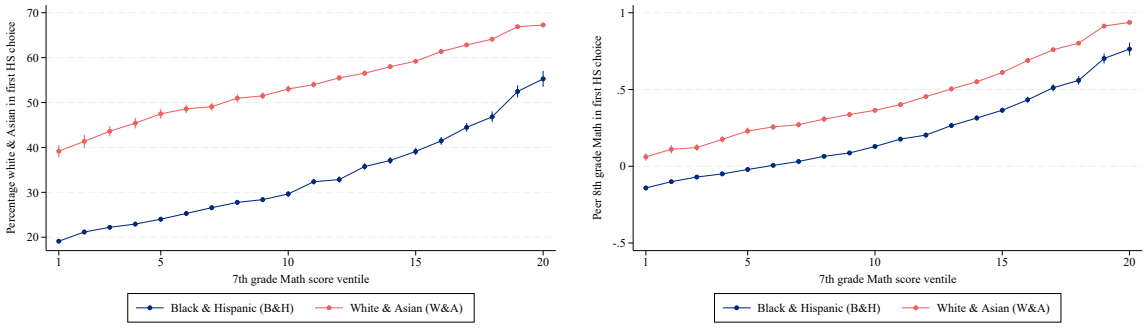
	Any school		Popular		High white+Asian %		High bl. Math		High VA	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(Black+Hispanic) × (High white+Asian MS)	0.08***	0.18	0.09***	0.02	0.10***	0.24*	0.11***	0.18	0.09***	0.21*
	(0.02)	(0.11)	(0.03)	(0.18)	(0.02)	(0.13)	(0.02)	(0.11)	(0.02)	(0.13)
High white+Asian MS	-0.04***	-0.09	0.00	-0.05	-0.04**	-0.09	-0.03**	-0.11	-0.02	-0.14
	(0.01)	(0.07)	(0.02)	(0.11)	(0.02)	(0.08)	(0.01)	(0.08)	(0.02)	(0.09)
Black+Hispanic	-0.04***	-0.09***	-0.07***	-0.09***	-0.08***	-0.11***	-0.07***	-0.08***	-0.07***	-0.06***
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
N	25,690	25,690	5,138	5,138	13,801	13,801	15,683	15,683	10,161	10,161
mean white+Asian	0.370	0.370	0.680	0.680	0.510	0.510	0.450	0.450	0.440	0.440

Notes: This table reports OLS and 2SLS estimates of middle school demographic composition effects on survey respondents' awareness sets. All regressions control for residential district fixed effects and 4th grade test score tercile. Endogeneous variables are defined in Appendix. Clustered standard errors at the student-level in parenthesis. Panel A uses data from survey question Q9.

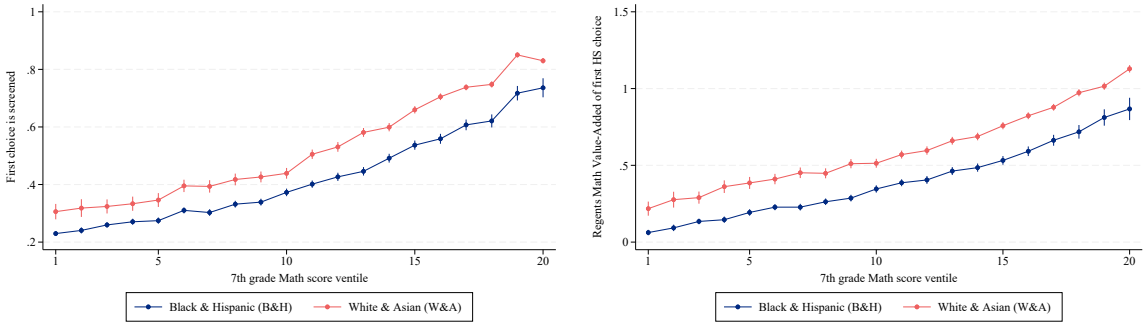
9 Figures

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

(a) Differences in peer composition



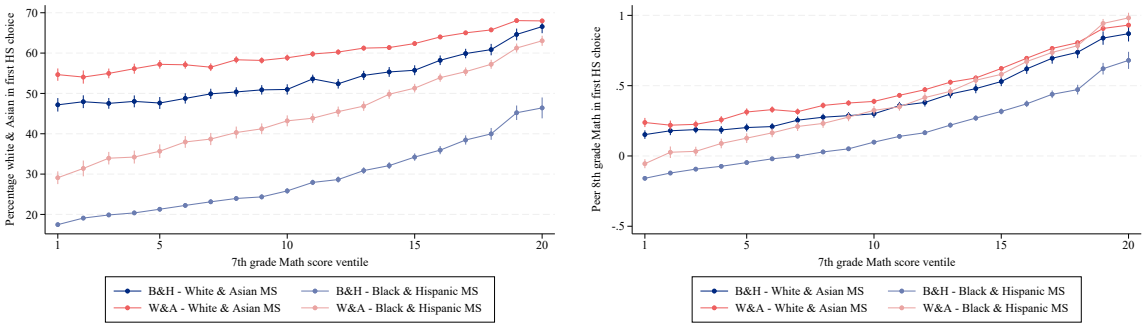
(b) Differences in school selectivity and quality



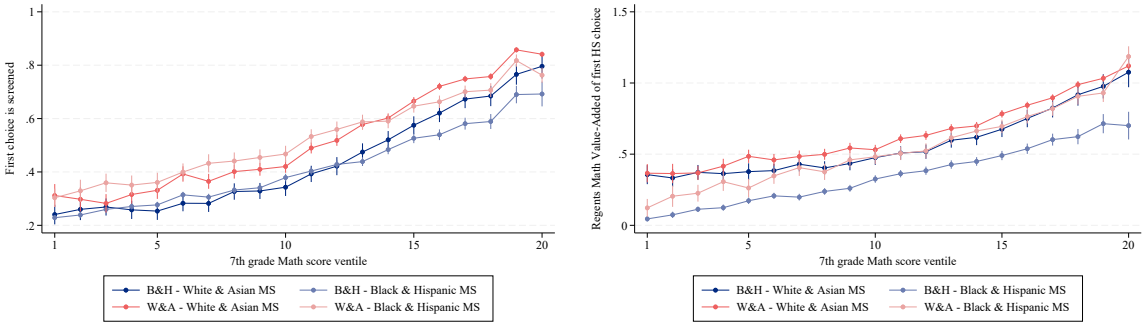
Notes: This figures plots mean characteristics of the school ranked first on the high school application by applicants' race and middle school test score ventiles, for application years 2015-17 and 2020. Panel (a) considers the percentage of white and Asian students and the mean 8th grade math scores of students at the school. Panel (b) considers the probability the school is screened and the RC math value-added (RC-VA) of the school.

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

(a) Differences in peer composition

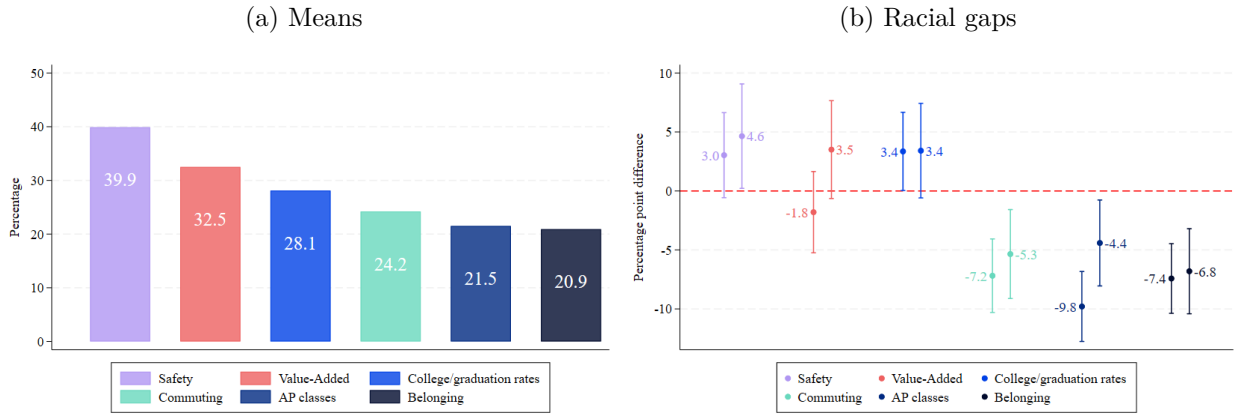


(b) Differences in school selectivity and quality



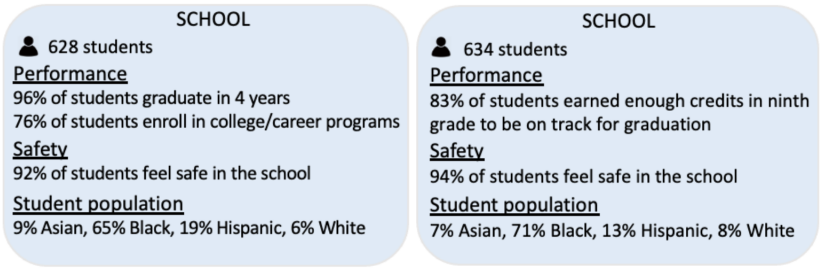
Notes: This figures plots mean characteristics of the school ranked first on the high school application by applicants' race and middle school test score ventiles and by the the racial composition of the middle school attended by the applicant. Characteristics of high school choices are depicted using a lighter shade for students enrolled in a majority-white and Asian middle school ($\geq 50\%$ of white and Asian enrollment) and in a darker shade for students enrolled in a majority-Black and Hispanic Middle School. The characteristics considered are the same as in Figure 1.

Figure 3: Differences in Most Important School Features



Notes: This figure reports differences in stated preference for school characteristics. Panel (a) reports the percentage of respondents who mentioned each school feature among their three most important when deciding which school to include in their list. Panel (b) reports the differences in the percentage of respondents who mentioned each school feature among Black and Hispanic respondents compared to white and Asian respondents. For each feature, the first bar depicts the raw percentage point difference while the second bar depicts the percentage point difference controlling for district of residence and middle school baseline test score. The capped lines display 95% confidence intervals. This figure uses data from survey question Q8.

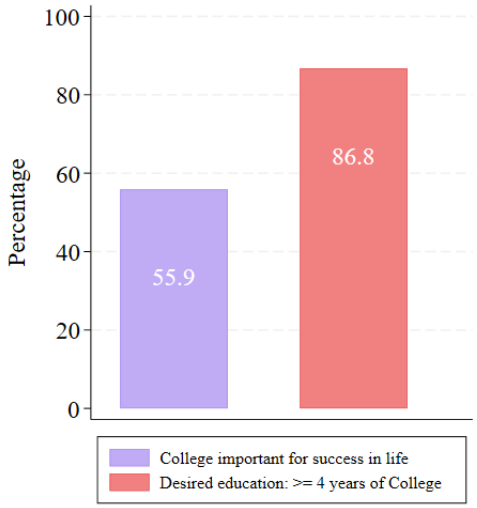
Figure 4: School Cards for Vignette Experiment



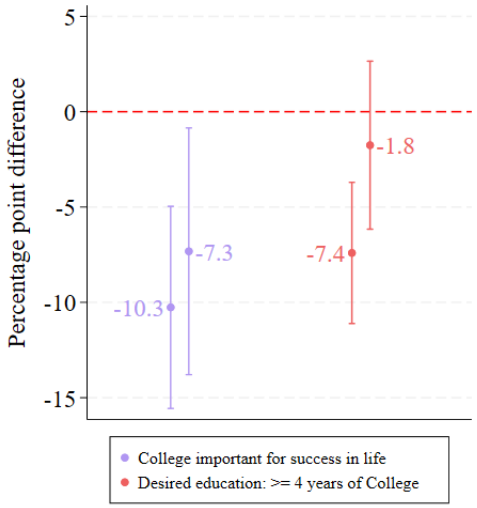
Notes: This figure displays an example of two cards used in the vignette experiment. The left card displays precise academic information (Treatment 1, received by around 60% of the experiment participants). The right card shows imprecise academic information (Treatment 2, received by around 40% of the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

Figure 5: Differences in Aspirations

(a) Means



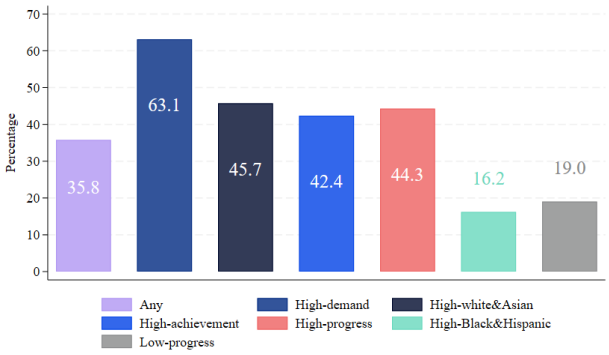
(b) Racial gaps



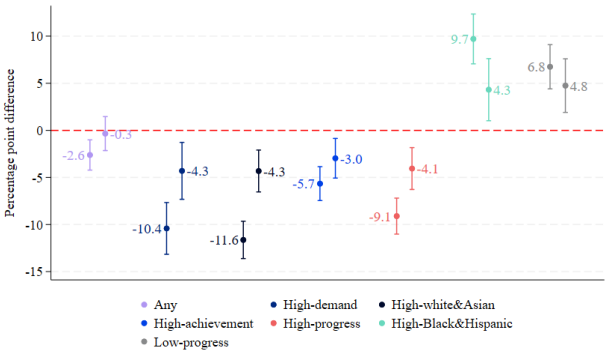
Notes: This figure reports differences respondents' aspirations for their students' future academic pursuits. Panel (a) reports the percentage of respondents who view college as important for success in life and who would like their kids to pursue at least 4 years of college. Panel (b) compares the academic aspirations of Black and Hispanic respondents to those of white and Asian respondents. For each answer, the first bar depicts the raw percentage point difference while the second bar depicts the percentage point difference controlling for district of residence and middle school baseline test score. The capped lines display 95% confidence intervals. This figure uses data from survey questions Q14 and Q15.

Figure 6: Differences in Awareness Sets

(a) Means



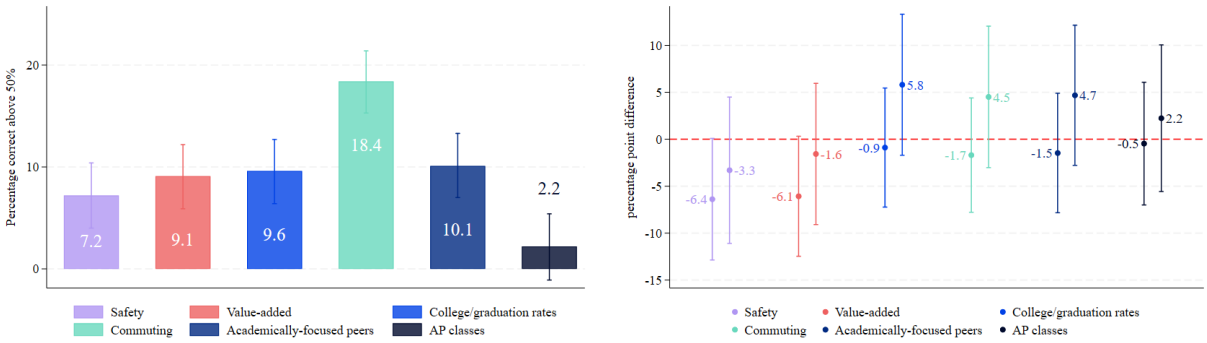
(b) Racial gaps



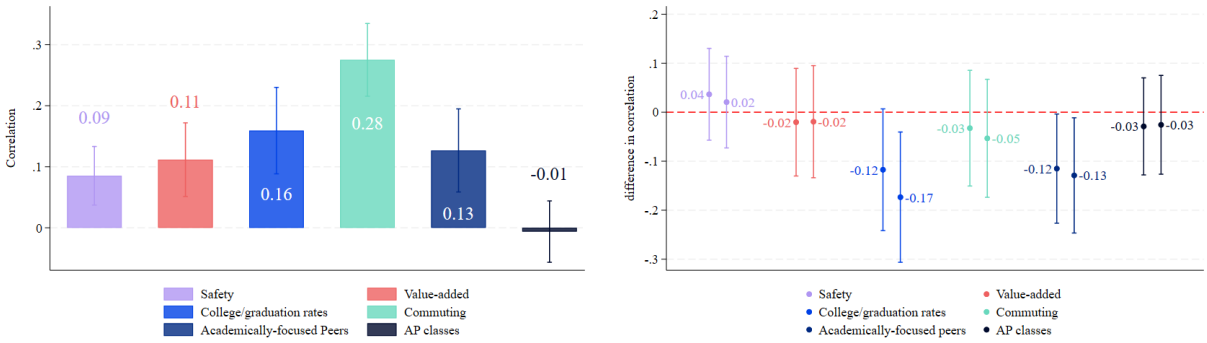
Notes: This figure reports differences in respondents' awareness sets. Panel (a) reports the mean share of schools respondents were aware of by school type. Panel (b) reports the differences in the share of schools Black and Hispanic respondents were aware of compared to white and Asian respondents. For each school type, the first bar depicts the raw percentage point difference while the second bar depicts the percentage point difference controlling for district of residence and middle school baseline test score. The capped lines display 95% confidence intervals. This figure uses data from survey question Q9.

Figure 7: Differences in Information About School Characteristics

(a) Excess p(correct) - pairwise comparisons



(b) Correlation between real quartile and answer



Notes: This figure reports differences in information about school characteristics. Panel (a) reports the percentage of respondents who responded correctly above 50% (which would correspond to random guesses only). Panel (b) reports the correlation of respondents' rankings with the true ranking of the school they were shown among schools in the same borough. For each school characteristics in each panel, the second bar corresponds to the differences in accuracy between Black and Hispanic respondents and white and Asian respondents. The capped lines display 95% confidence intervals. This figure uses data from survey questions Q10a-g.

A Appendix Tables

Table A1: Sample Construction

	N
Non-swd 6th grade applicants, applying to at least one program	336,083
Who enroll in 6th grade in the district, with demographics	303,299
Who apply to 9th grade in the district	268,446
Has non-degenerate risk of school assignment	145,101
Who are offered a 6th grade seat	127,104
Who have baseline scores	118,078

Notes: The initial administrative data sample includes students who applied for middle school enrollment in the school years from 2015-2016 to 2020-2021.

Table A2: Survey Summary Statistics

	Applicants with baselines			Respondents			
				Answered > 0%	Answered > 50%		
	NYC (1)	Wave 2 (2)	Wave 3 (3)	All (4)	All (5)	Wave 2 (6)	Wave 3 (7)
Asian	20%	27%	22%	29%	29%	30%	28%
Black	19%	15%	17%	14%	14%	13%	16%
Hispanic	42%	39%	39%	33%	32%	32%	31%
White	16%	16%	19%	21%	22%	21%	23%
Poverty	76%	73%	74%	66%	64%	62%	66%
Brooklyn	29%	33%	28%	31%	31%	33%	28%
Queens	33%	36%	32%	37%	36%	38%	33%
Manhattan	10%	7%	11%	10%	11%	9%	15%
Bronx	21%	16%	20%	14%	14%	12%	17%
Math 4^{th}	.06	.32	.29	.51	.54	.54	.54
ELA 4^{th}	.02	.28	.28	.46	.49	.50	.49
N	47,618	11,415	9,986	3,628	3,099	1,935	1,164
Response %				17%	15%	17%	12%

Notes: This table reports summary statistics for students applying to enroll in high school in the fall of 2023. The first column restricts the sample to high school applicants with non missing baseline demographics and

achievement outcomes, while the second and third column to applicants selected to receive our survey. Columns (4) to (7) report summary statistics for the survey respondents.

Table A3: Differences in Information Accuracy by Race

	Safety	VA	College rate	Commuting	Peers	AP classes
	(1)	(2)	(3)	(4)	(5)	(6)
Black+Hispanic	-0.01	-0.02	0.05*	-0.02	0.02	0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
N	1,767	1,765	1,768	1,798	1,748	1,735

Notes: This table reports OLS estimates of racial differences in accuracy of beliefs about school attributes. The dependent variable is a dummy indicating correct or approximately correct beliefs. Regression estimates control for student baseline achievement and district of residence.

Table A4: Attrition and Covariate Balance - Discrete Treatment

	Offered majority white+Asian MS				
	All	White	Minority	Black	Hispanic
Dependent variable	(1)	(2)	(3)	(4)	(5)
Panel A: Attrition					
Enrolls in district (6th grade)	0.030***	0.020**	0.033***	0.069***	0.016
	(0.007)	(0.011)	(0.009)	(0.016)	(0.011)
mean	0.92	0.93	0.91	0.89	0.92
N	144,251	54,399	87,110	28,143	58,697
Has 9th grade application	0.004	-0.008	0.016	0.013	0.022
	(0.008)	(0.012)	(0.011)	(0.020)	(0.013)
mean	0.89	0.90	0.88	0.86	0.89
N	132,592	50,686	79,391	25,393	53,998
Panel B: Covariates balance					
Black	0.006		0.007		
	(0.010)		(0.016)		
mean	0.19		0.31		
Hispanic	0.016		-0.007		
	(0.012)		(0.016)		
mean	0.41		0.69		
White+Asian	-0.027**				
	(0.013)				
mean	0.39				
Female	-0.001	-0.021	0.013	0.010	0.016
	(0.013)	(0.019)	(0.018)	(0.031)	(0.023)
mean	0.51	0.50	0.53	0.53	0.52
English Language Learner	-0.007	-0.007	-0.009	-0.004	-0.011
	(0.006)	(0.008)	(0.008)	(0.005)	(0.013)
mean	0.09	0.07	0.10	0.02	0.14
Low-income	0.004	-0.004	0.008	0.020	0.002
	(0.012)	(0.019)	(0.016)	(0.029)	(0.020)
mean	0.72	0.58	0.82	0.79	0.84
Baseline English	0.035*	0.043	0.036	-0.038	0.083**
	(0.020)	(0.029)	(0.028)	(0.046)	(0.037)
mean	0.19	0.54	-0.05	-0.05	-0.05
Baseline Math	0.008	0.040	-0.008	-0.044	0.018
	(0.020)	(0.027)	(0.028)	(0.046)	(0.037)
mean	0.19	0.66	-0.12	-0.21	-0.09
N	118,078	45,649	70,219	21,944	48,275

Notes: This table reports coefficients from regressions of the variables listed to the left on an indicator for being offered a seat at a majority white and Asian middle school. Column heading labels refer to different estimation samples. The sample is always limited to applicants with non-degenerate risk of middle school assignment.

Table A5: Attrition and Covariate Balance - Continuous Treatment

	Offered 10pp more White+Asian				
	All	White	Minority	Black	Hispanic
Dependent variable	(1)	(2)	(3)	(4)	(5)
Panel A: Attrition					
Enrolls in district (6th grade)	0.007***	0.005**	0.007***	0.014***	0.004*
	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)
mean	0.92	0.93	0.91	0.89	0.92
N	144,251	54,399	87,110	28,413	58,697
Has 9th grade application	0.003*	0.002	0.003	0.003	0.004
	(0.002)	(0.003)	(0.002)	(0.004)	(0.003)
mean	0.89	0.90	0.88	0.86	0.89
N	132,592	50,686	79,391	25,393	53,998
Panel B: Covariates balance					
Black	-0.001		-0.001		
	(0.002)		(0.003)		
mean	0.19		0.31		
Hispanic	0.000		0.001		
	(0.002)		(0.003)		
mean	0.41		0.69		
White+Asian	0.000				
	(0.002)				
mean	0.39				
Female	0.003	-0.001	0.005	-0.003	0.009**
	(0.003)	(0.005)	(0.004)	(0.006)	(0.004)
mean	0.51	0.50	0.53	0.53	0.52
English Language Learner	-0.001	-0.003	0.000	-0.002	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
mean	0.09	0.07	0.10	0.02	0.14
Low-income	-0.003	-0.004	-0.002	-0.007	-0.002
	(0.002)	(0.004)	(0.003)	(0.005)	(0.003)
mean	0.72	0.58	0.82	0.79	0.84
Baseline English	0.007	0.013*	0.001	-0.012	0.010
	(0.004)	(0.007)	(0.006)	(0.009)	(0.007)
mean	0.19	0.54	-0.05	-0.05	-0.05
Baseline Math	-0.001	0.008	-0.007	-0.011	-0.003
	(0.004)	(0.007)	(0.005)	(0.009)	(0.007)
mean	0.19	0.66	-0.12	-0.21	-0.09
N	118,078	45,649	70,219	21,944	48,275

Notes: This table reports coefficients from regressions of the variables listed to the left on the continuous version of our instrument for exposure to white and Asian peers, as defined by the potential share of white and Asian peers in the offered middle school. Column heading labels refer to different estimation samples. The sample is always limited to applicants with non-degenerate risk of middle school assignment.

Table A6: OLS Estimates of Peer Effects on Students' Top 3 High School Choices

	% Black (1)	% Hispanic (2)	% white (3)	Peer Ela (4)	Peer Math (5)	Popularity (6)	Screened (7)	Regents math VA (8)	Length of rol (9)
Panel A: Black & Hispanic students									
Majority White+Asian MS	-6.368***	-4.510***	10.903***	0.115***	0.141***	0.480***	-0.020**	0.185***	-0.299***
	(0.161)	(0.168)	(0.217)	(0.004)	(0.004)	(0.037)	(0.006)	(0.009)	(0.050)
Share White+Asian (10pp)	-1.403***	-0.798***	2.198***	0.018***	0.022***	0.057***	0.005**	0.032***	-0.082***
	(0.057)	(0.057)	(0.069)	(0.001)	(0.001)	(0.012)	(0.002)	(0.003)	(0.015)
Mean	23.25	45.90	29.00	-0.01	0.05	3.43	0.59	0.44	8.22
N	70,219	70,219	70,219	70,219	70,219	69,834	70,219	70,208	70,219
Panel B: White & Asian students									
Majority Black+Hispanic MS	4.112***	4.081***	-8.372***	-0.083***	-0.096***	0.001	0.040***	-0.111***	0.557***
	(0.120)	(0.153)	(0.203)	(0.005)	(0.005)	(0.036)	(0.005)	(0.009)	(0.048)
Share Black+Hispanic (10pp)	0.945***	1.241***	-2.204***	-0.017***	-0.027***	0.024*	0.006**	-0.022***	0.139***
	(0.046)	(0.059)	(0.076)	(0.002)	(0.002)	(0.014)	(0.002)	(0.004)	(0.019)
mean	12.82	27.67	57.41	0.35	0.45	5.00	0.79	1.13	7.51
N	45,649	45,649	45,649	45,649	45,649	45,595	45,649	45,626	45,649

Notes: This table reports OLS estimates of middle school demographic composition effects on the characteristics of top 3 high school choices. Panel A focuses on Black and Hispanic applicants, while panel B on White and Asian applicants. The sample and controls are as defined in the notes of Table 5. Robust standard errors in parenthesis.

Table A7: 2SLS Estimates of Peer Share Effects on Black Students' High School Choices

	% Black (1)	% Hispanic (2)	% white (3)	Peer Ela (4)	Peer Math (5)	Popularity (6)	Screened (7)	Regents math VA (8)	Length of rol (9)
Top 3 choices									
Majority White+Asian MS	-5.247**	-1.261	6.778***	0.091**	0.149***	0.538**	-0.068	0.125**	0.267
	(1.914)	(1.202)	(1.987)	(0.033)	(0.035)	(0.264)	(0.049)	(0.059)	(0.280)
Share White+Asian (10pp)	-1.604***	0.038	1.570***	0.014**	0.020**	0.143**	0.008	0.019*	-0.004
	(0.352)	(0.232)	(0.354)	(0.006)	(0.007)	(0.052)	(0.009)	(0.011)	(0.073)
Mean	33.76	37.19	26.95	-0.02	0.02	3.06	0.61	0.16	8.43
All choices									
Majority White+Asian MS	-2.871***	-1.257*	4.364***	0.043**	0.092***	0.099	0.038	0.088**	
	(0.686)	(0.762)	(0.924)	(0.018)	(0.020)	(0.154)	(0.027)	(0.029)	
Share White+Asian (10pp)	-1.313***	-0.155	1.472***	0.016**	0.020***	0.127***	0.011	0.019**	
	(0.291)	(0.192)	(0.274)	(0.005)	(0.005)	(0.038)	(0.007)	(0.008)	
Mean	34.31	38.57	25.06	-0.09	-0.04	2.69	0.80	0.10	
N	21,944	21,944	21,944	21,944	21,944	21,850	21,944	21,940	21,944

Notes: This table reports 2SLS estimates of middle school demographic composition effects on Black students' high school choices. Panel A focuses on each applicant's top 3 choices, panel B includes all the choices. The sample, controls and endogeneous variables are as defined in the notes of Table 5. Robust standard errors in parenthesis.

Table A8: 2SLS Estimates of Peer Share Effects on Hispanic Students' High School Choices

	% Black (1)	% Hispanic (2)	% white (3)	Peer Ela (4)	Peer Math (5)	Popularity (6)	Screened (7)	Regents math VA (8)	Length of rol (9)
				Top 3 choices					
Majority White+Asian MS	-3.901***	-2.027**	6.106***	0.061**	0.115***	0.192	0.037	0.118**	0.267
	(0.760)	(1.002)	(1.157)	(0.024)	(0.026)	(0.214)	(0.034)	(0.039)	(0.280)
Share White+Asian (10pp)	-0.576***	-0.639**	1.192***	0.017***	0.023***	0.097**	0.012*	0.029***	0.014
	(0.150)	(0.209)	(0.230)	(0.005)	(0.005)	(0.045)	(0.007)	(0.008)	(0.053)
Mean	18.47	49.86	29.93	-0.00	0.07	3.60	0.58	0.29	8.12
				All choices					
Majority White+Asian MS	-2.871***	-1.257*	4.364***	0.043**	0.092***	0.099	0.038	0.088**	
	(0.686)	(0.762)	(0.924)	(0.018)	(0.020)	(0.154)	(0.027)	(0.029)	
Share White+Asian (10pp)	-0.393**	-0.468**	0.846***	0.011**	0.016***	0.044	0.009*	0.027***	
	(0.128)	(0.157)	(0.178)	(0.004)	(0.004)	(0.033)	(0.005)	(0.006)	
Mean	19.37	50.51	28.42	-0.06	0.01	3.22	0.77	0.23	
N	48,275	48,275	48,275	48,275	48,275	47,984	48,275	48,268	48,275

Notes: This table reports 2SLS estimates of middle school demographic composition effects on Hispanic students' high school choices. Panel A focuses on each applicant's top 3 choices, panel B includes all the choices. The sample, controls and endogeneous variables are as defined in the notes of Table 5. Robust standard errors in parenthesis.

Table A9: 2SLS Estimates of Peer Share Effects on Black & Hispanic Students' Top 3 High School Choices - Multiple Treatment

	% Black (1)	% Hispanic (2)	% white (3)	Peer Ela (4)	Peer Math (5)	Popularity (6)	Screened (7)	Regents math VA (8)	Length of rol (9)
Panel A: Black applicants									
Share white (10pp)	-1.700***	0.156	1.554***	0.013**	0.019**	0.147**	0.005	0.034**	0.002
	(0.354)	(0.225)	(0.358)	(0.006)	(0.007)	(0.053)	(0.009)	(0.013)	(0.074)
Share Hispanic (10pp)	-1.367***	1.657***	-0.223	-0.007	-0.009	0.048	-0.035**	-0.009	0.090
	(0.410)	(0.310)	(0.390)	(0.009)	(0.009)	(0.076)	(0.011)	(0.018)	(0.095)
Mean	33.76	37.19	26.95	-0.02	0.02	3.06	0.61	0.36	8.43
N	21,944	21,944	21,944	21,944	21,944	21,848	21,944	21,940	21,944
Panel B: Hispanic applicants									
Share white (10pp)	-0.055	-1.244***	1.252***	0.018***	0.021***	0.023	0.022**	0.023**	0.141**
	(0.141)	(0.229)	(0.244)	(0.005)	(0.006)	(0.050)	(0.007)	(0.011)	(0.058)
Share Black (10pp)	1.879***	-2.199***	0.237	0.003	-0.006	-0.271***	0.037***	-0.011	0.472***
	(0.262)	(0.303)	(0.335)	(0.008)	(0.008)	(0.080)	(0.011)	(0.016)	(0.088)
Mean	33.76	37.19	26.95	-0.02	0.02	3.06	0.61	0.36	8.43
N	48,275	48,275	48,275	48,275	48,275	47,973	48,275	48,266	48,275
Panel C: White applicants									
Share Black (10pp)	0.748**	0.028	-0.860**	-0.008	-0.013	0.048	0.009	-0.016	-0.288**
	(0.274)	(0.267)	(0.370)	(0.009)	(0.010)	(0.077)	(0.010)	(0.018)	(0.094)
Share Hispanic (10pp)	-0.247	0.731***	-0.508*	-0.014**	-0.014**	0.038	0.003	-0.021	0.034
	(0.155)	(0.217)	(0.261)	(0.006)	(0.007)	(0.057)	(0.007)	(0.013)	(0.069)
Mean	33.76	37.19	26.95	-0.02	0.02	3.06	0.61	0.36	8.43
N	45,649	45,649	45,649	45,649	45,649	45,533	45,649	45,626	45,649

Notes: This table reports 2SLS estimates of middle school demographic composition effects on high school choices for models with two endogenous regressors, one for each race share different from own. Panel A focuses on Black applicants, panel B on Hispanic applicants, while panel C on white and Asian applicants. The sample, controls and endogeneous variables are as defined in the notes of Table 5. Robust standard errors in parenthesis.

Table A10: Peer Effects on Perceived Discrimination

	Peer discrimination		Teacher discrimination		Act on fear discrimination		Fit well other races	
	OLS (1)	IV (2)	OLS (3)	IV (4)	OLS (5)	IV (6)	OLS (7)	IV (8)
(Asian) × (High other race MS)	-0.05	-0.30*	-0.02	-0.34*	-0.15***	-0.35***	0.09	-0.11*
	(0.05)	(0.18)	(0.04)	(0.18)	(0.05)	(0.16)	(0.05)	(0.16)
(Black) × (High other race MS)	0.03	0.02	0.07	0.10	0.04	0.23	0.09	0.16
	(0.07)	(0.13)	(0.08)	(0.14)	(0.07)	(0.16)	(0.08)	(0.14)
(Hispanic) × (High other race MS)	-0.00	-0.22*	0.05	-0.12	0.04	0.00	0.01	0.41**
	(0.04)	(0.13)	(0.04)	(0.17)	(0.04)	(0.16)	(0.05)	(0.15)
(White) × (High other race MS)	-0.00	-0.05	0.00	0.04	-0.05	-0.02	0.06	0.15
	(0.04)	(0.09)	(0.03)	(0.08)	(0.04)	(0.08)	(0.05)	(0.12)
N	1,934	1,934	1,937	1,937	1,937	1,937	1,932	1,932

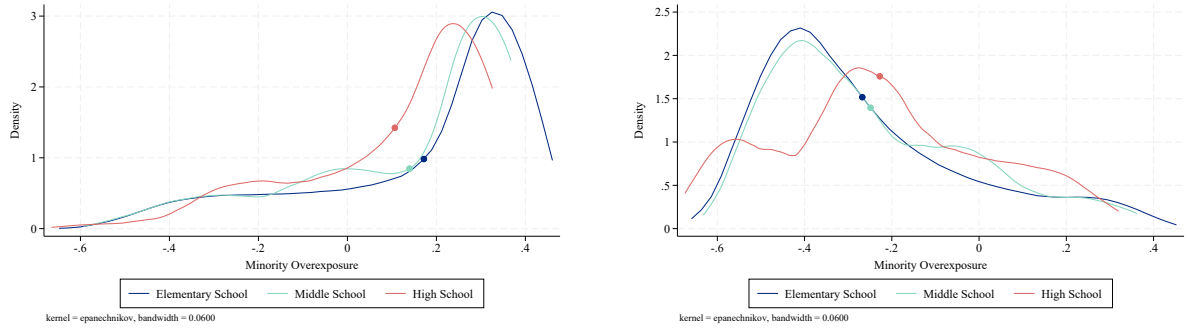
Notes: This table reports OLS and 2SLS estimates of the effect of enrolling in middle schools where the majority of peers are from a different race, relative to attending schools enrolling a majority of same-race peers, on measures of perceived discrimination. The effect of exposure to other-race peers 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. 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", "My student is likely to be treated negatively by their teachers based on their race", "My student would feel like they belong in a school where the majority of peers are from a different race". The last outcomes 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?".

B Appendix Figures

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

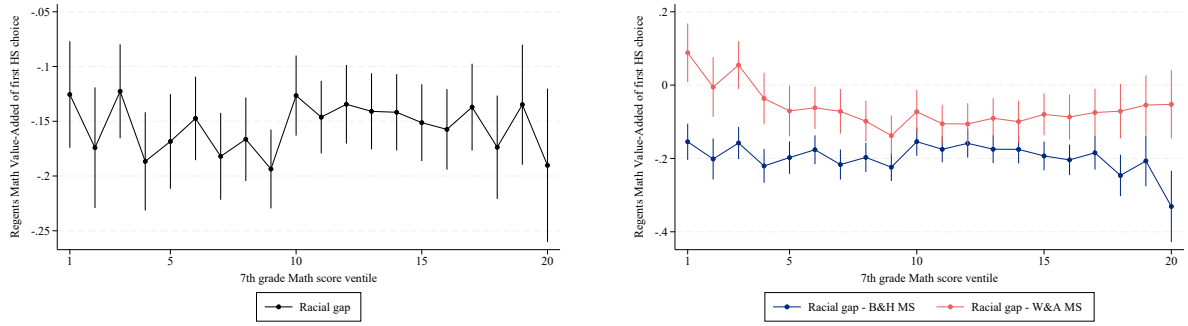
(a) Black and Hispanic students

(b) white and Asian students



Notes: These figures show the distribution of own-race overexposure, a measure of school segregation, for Black and Hispanic students (panel a) and white and Asian students (panel b) in different grade levels (elementary, middle and high school).

Figure B2: Differences in Value-Added of High School Choices by Race, Controlling for District

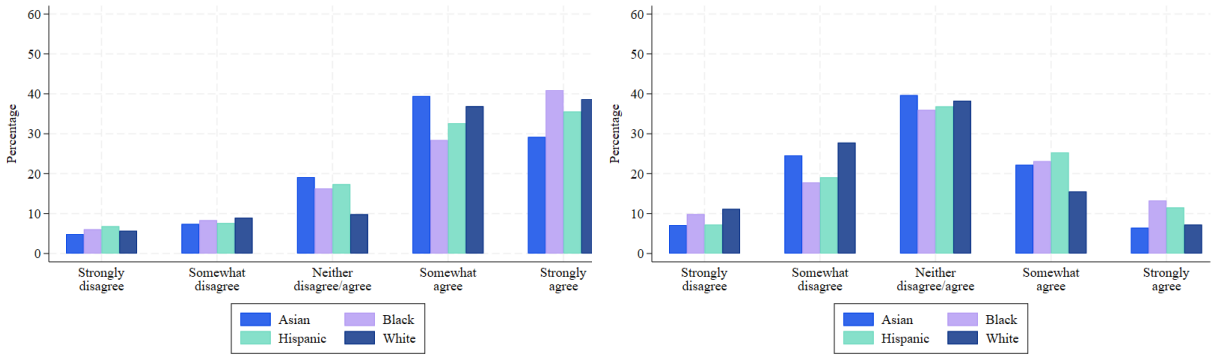


Notes: This figure shows how the racial gap varies across student baseline achievement for all students (panel a), and separately for students enrolled in majority Black and Hispanic middle schools and majority white and Asian middle schools (panel b). The two panels plot the coefficient on a dummy indicating Black and Hispanic applicants in regressions of school value-added in first choices on race and 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: Perceptions of Fitness in School

(a) Student would fit well in school with strong academics

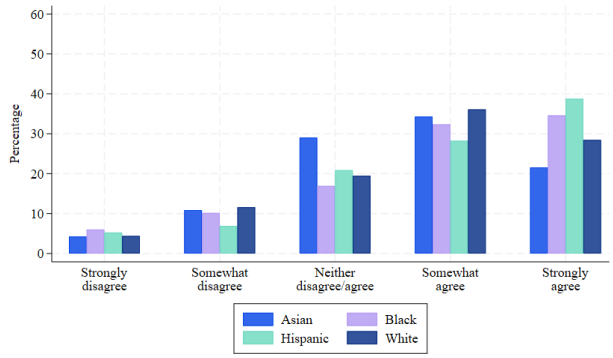
(b) Student would fit well with peers of different achievement



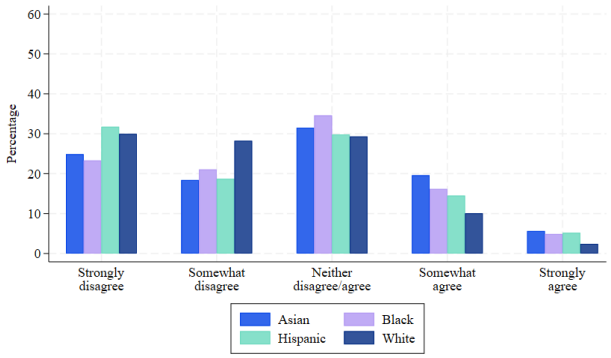
Notes: This figure shows the raw distribution of answers to survey questions asking respondents how well they feel they would fit within a school community, separately by respondent race.

Figure B4: Perceptions of Racial Belonging and Discrimination

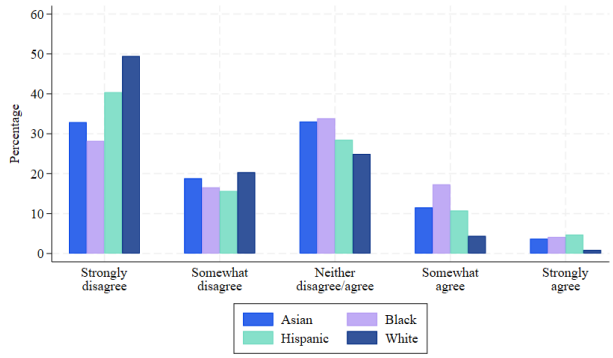
(a) Student would belong with peers of other race



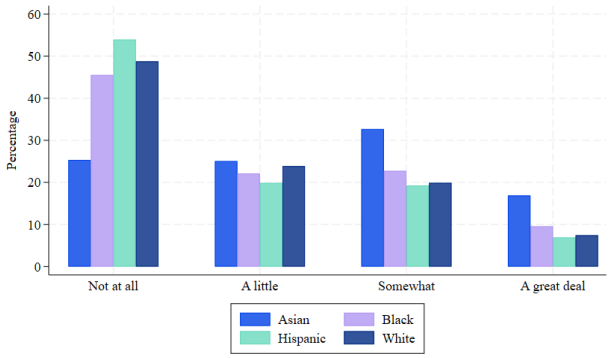
(b) Discrimination by classmates



(c) Discrimination by teachers



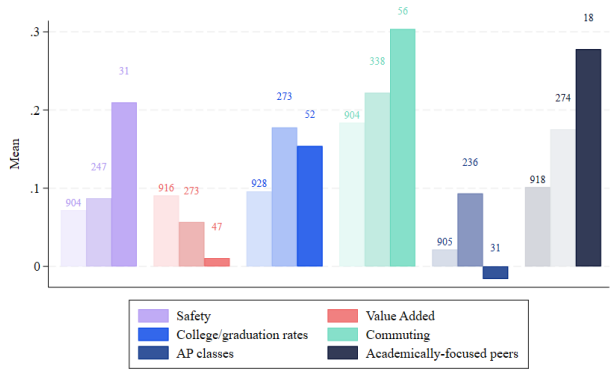
(d) Discrimination influenced choices



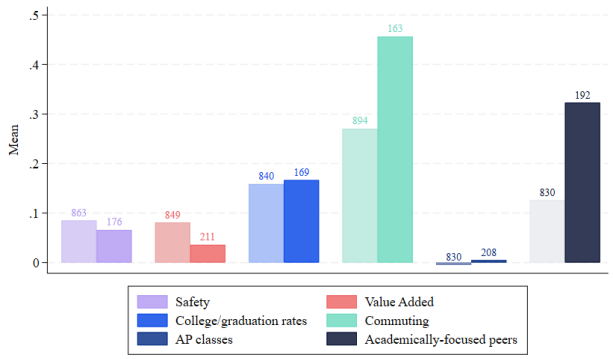
Notes: This figure shows the raw distribution of answers to survey questions asking respondents about perceptions of discrimination or about how well they feel they would belong to a school community, separately by respondent race.

Figure B5: Differences in Information Accuracy About School Characteristics if School is in Awareness Set

(a) Excess p(correct) - pairwise comparisons



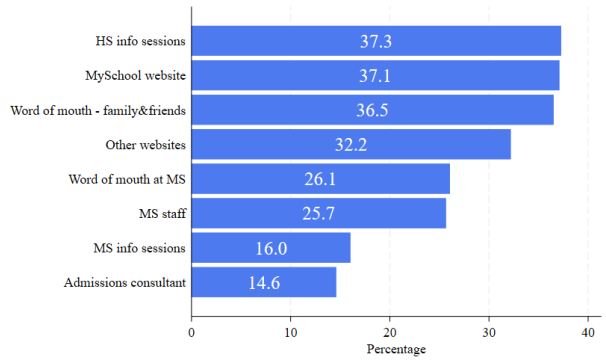
(b) Correlation between real quartile and answer



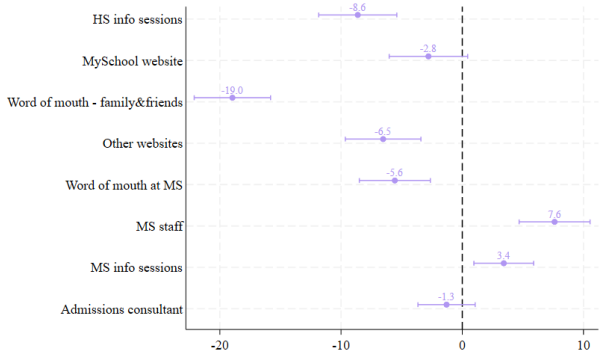
Notes: This figure reports measures of accuracy of information about school characteristics and how these vary for schools that are for sure in respondents' awareness sets. Panel (a) reports the percentage of respondents who responded correctly above 50% (which would correspond to random guesses only). Panel (b) reports the correlation of respondents' rankings with the true ranking of the school they were shown among schools in the same borough. For each school characteristics in panel (a), the first bar is for all questions, the second bar restricts the sample to questions in which one school in for sure known and the third bar to questions in which both schools are for sure known. For each school characteristics in panel (b), the first bar is for all questions, while the second bar restricts the sample to questions in which the school in for sure known. This figure uses data from survey questions Q10a-g.

Figure B6: Information sources

(a) Means



(b) Racial gaps



Notes: This figure shows racial differences in the use of different sources of information about high schools, which are listed on the left of the figure. It plots the regression coefficient of a dummy indicating Black and Hispanic respondents in separate regressions of indicators for having selected each source of information as one of the three most important on respondent race, controlling for district of residence and baseline achievement.

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. Finally, section C.5 includes all survey questions as shown to participants.

C.1 Survey logistics

C.1.1 Survey timeline

High school applications in NYC closed on December 5th, 2023. 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.

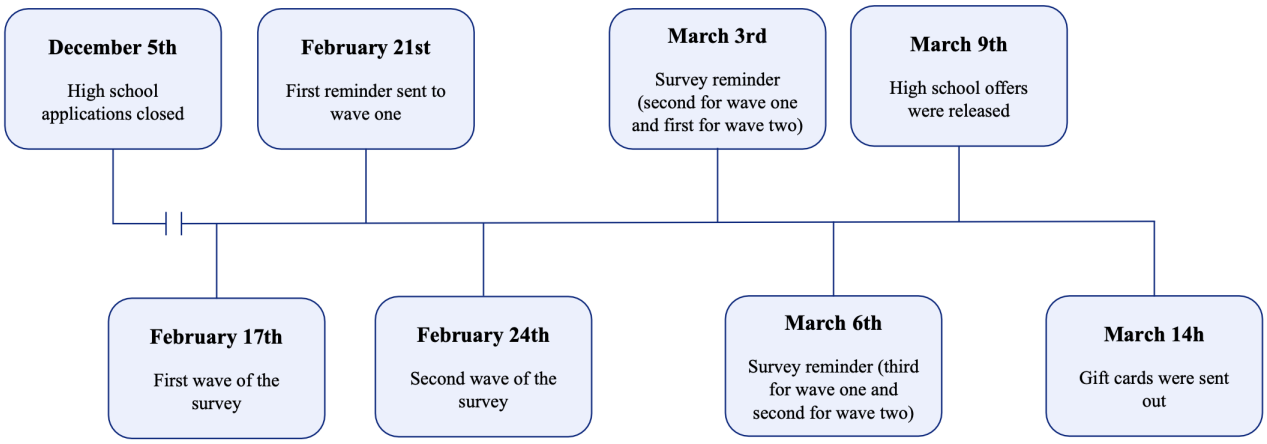
C.1.2 Survey emails

The survey was sent by email in the top three most spoken languages in NYC[30]: English, Spanish, and Simplified Chinese. Figures C2, C3, and C4 show the first email sent to the

[30]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.

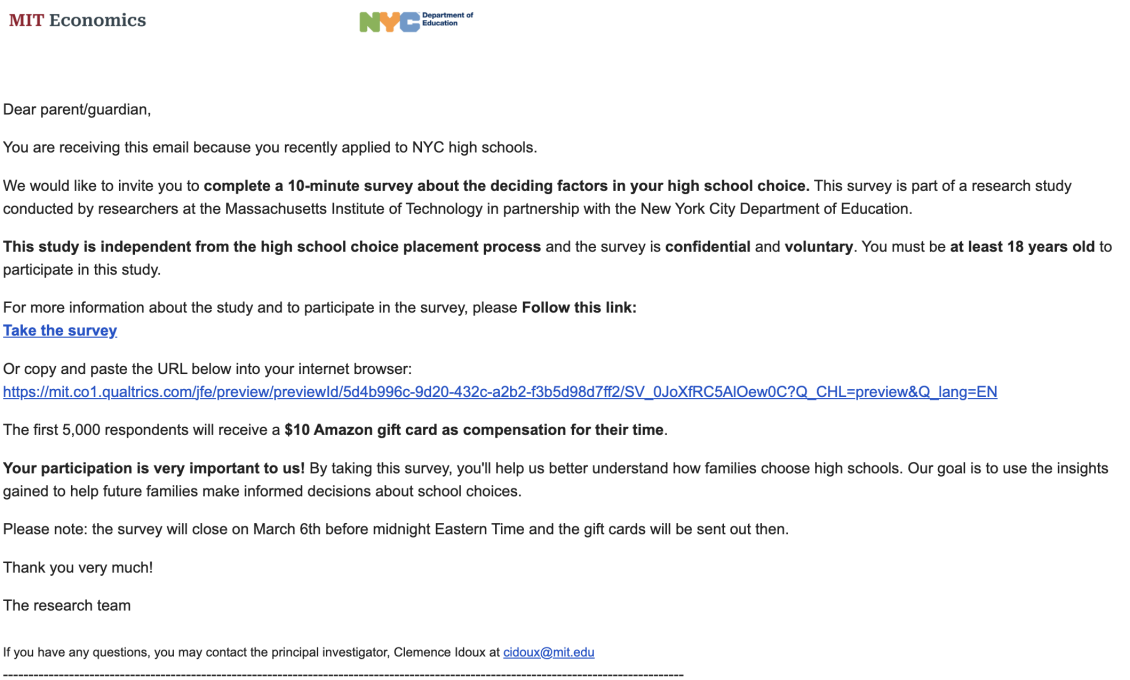
potential participants. Figures C5, C6, and C7 show the survey reminder email.

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

---

Figure C3: Invitation to Participate. Spanish Version

Estimado Padre/Tutor,

Está recibiendo este correo electrónico porque recientemente solicitó el ingreso a escuelas secundarias de la Ciudad de Nueva York.

Nos gustaría invitarlo a **completar una encuesta de 10 minutos sobre los factores decisivos en su elección de escuela secundaria**. Esta encuesta es parte de un estudio de investigación realizado por el Massachusetts Institute of Technology (MIT) con el Departamento de Educación de la Ciudad de Nueva York.

El estudio es **independiente del proceso de asignación a escuelas secundarias** y la encuesta es **confidencial** y **voluntaria**. Debe tener **al menos 18 años** para participar en este estudio.

Para obtener más información sobre el estudio y participar en la encuesta, **siga este enlace:**
Tome la encuesta

O copie y pegue el siguiente enlace (URL) en su navegador de Internet:
https://mit.co1.qualtrics.com/jfe/preview/previewId/5d4b996c-9d20-432c-a2b2-f3b5d98d7ff2/SV_0JoXfRC5AlOew0C?Q_CHL=preview&Q_lang=ES

Los primeros 5,000 encuestados recibirán **una tarjeta de regalo de Amazon de $10 dólares como pago por su tiempo.**

¡Su participación es muy importante para nosotros! Al completar esta encuesta, usted nos ayudará a entender mejor cómo las familias eligen escuelas secundarias. Nuestro objetivo es usar los conocimientos adquiridos para ayudar a futuras familias a tomar decisiones informadas sobre sus opciones de escuelas.

Tenga en cuenta que la encuesta cerrará el 6 de Marzo antes de la media noche, hora del este.

¡Muchas gracias!

El equipo de investigación

Si usted tiene alguna pregunta, puede contactar a la investigadora principal, Clemence Idoux en el correo electrónico cidoux@mit.edu

Figure C4: Invitation to Participate. Simplified Chinese Version

尊敬的家长和监护人:

您收到此调查问卷是因为您最近申请了纽约市的高中。

我们想请您**填写一份10分钟的有关您高中选择的决定因素的调查问卷**。这份调查问卷是麻省理工学院与纽约市教育部开展的调研的一部分。

本调研独立于高中择校流程。本调查问卷**为自愿参加**并且您通过本调查问卷**向我们提供的信息将予以保密**。您必须**年满18岁**来参与本次调研。

了解有关本调研的更多信息和填写调查问卷，**请使用以下链接**:
填写问卷

或复制粘贴以下链接至您的网络浏览器:
https://mit.co1.qualtrics.com/jfe/preview/previewId/5d4b996c-9d20-432c-a2b2-f3b5d98d7ff2/SV_0JoXfRC5AlOew0C?Q_CHL=preview&Q_lang=ZH-T

调查问卷的前 5,000位回答者将会获得一**张价值10美元的亚马逊礼品卡，为答谢您为此付出的时间。**

您的参与对我们非常重要！ 通过填写这份调查， 您将会帮助我们更好地了解家庭如何选择高中。我们的目标是用我们获得的了解来帮助未来的家庭在择校中做出明智的决定。

请注意：调查问卷会在美东时间3月6日午夜前关闭并且礼品卡会届时发出。

谢谢!

研究团队

如果您有任何问题，您可以通过邮箱cidoux@mit.edu 联系项目负责人Clemence Idoux 。

Follow the link to opt out of future emails:
Click here to unsubscribe

Figure C5: Survey Reminder. English Version

Dear parent/guardian,

You have 3 days left to **complete a 10-minute survey about the deciding factors in your high school choice.**

Your opinion matters! By taking this survey, you'll help us better understand how families choose high schools and improve the application experience for future families.

To compensate you for your time, **you will receive a $10 Amazon gift card**, after the survey is closed on March 6th.

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/81e99fbf-5fc8-43f4-a64b-fa87bf0f6fb4/SV_4OysVkOlp8xjaZM?Q_CHL=preview&Q_lang=EN

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.

Thank you very much! Your participation is very important to us.

The research team

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

Figure C6: Survey Reminder. Spanish Version

Estimado Padre/Tutor,

Le quedan 3 días para completar **una encuesta de 10 minutos sobre los factores decisivos en su elección de escuela secundaria.**

¡Su participación es muy importante! Al completar esta encuesta, usted nos ayudará a entender mejor cómo las familias eligen escuelas secundarias y a mejorar la experiencia de solicitud de ingreso de futuras familias.

Como compensación por su tiempo, **usted recibirá una tarjeta de regalo de Amazon de $10 dólares** después de que la encuesta cierre el 6 de Marzo.

Para obtener más información sobre el estudio y participar en la encuesta, **siga este enlace:**
Tome la encuesta

O copie y pegue el siguiente enlace (URL) en su navegador de Internet:
https://mit.co1.qualtrics.com/jfe/preview/previewId/81e99fbf-5fc8-43f4-a64b-fa87bf0f6fb4/SV_4OysVkOlp8xjaZM?Q_CHL=preview&Q_lang=ES

Esta encuesta es parte de un estudio de investigación realizado por el Massachusetts Institute of Technology (MIT) con el Departamento de Educación de la Ciudad de Nueva York. El estudio es **independiente del proceso de asignación a escuelas secundarias** y la encuesta es **confidencial** y **voluntaria**. Debe tener **al menos 18 años** para participar en este estudio.

¡Muchas gracias! Su participación es muy importante para nosotros.

El equipo de investigación

Si usted tiene alguna pregunta, puede contactar a la investigadora principal, Clemence Idoux en el correo electrónico cidoux@mit.edu
---

Figure C7: Survey Reminder. Simplified Chinese Version

尊敬的家长和监护人:

您还有3天时间**填写一份10分钟的有关您高中选择的决定因素的调查问卷。**

您的意见很重要！ 通过填写这份调查，您将会帮助我们更好地了解家庭如何选择高中并改善未来家庭的择校体验。

为答谢您为此付出的时间，您将会获得**一张价值10美元的亚马逊礼品卡**，在三月六日调查问卷关闭之后。

了解有关本调研的更多信息和填写调查问卷，**请使用以下链接**:
填写问卷

或复制粘贴以下链接至您的网络浏览器:
https://mit.co1.qualtrics.com/jfe/preview/previewId/81e99fbf-5fc8-43f4-a64b-fa87bf0f6fb4/SV_4OysVkOlp8xjaZM?Q_CHL=preview&Q_lang=ZH-T

这份调查问卷是麻省理工学院与纽约市教育部开展的调研的一部分。**本调研独立于高中择校流程。**本调查问卷**为自愿参加**并且您通过本调查问卷**向我们提供的信息将予以保密。**您必须**年满18岁**来参与本次调研。

谢谢！ 您的参与对我们非常重要。

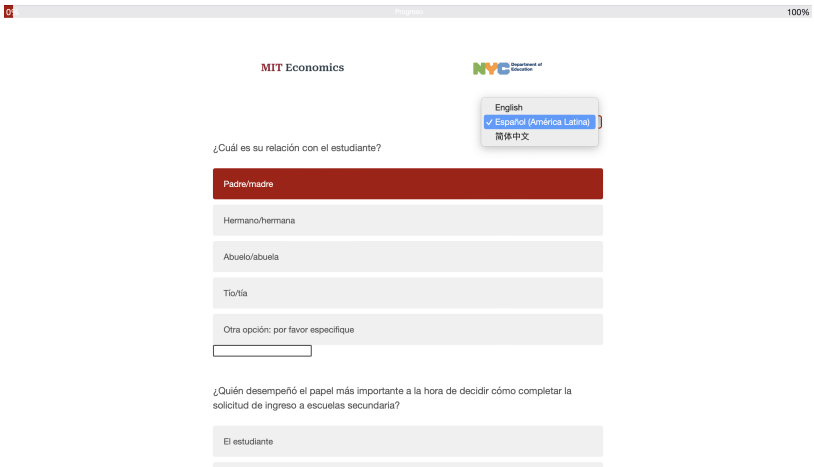
研究团队

如果您有任何问题，您可以通过邮箱cidoux@mit.edu 联系项目负责人Clemence Idoux 。

C.1.3 Qualtrics design

The survey was designed on Qualtrics and it was available in English, Spanish, and Simplified Chinese (see Figure C8). 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.

Figure C8: Language Selection in the Survey

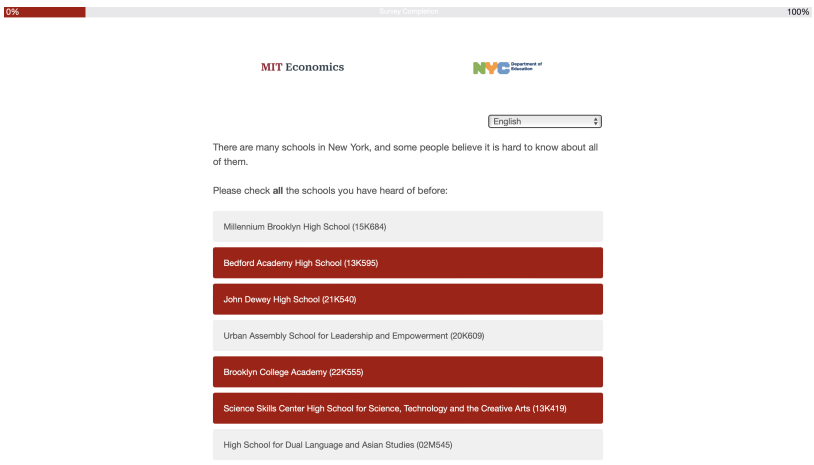


Notes: This figure illustrates how survey participants could change the survey language at any point when filling it out.

We personalized the survey by using JavaScript 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 Figure C9 and section C.4 of the main text for more details.

Figure C9: 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

C.2.1 Assignment of participants to survey waves

The sample of all eligible participants consisted of parents or guardians of students applying to start 9th grade in fall 2023. Participant population data was selected as described in Section 4 of the paper. Summary statistics can be found in Table 1. 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, and has all baseline scores.

3. Survey priority is 3 if the student was enrolled in a DOE school in 6th grade, is at risk of middle school assignment, and is missing some demographic information or baseline scores.

4. Survey priority is 4 if any of the following is true:

    - The student was enrolled in a DOE school in 6th grade and is missing some demographic information or baseline scores.
    - The student has disabilities, was enrolled in a DOE school in 6th grade, and has any risk of middle school assignment.

5. Survey priority is 5 if any of the following is true:

    - The student was enrolled in a DOE school in 6th grade and does not have risk of middle school assignment.
    - The student has disabilities, was enrolled in a DOE school in 6th grade, and is missing risk of middle school assignment.

6. Survey priority is 6 if the student was not enrolled in a DOE school in 6th grade.

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. Additional waves of participants were created for potential expansion of the survey, although they were not used. The final sample of potential participants comprised 21,401 parents or guardians.

C.3 Survey block design and randomization to survey version

C.3.1 Description of survey blocks

The complete survey had a total of 47 questions, including the consent question and the end-of-survey comment box. We grouped all the questions by type and created five different blocks (1a, 1b, 2, 3a, 3b), which are shown in Table C1. 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 5.1.2 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.

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 C2). All survey versions included the consent to participate, general questions, and the end-of-survey comment box. All potential participants had an equal probability of receiving any of the eight survey versions (12.5% each). The marginal probability for each block was thus 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 C2).

Table C3 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 C1: 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 5.1.2). 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 5.1.2). 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).

Notes: This table presents the five distinct question blocks in the survey, including a general one. Each block groups different types of questions, as shown in the first column. The last column provides a breakdown of each question type, including the total number of questions, the questions format (checkbox, open-ended, or multiple choice), and the question numbers in the survey.

Table C2: 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

Notes: This table shows the survey blocks included in each of the eight survey versions. Specific questions within each block are detailed in Table C1. The third column provides the total number of questions for each survey version.

Table C3: Survey Attrition and Covariate Balance

Dependent Variable	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		
	All (1)	All (2)	W&A (3)	B&H (4)	All (5)	W&A (6)	B&H (7)	All (8)	W&A (9)	B&H (10)	All (11)	W&A (12)	B&H (13)	All (14)	W&A (15)	B&H (16)	All (17)	W&A (18)	B&H (19)	All (20)	W&A (21)	B&H (22)	All (23)	W&A (24)	B&H (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 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 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 SAT 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 (low-white) schools. If empty again, we selected the school with the highest (lowest) share of white students among high-VA schools. Finally, we randomized at the student level between these two high VA schools.

- **School 7:** A school with low SAT 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

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 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 (low-white) schools. If empty again, we selected the school with the highest (lowest) share of white 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: baseline test scores, college enrollment rates, Regents VA, college VA, language and ap stem classes. For each district and each school characteristic, we selected four pairs of schools:

1. Both are high-white-share
2. Both are non-high-white share
3. The first is high-white and the second is not
4. The second is high-white 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 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.
- **College VA:** Different by at least 5pp.
- **Language classes:** Discrete difference (at least 1 more/less class).
- **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 description of the vignette experiment is on Section 5.1.2 of the paper. The experiment includes a total of 24 possible vignettes, also referred to as school cards. First, 16 school cards show 3 school characteristics (called the "3-factor list"): academics (x2), safety (x2), and racial composition (x4). Second, we have 8 school cards that show 2 school characteristics (called the "2-factor list"): safety (x2) and racial composition (x4). Regarding academics and safety, hypothetical schools had either high-safety or low-safety ratings. In terms of student demographics, hypothetical schools had either a balanced racial composition representative of the school district, a majority of Black students, a majority of Hispanic students, or a

majority of white or Asian students. The average school characteristics are in Table C4. Examples of the 2- and 3-factor cards are in Figure C10.

Table C4: 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 on 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%			

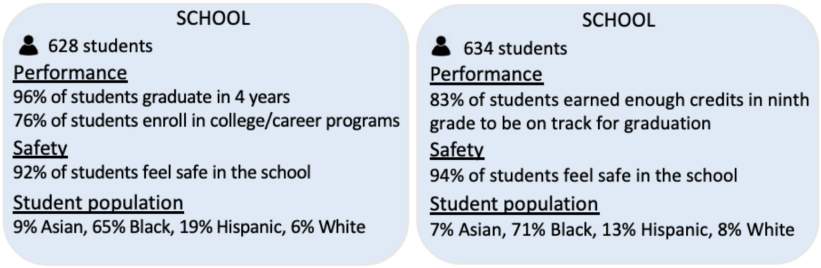
Notes: This table reports the characteristics of the school cards presented to respondents in the vignette experiments (questions Q17 and Q18).

For question 17, we randomly selected one school from the 3 factor list and one school from the 2 factor list for each student. In question 18 (relative scale, ranking of 3 schools), we randomly selected three schools from the 3 factor list and three schools from the 2 factor list, without replacement, for each student.

Finally, we randomized at the student level whether the student would receive the vignette with three or with two factors. We assigned 60% of students to the 3 factor version of the questions.

C.5 Survey images

Figure C10: 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).

Figure C11: Consent Question

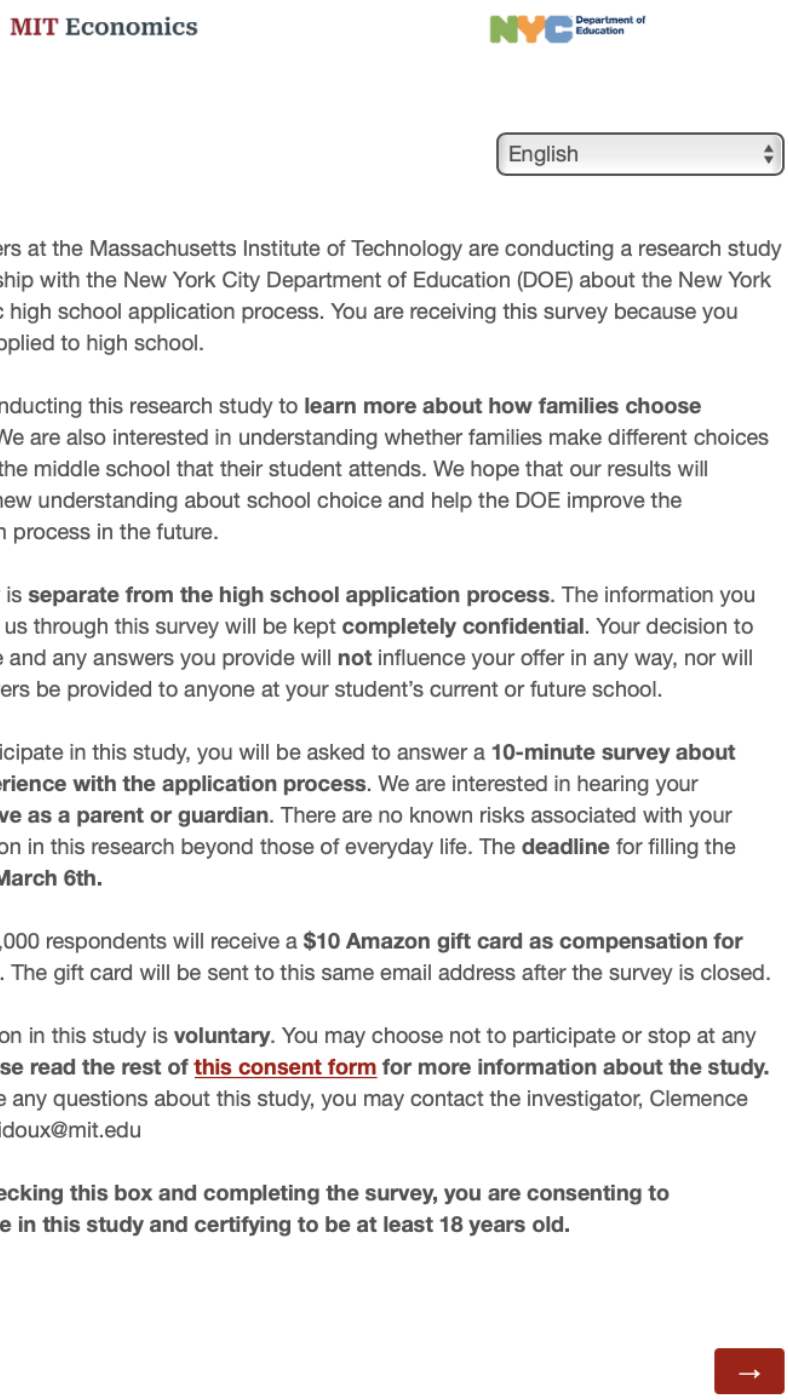


Figure C12: Question 1

Figure C13: Question 2

Figure C14: Question 3

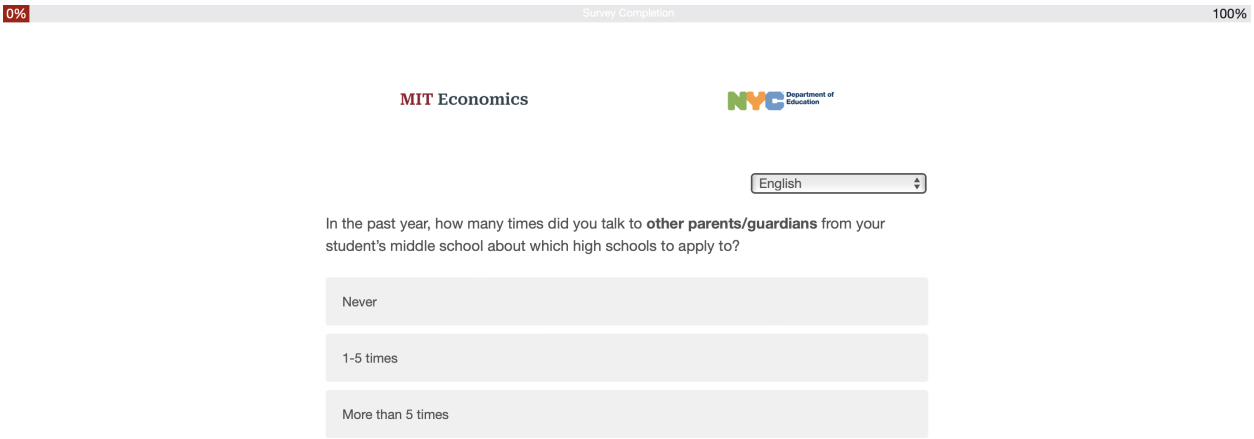


Figure C15: Question 3b

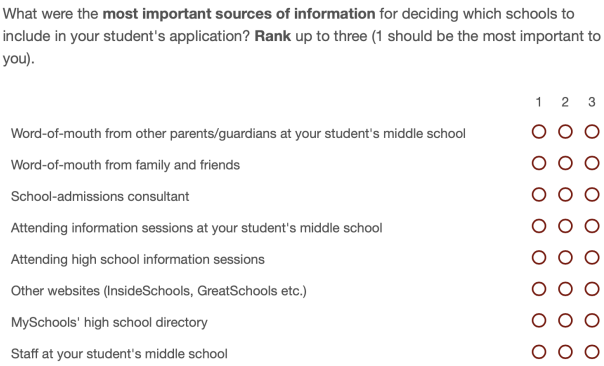


Figure C16: Question 3b.2

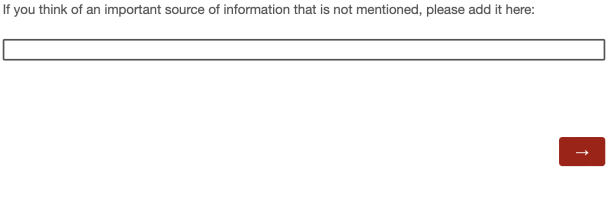


Figure C17: Question 4

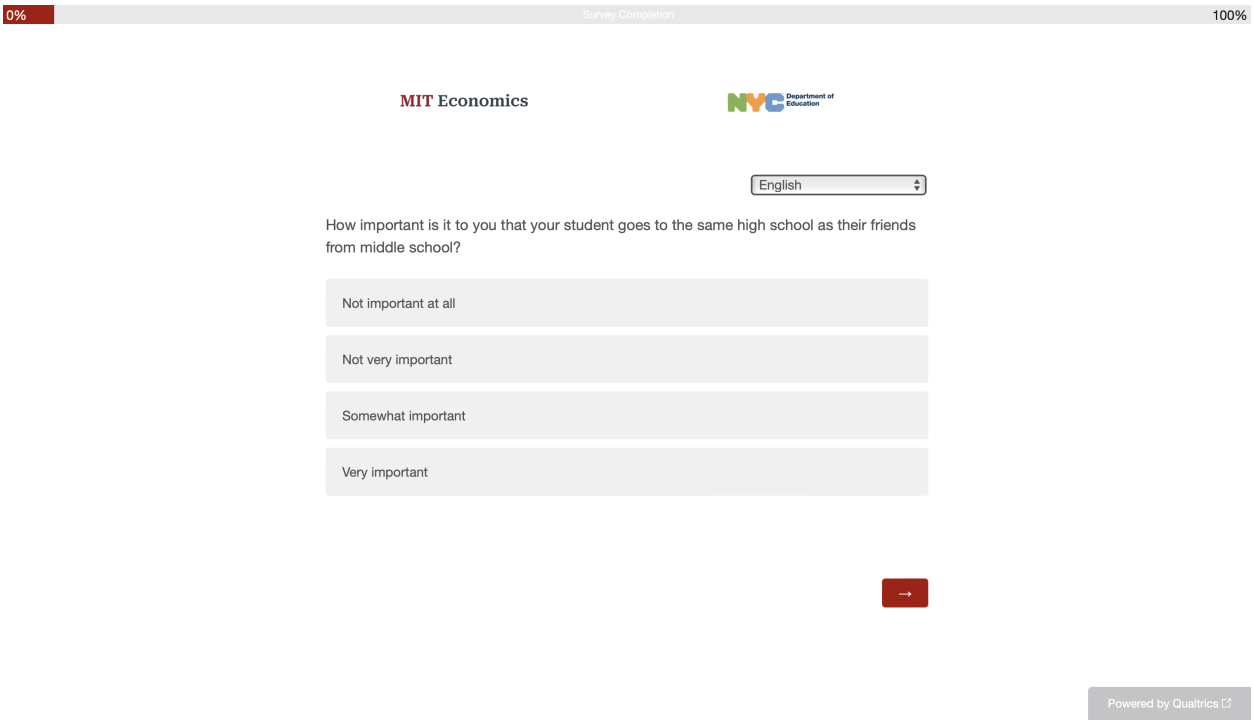


Figure C18: Question 5

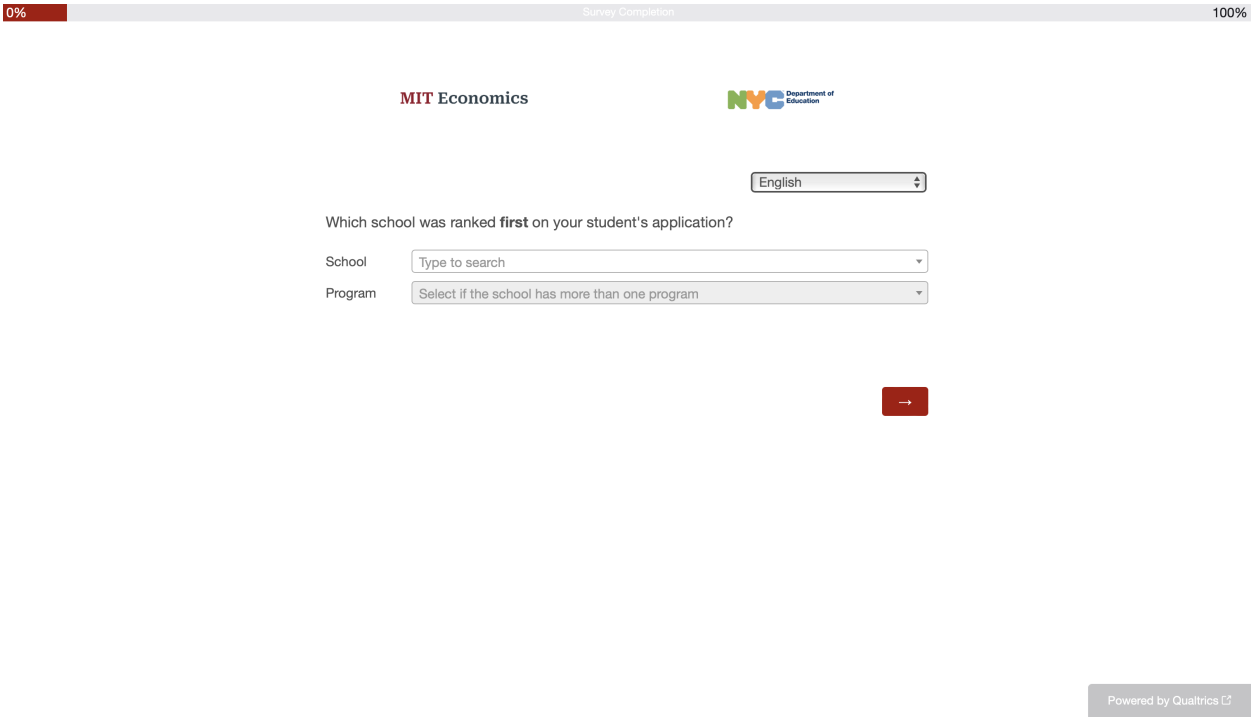


Figure C19: Question 6

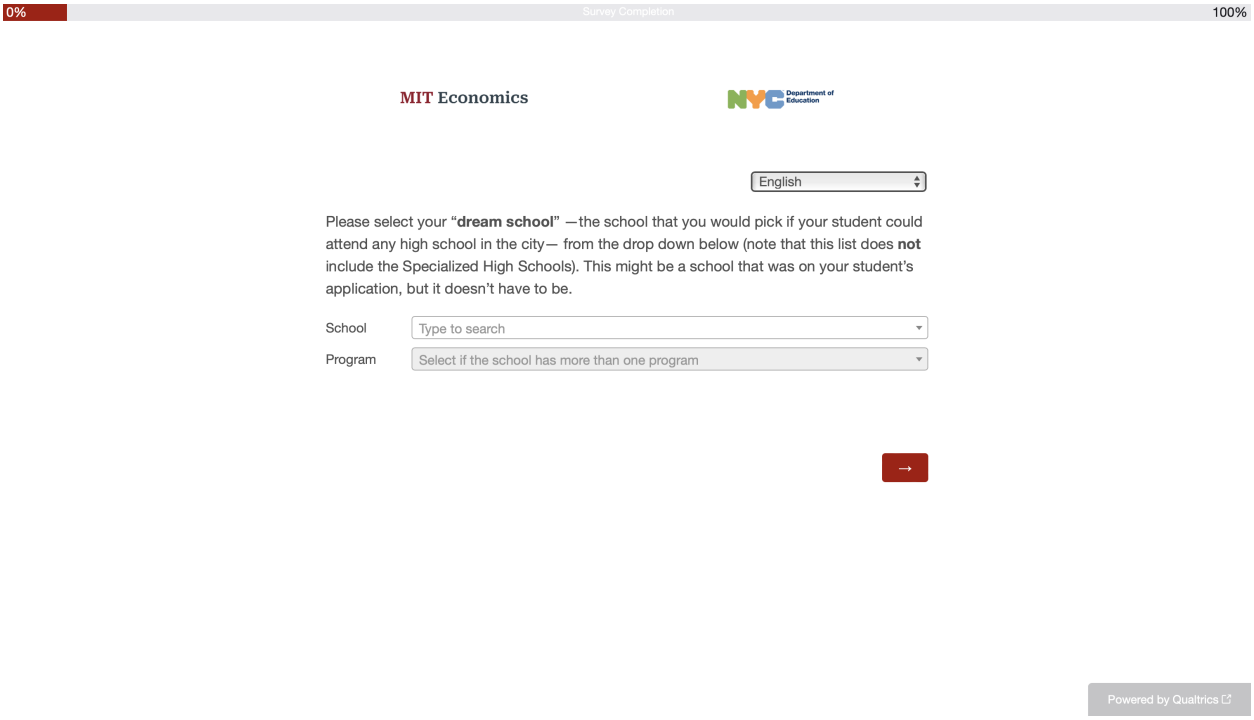


Figure C20: Question 7a

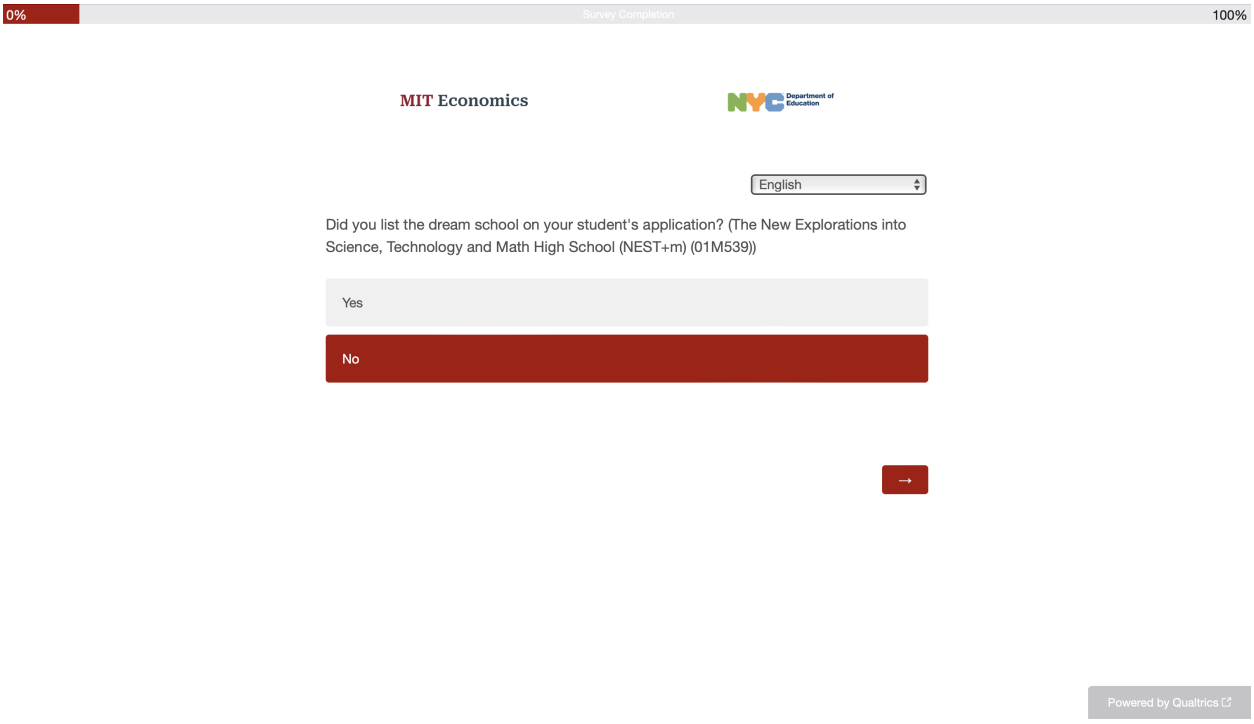


Figure C21: Question 7b

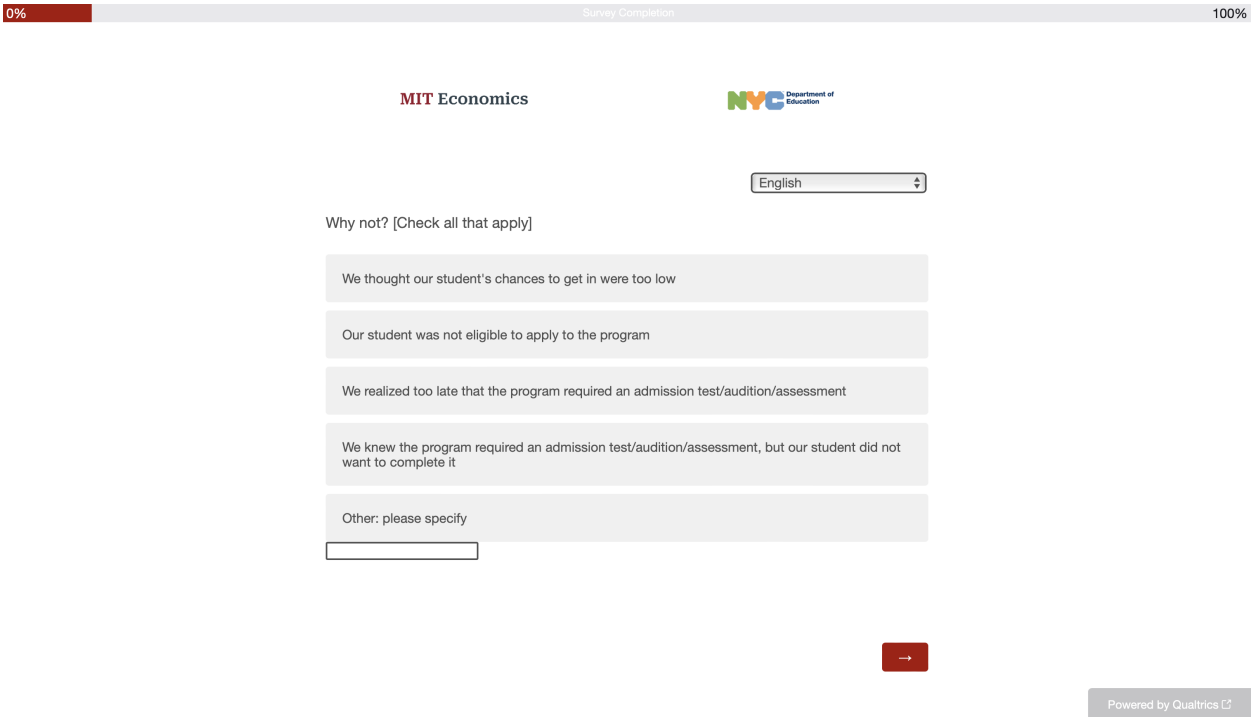


Figure C22: Question 7c

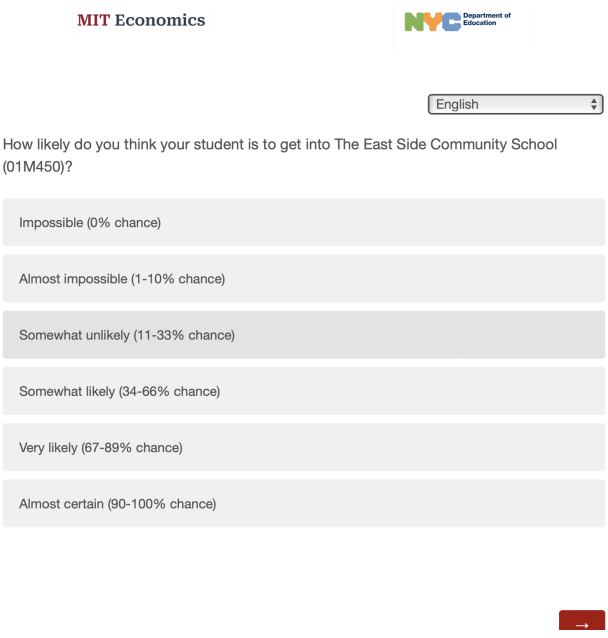


Figure C23: Question 7c.2

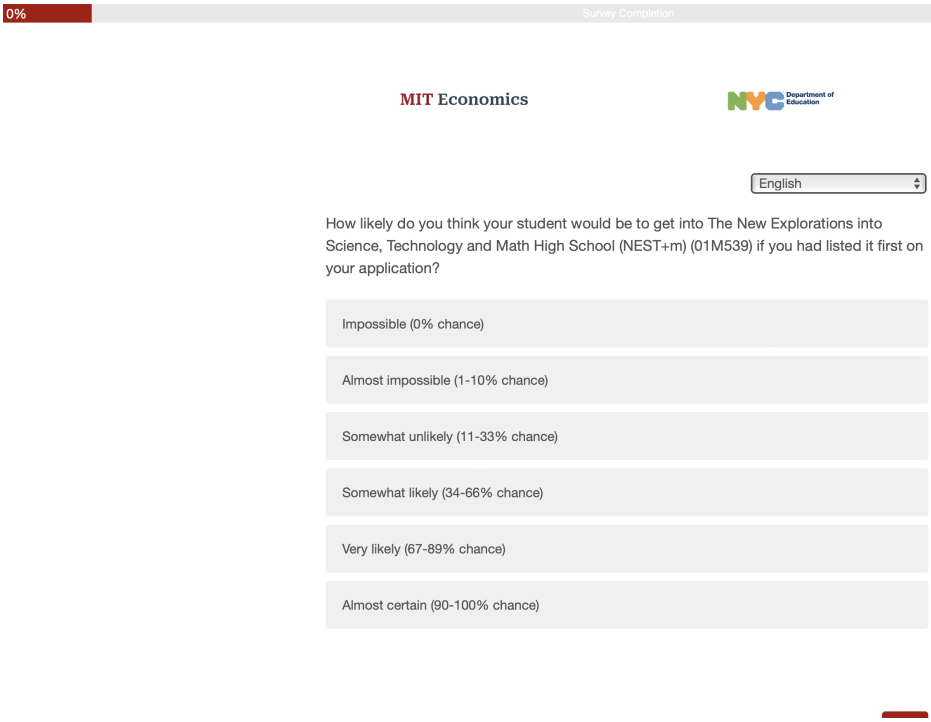


Figure C24: Question 8a

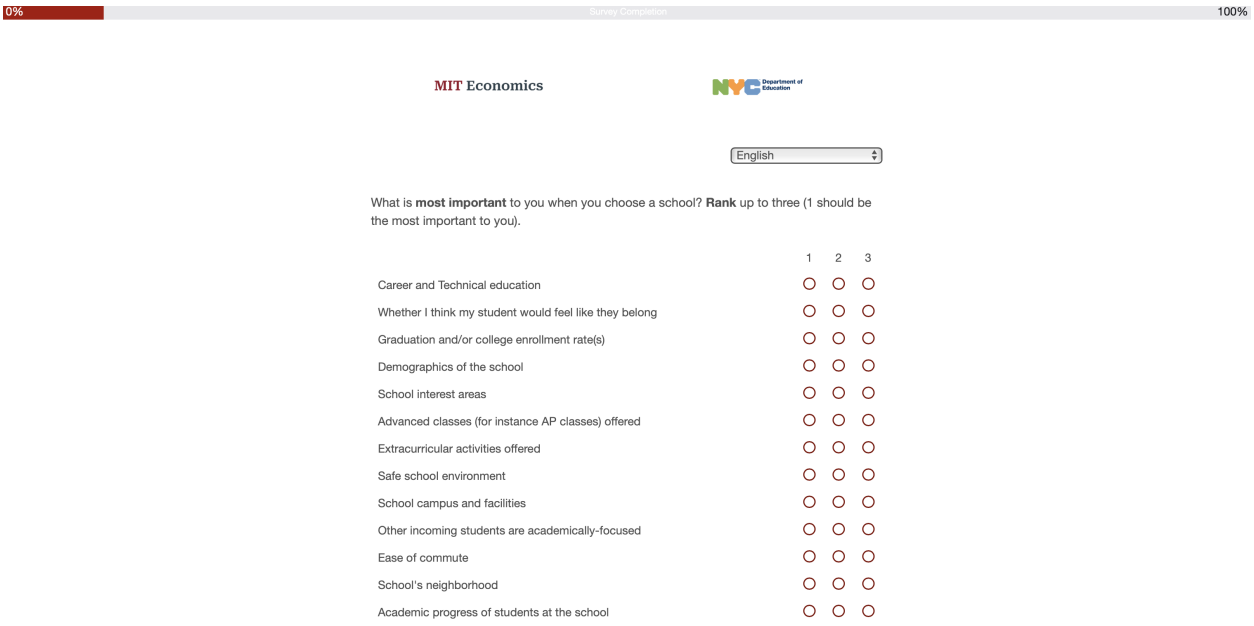


Figure C25: Question 8a.2

Figure C26: Question 9

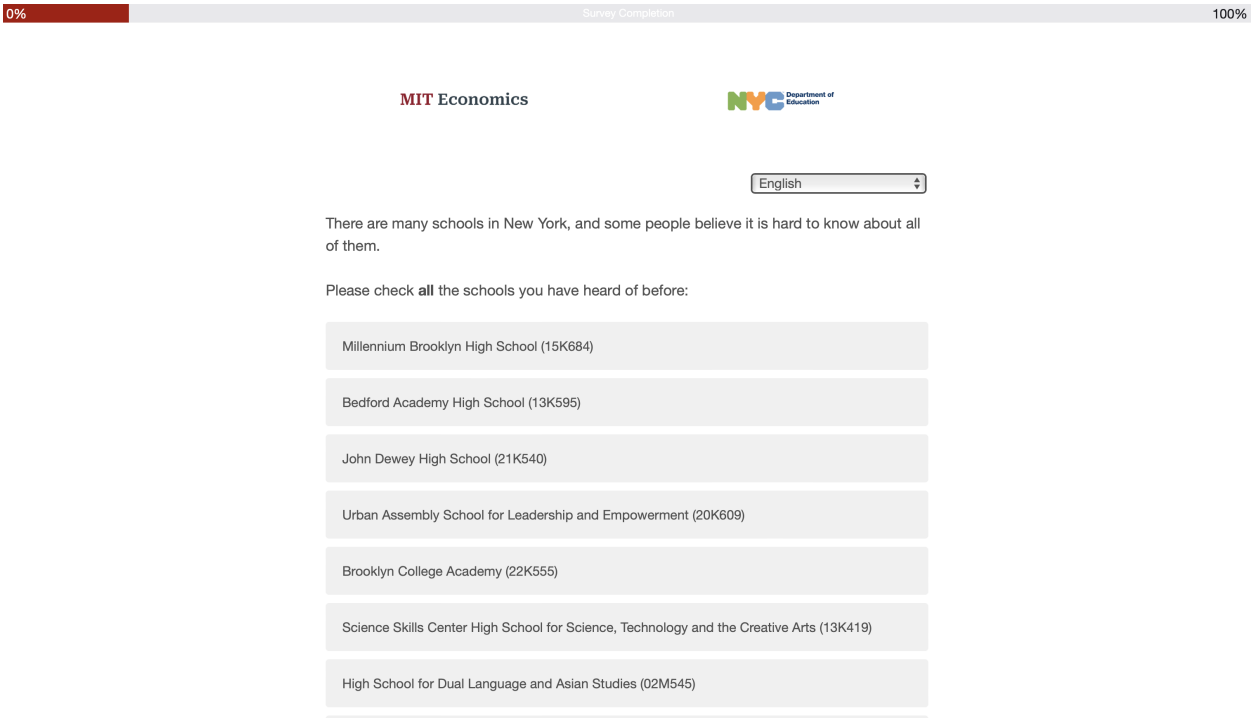


Figure C27: Question 10

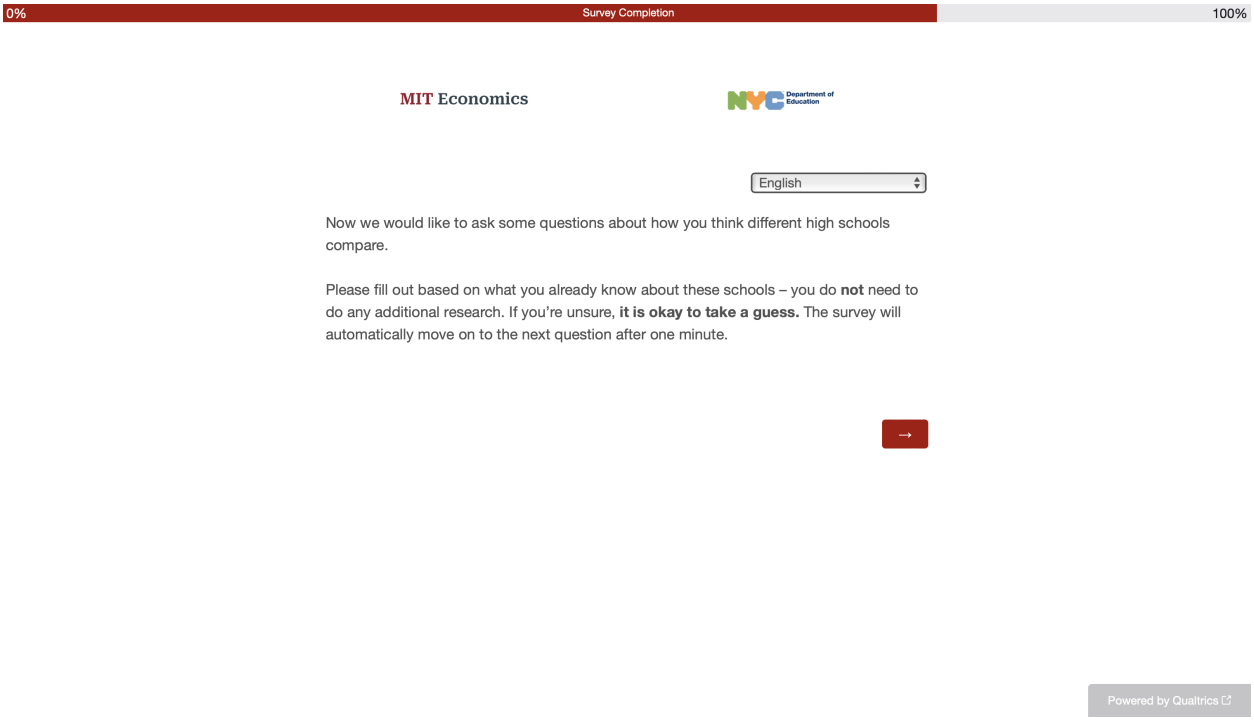


Figure C28: Question 10a

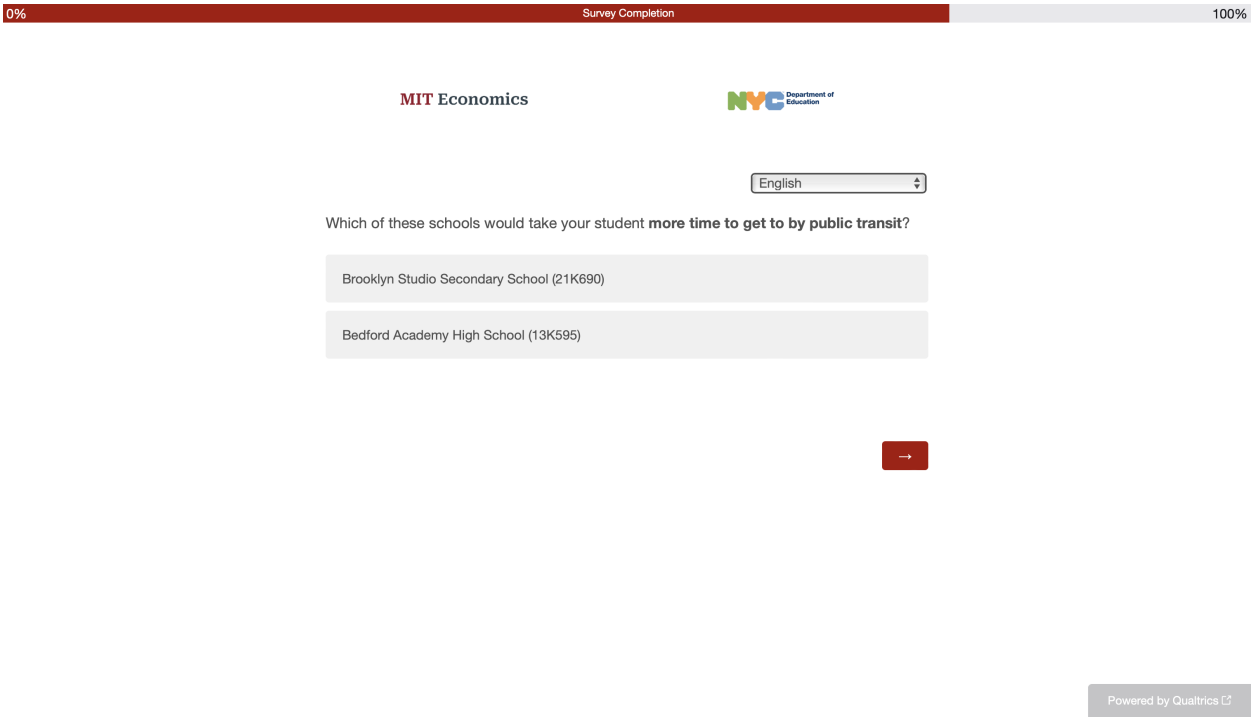


Figure C29: Question 10b

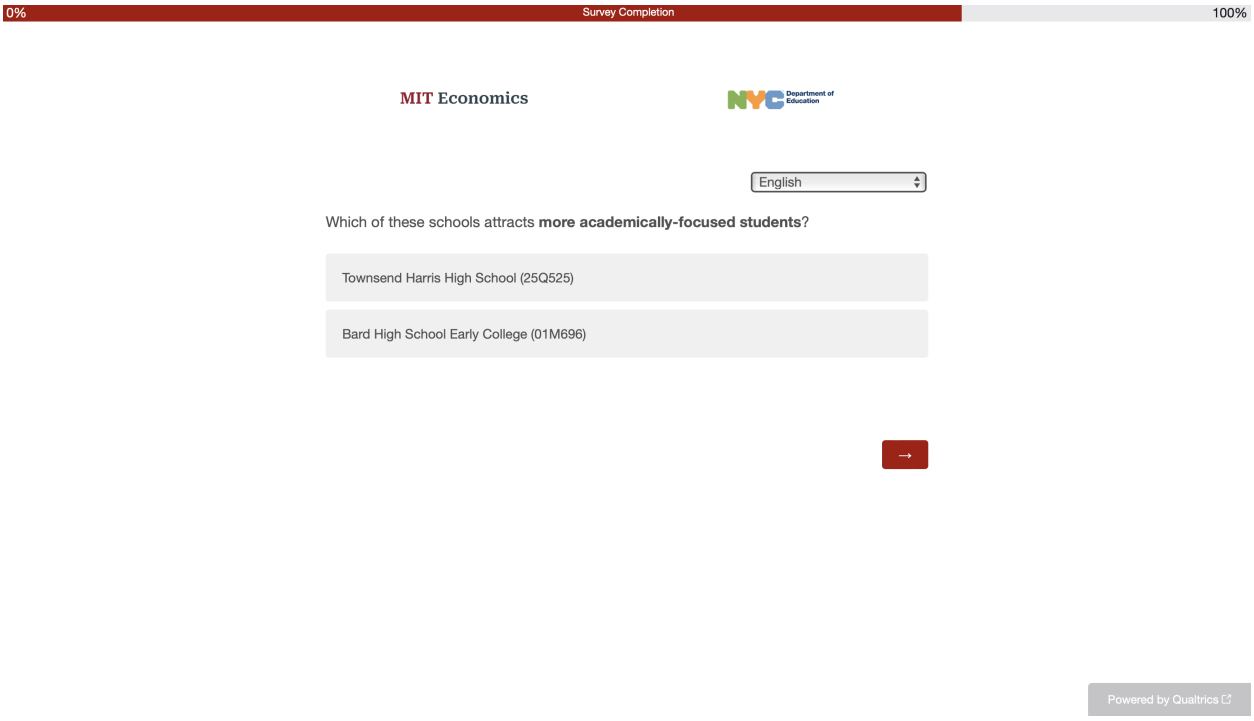


Figure C30: Question 10c

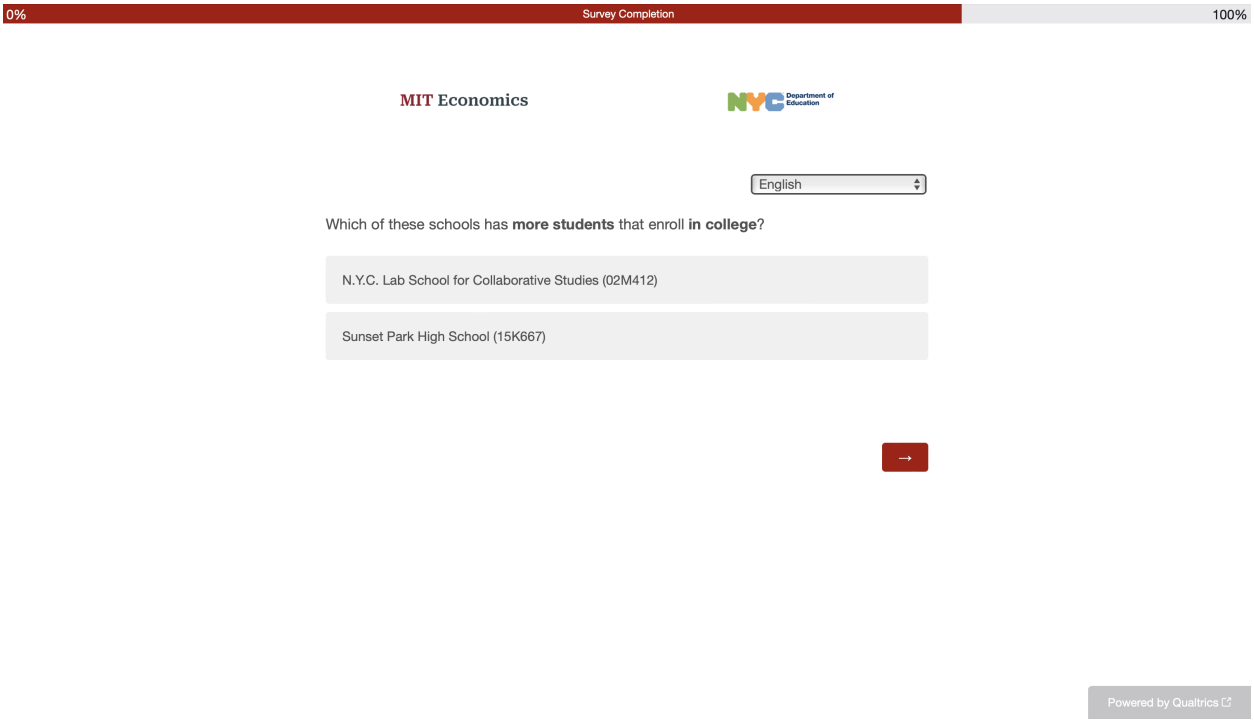


Figure C31: Question 10d

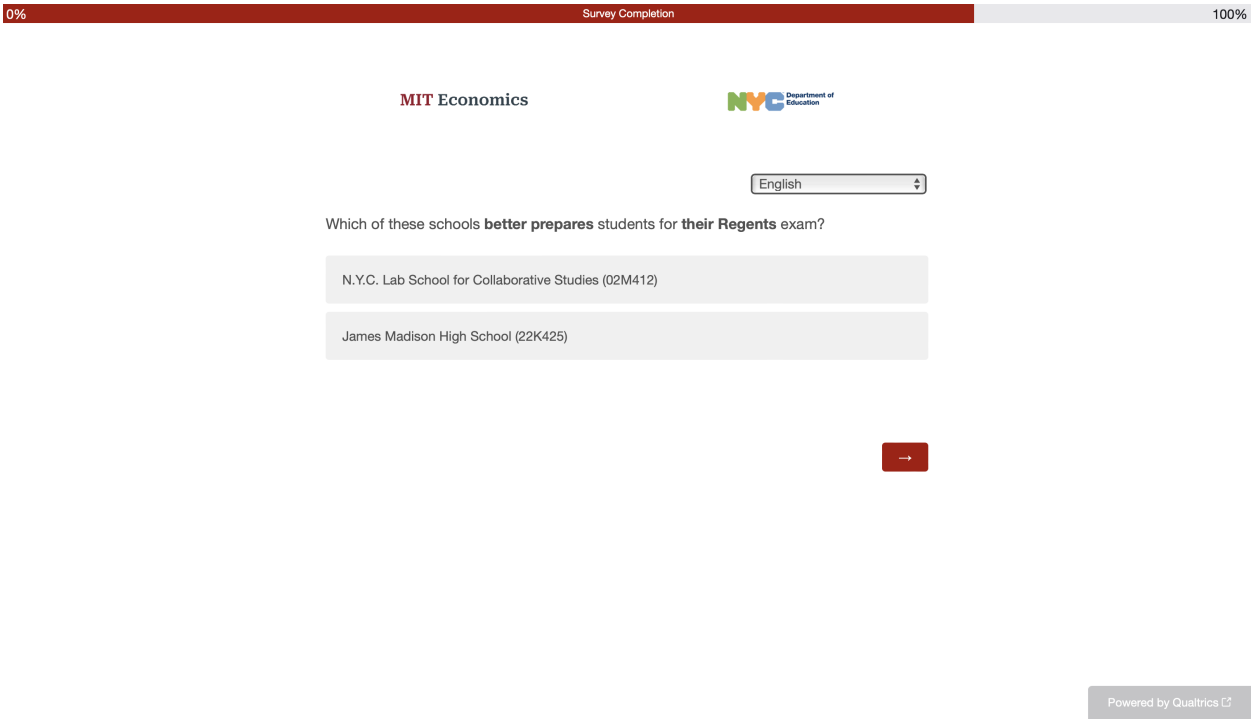


Figure C32: Question 10f

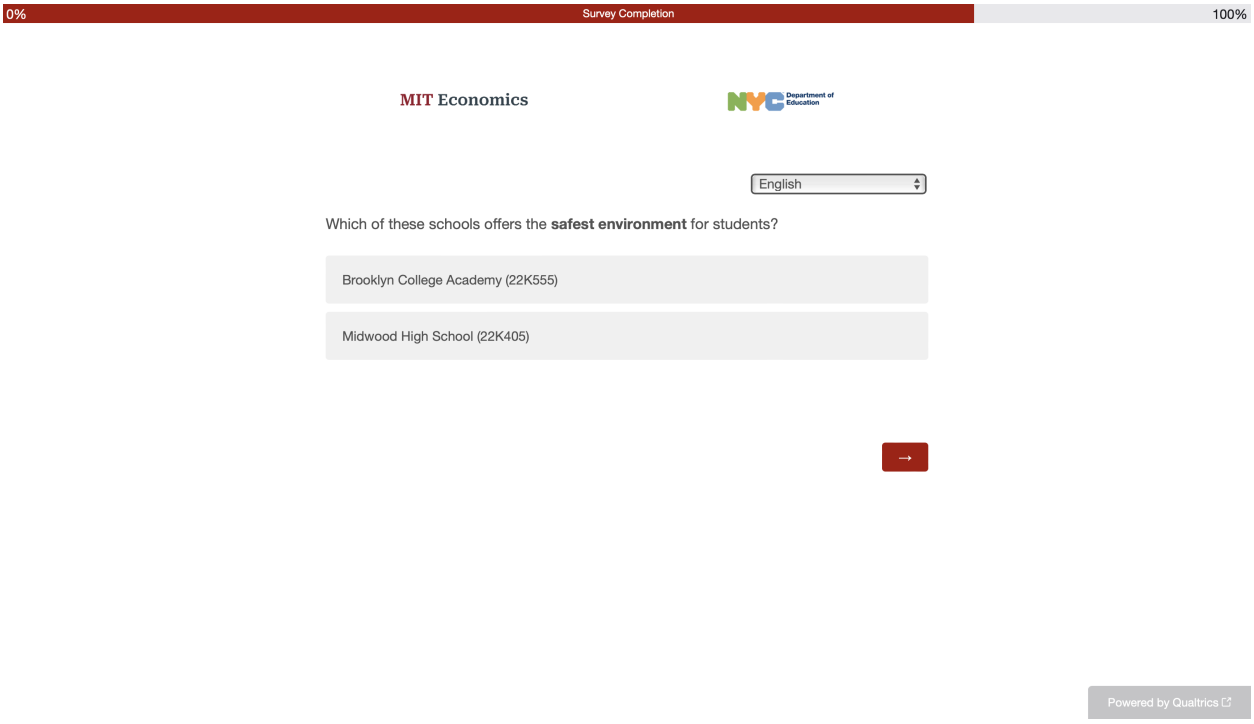


Figure C33: Question 10g

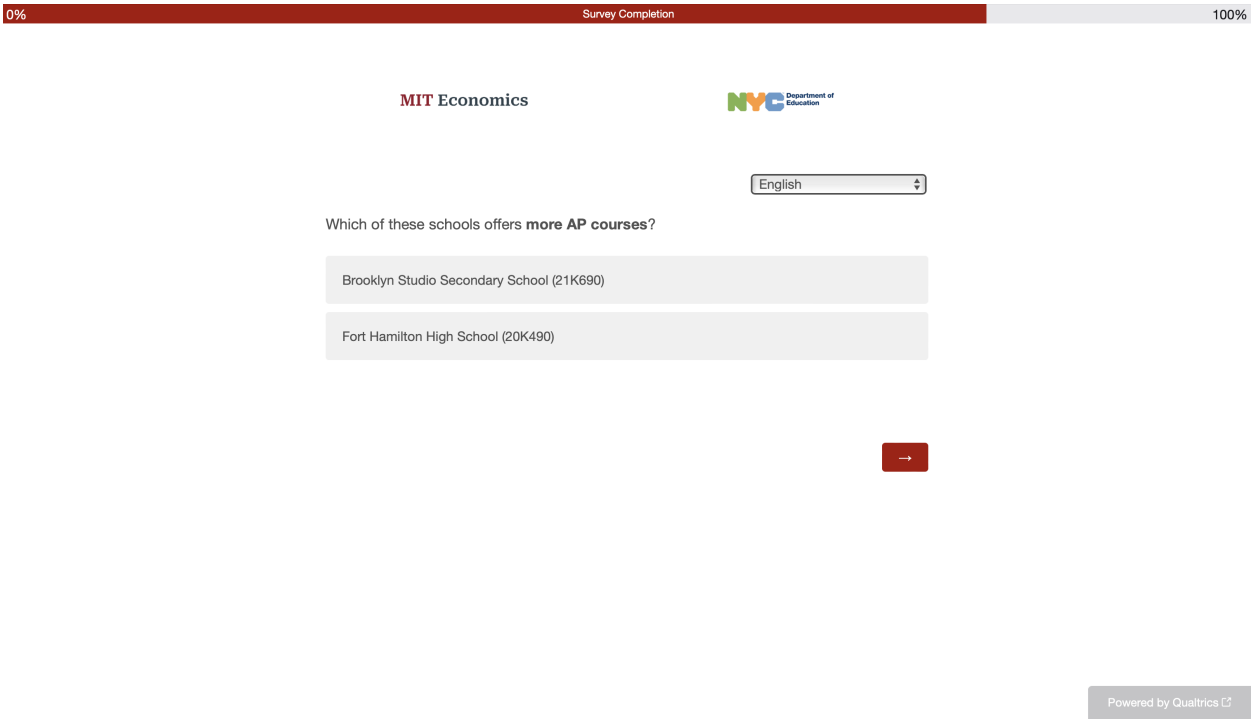


Figure C34: Question 11a

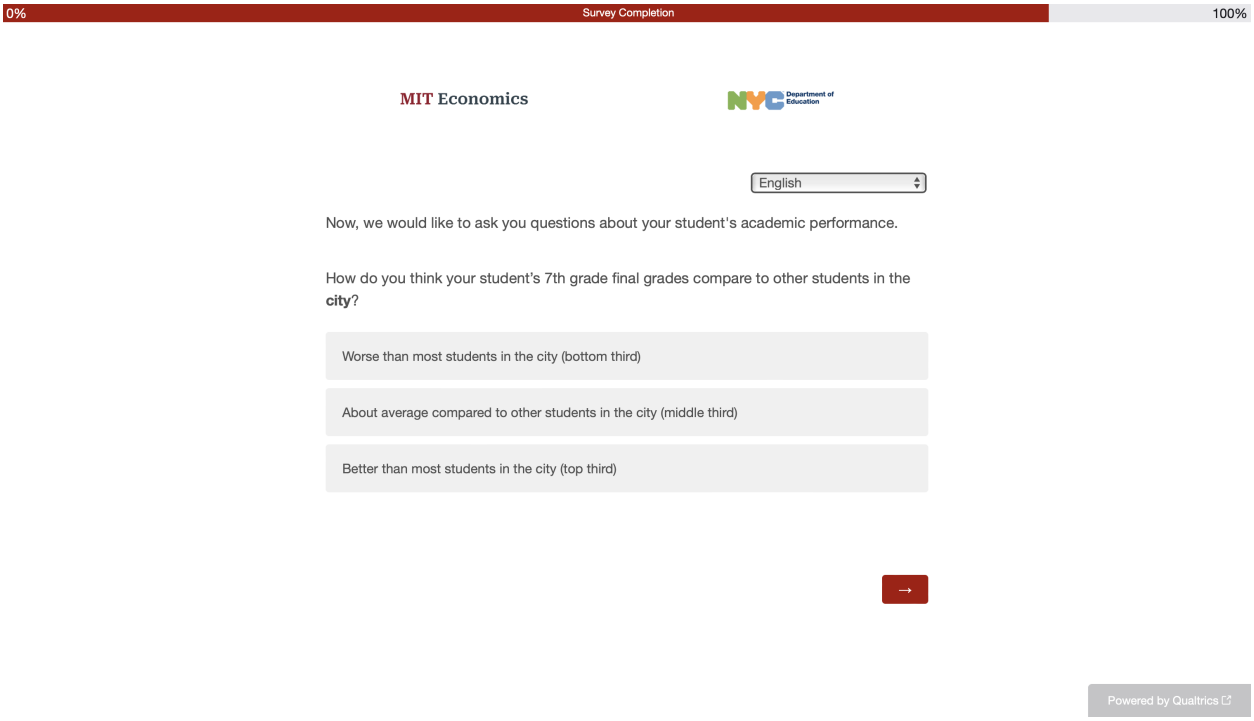


Figure C35: Question 12

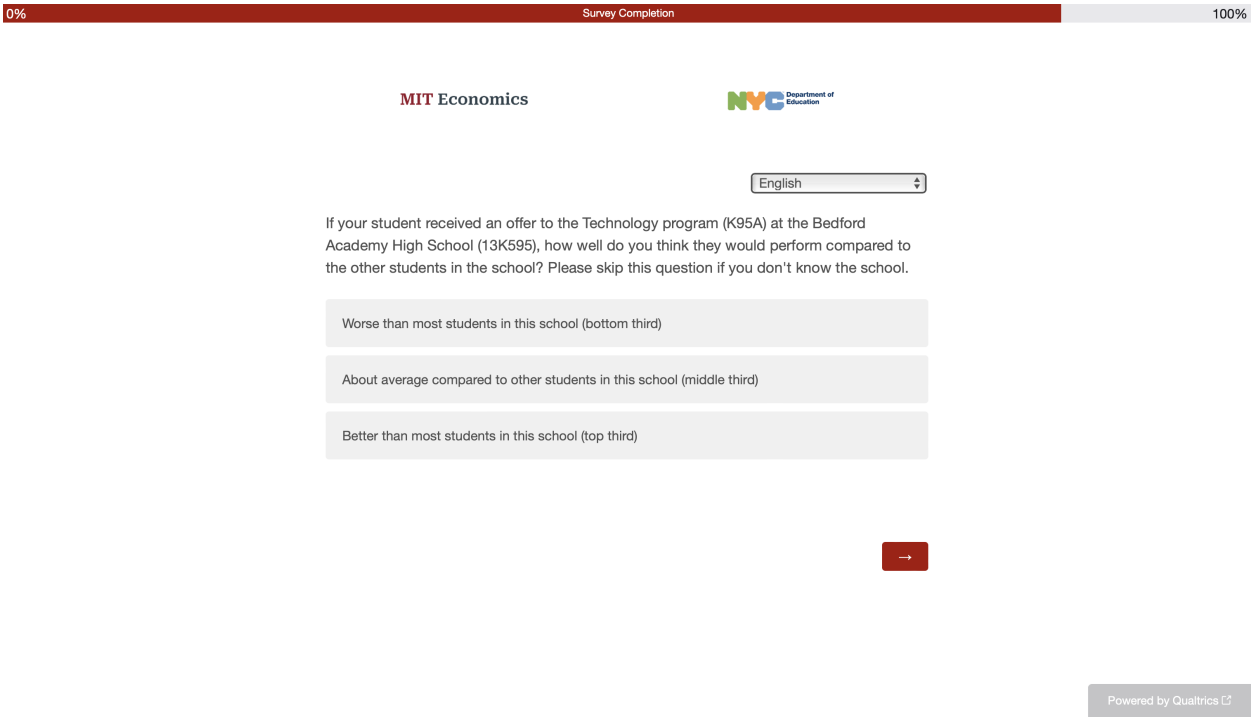


Figure C36: Question 13

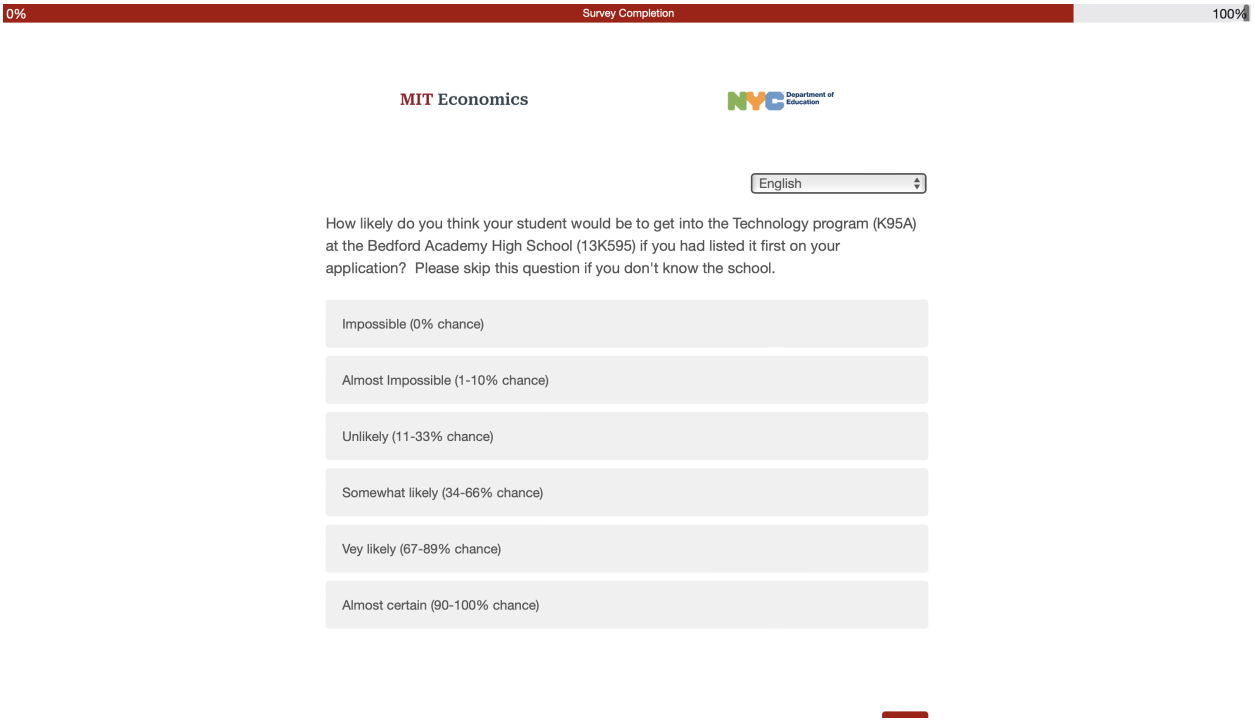


Figure C37: Question 14a

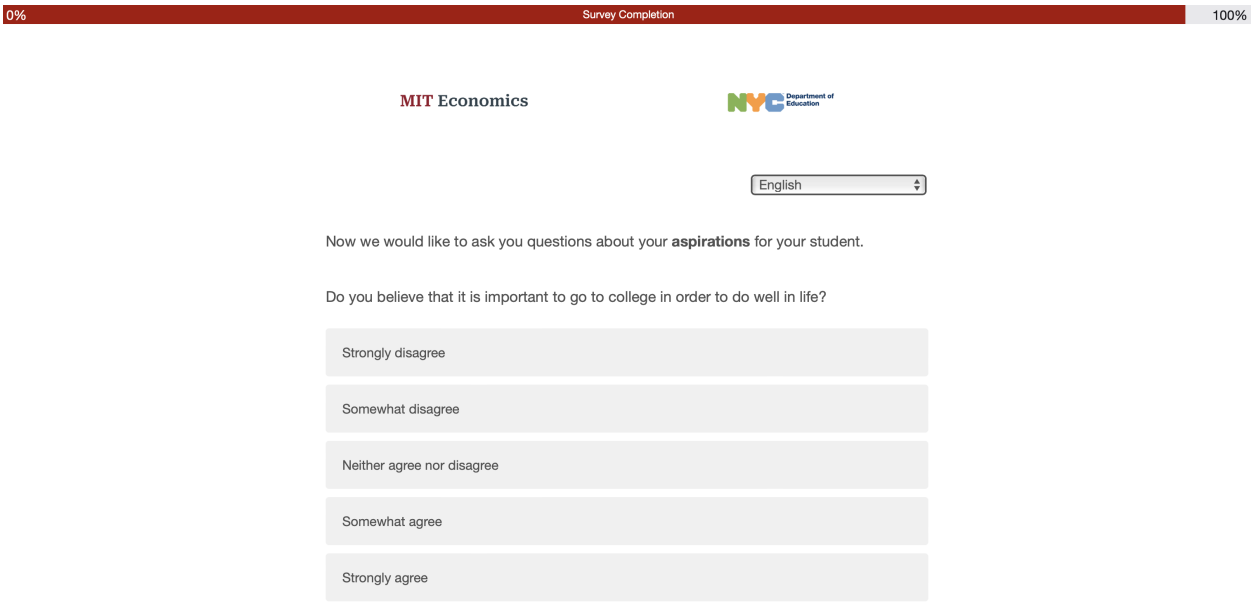


Figure C38: Question 15

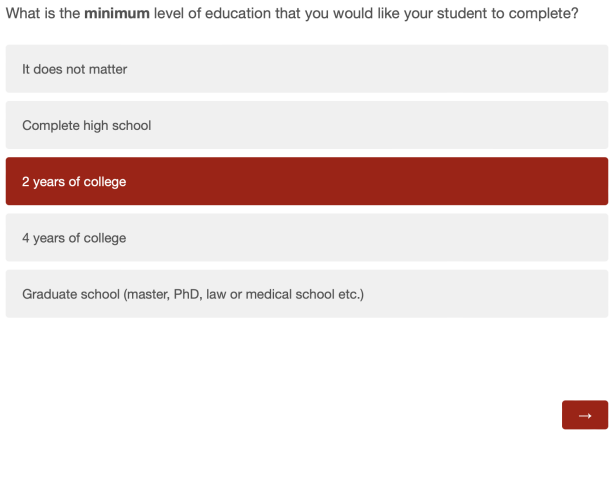


Figure C39: Question 16a

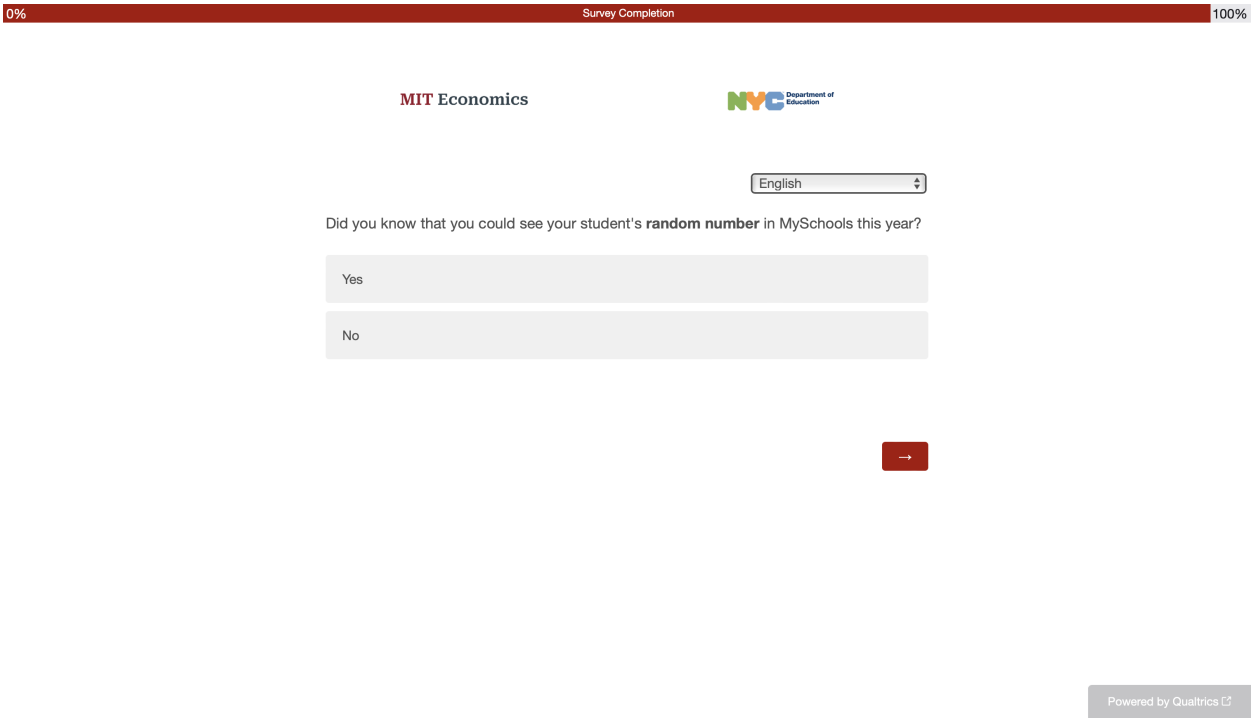


Figure C40: Question 16b

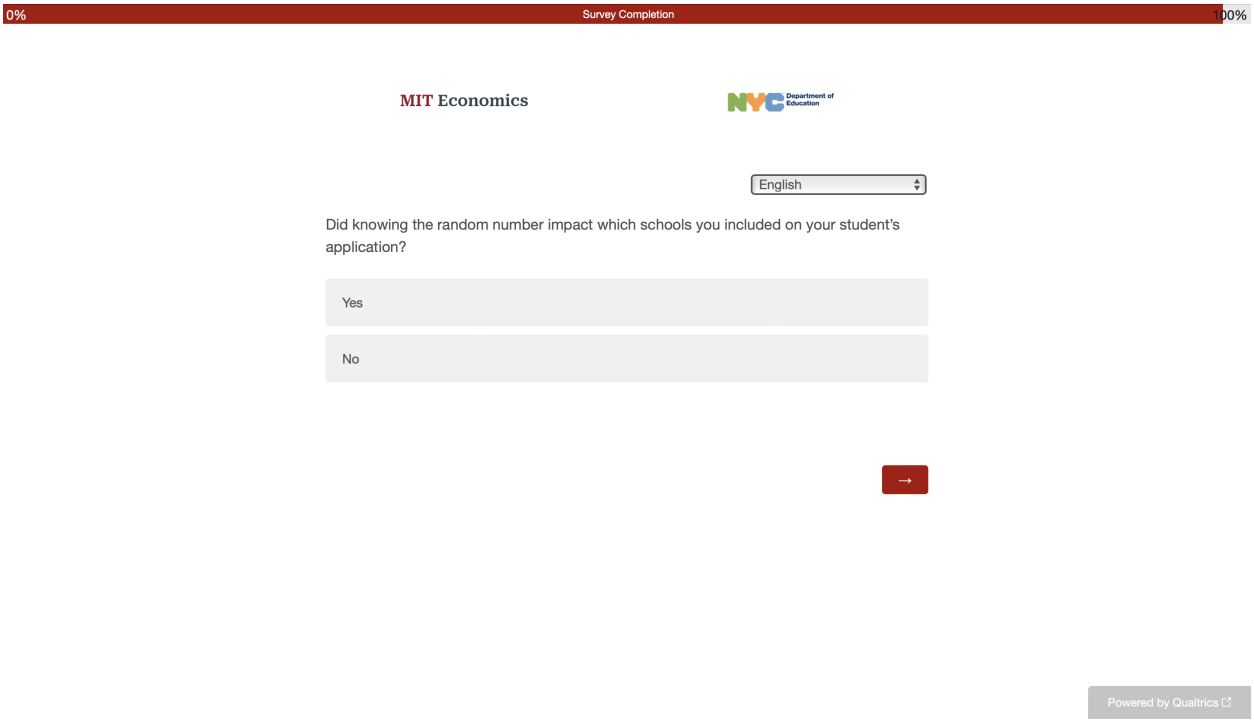


Figure C41: Question 16c

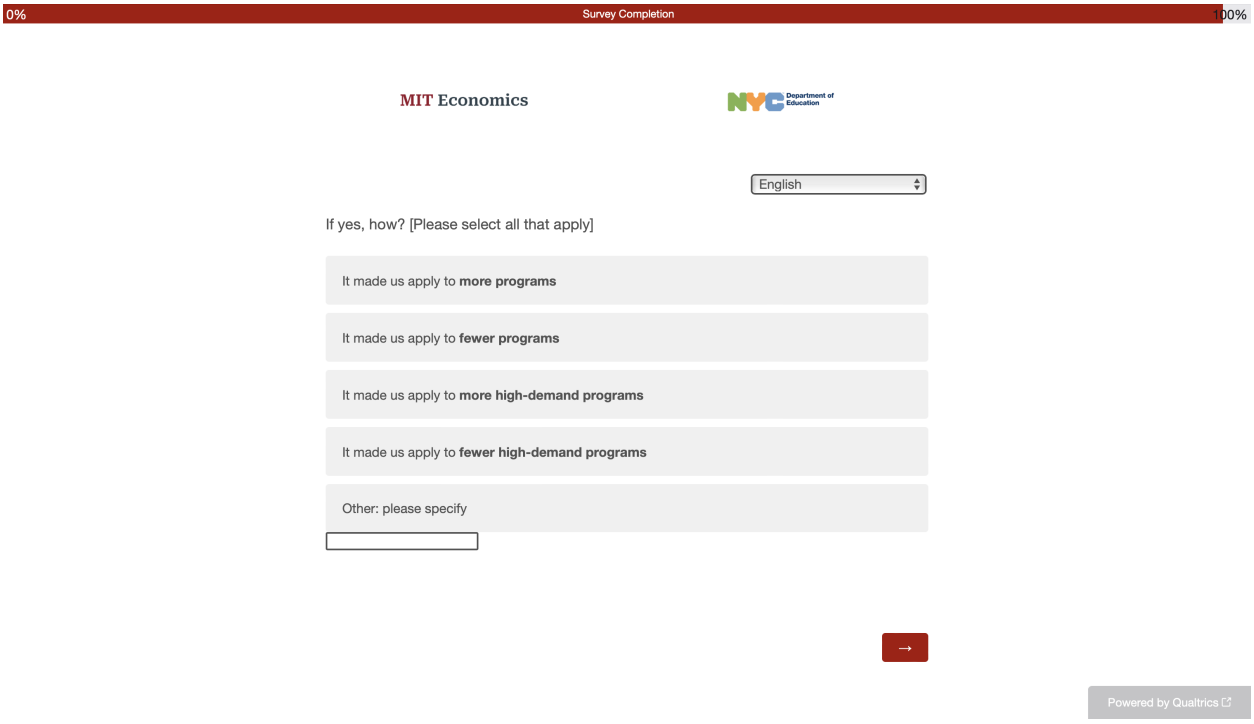


Figure C42: Question 17

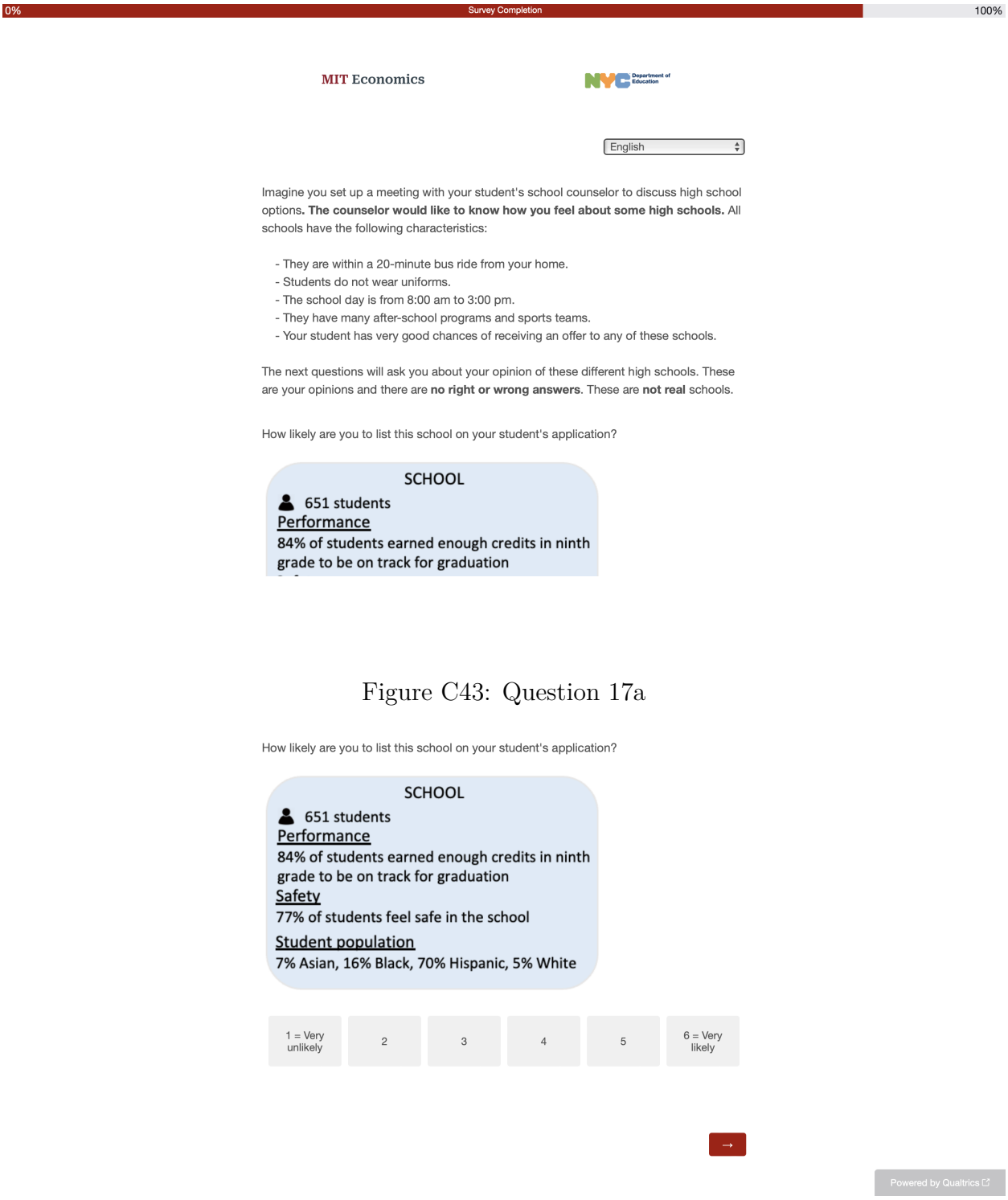


Figure C43: Question 17a

Figure C44: Question 17b

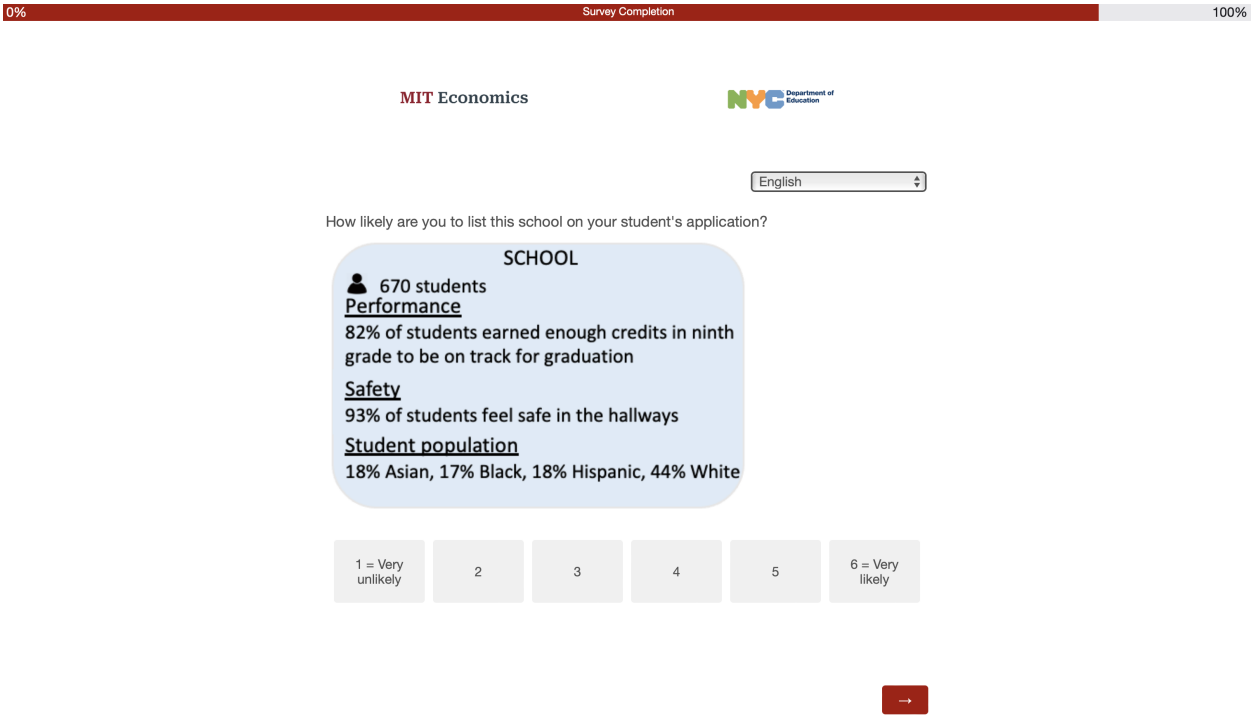


Figure C45: Question 18a

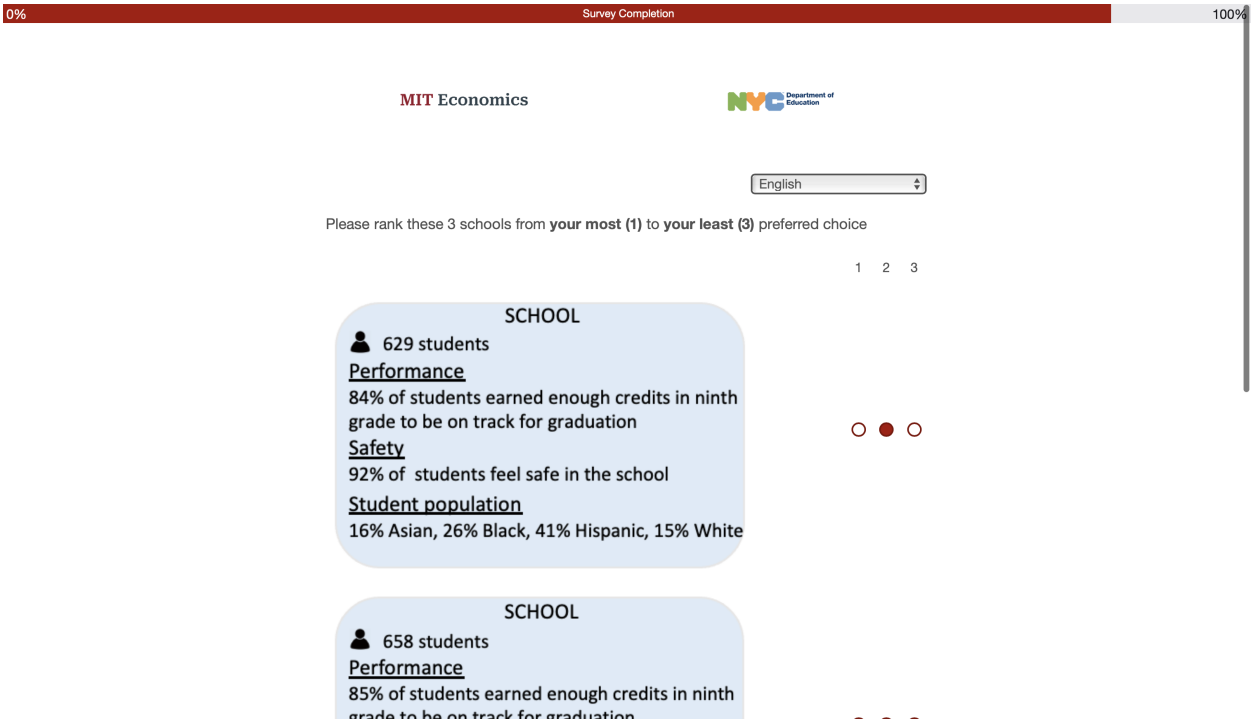


Figure C46: Question 18b

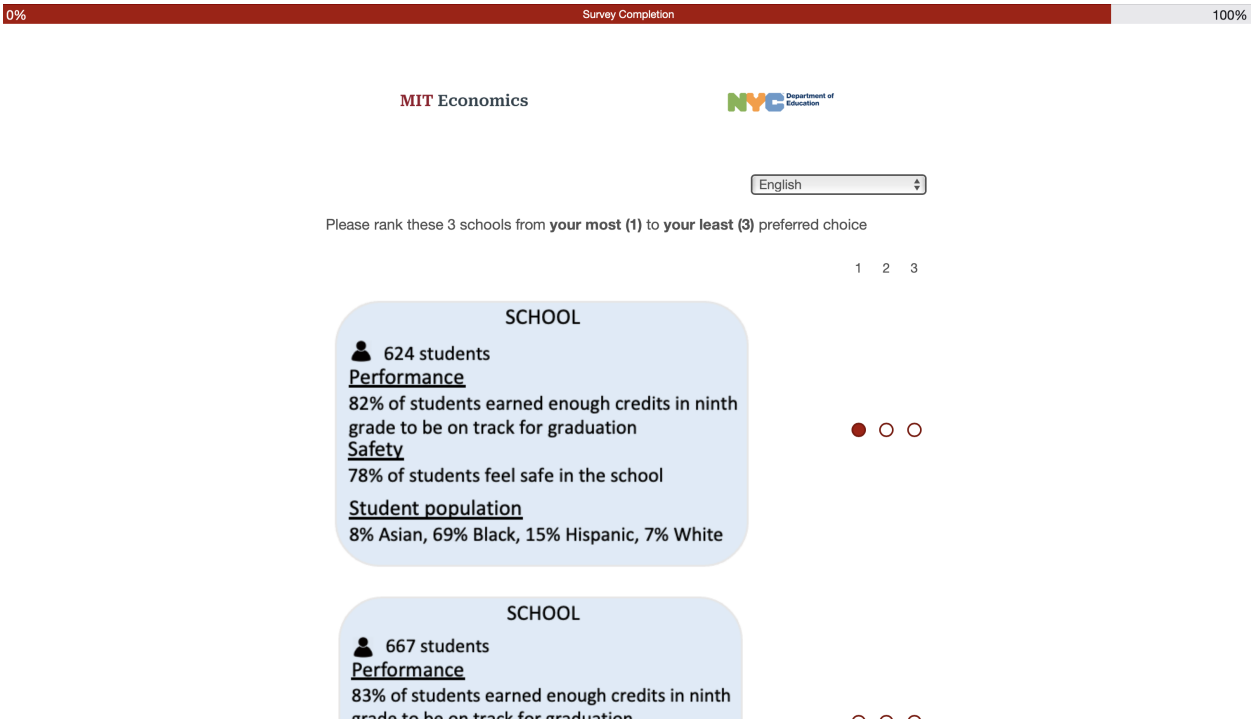


Figure C47: Question 19a

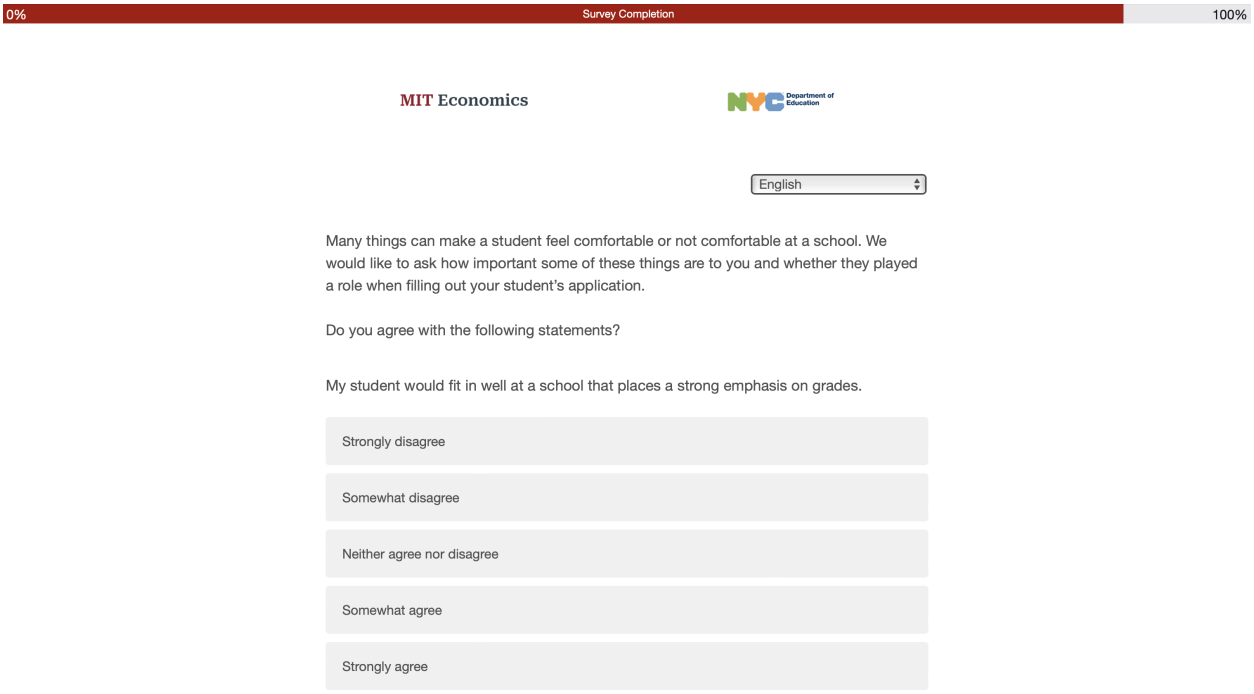


Figure C48: Question 19b

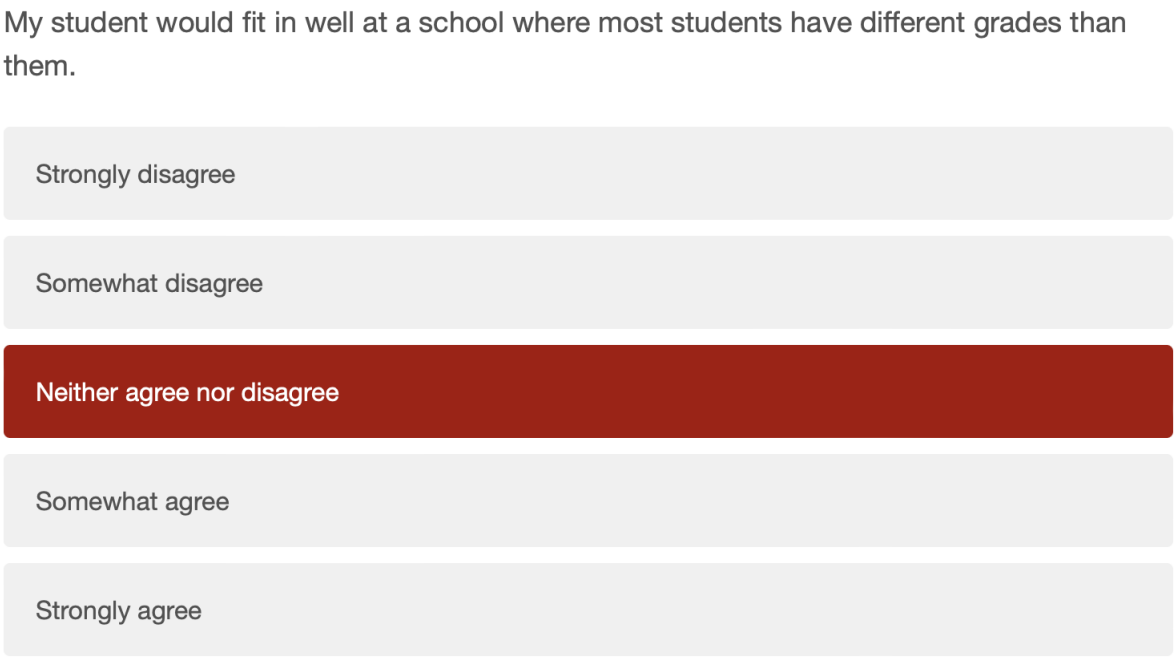


Figure C49: Question 19c

Figure C50: Question 20a

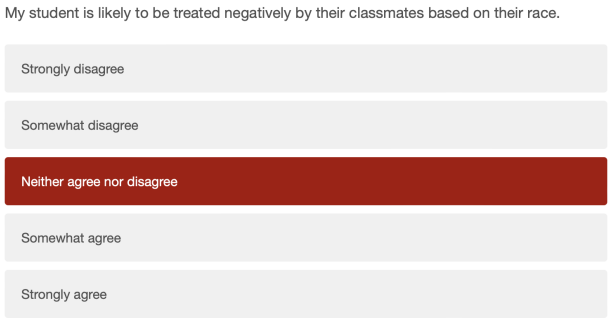


Figure C51: Question 20b

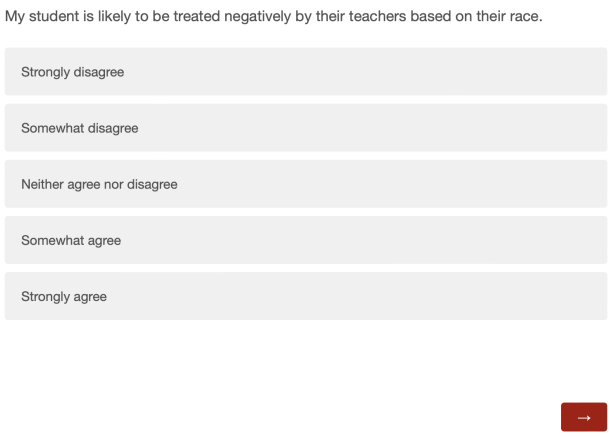


Figure C52: Question 20c

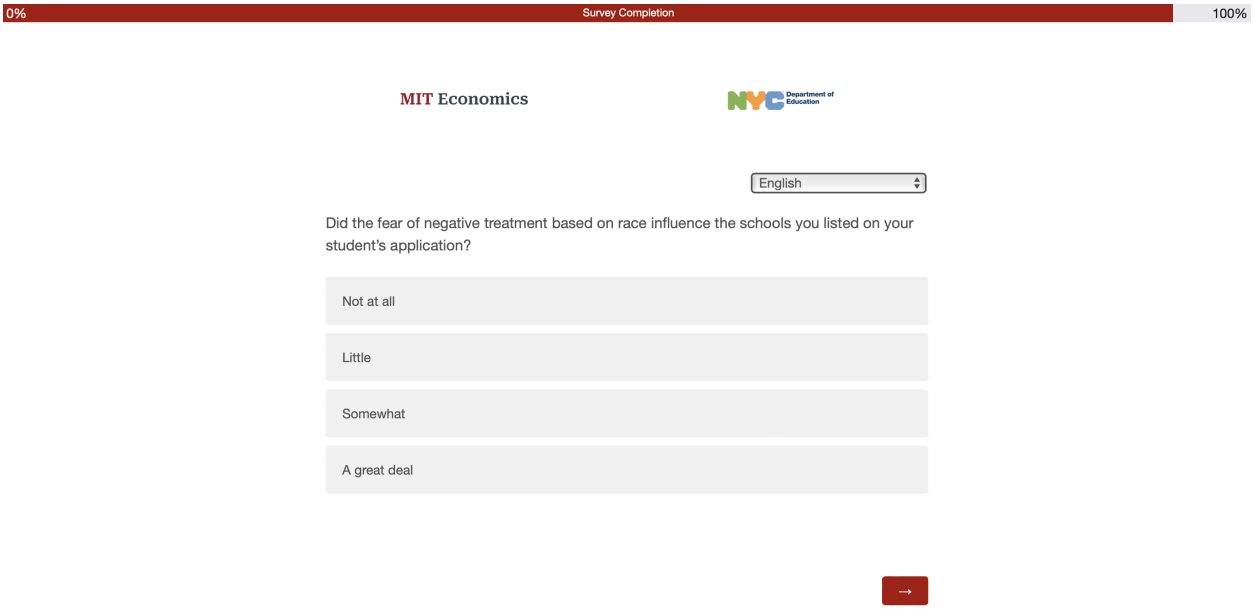
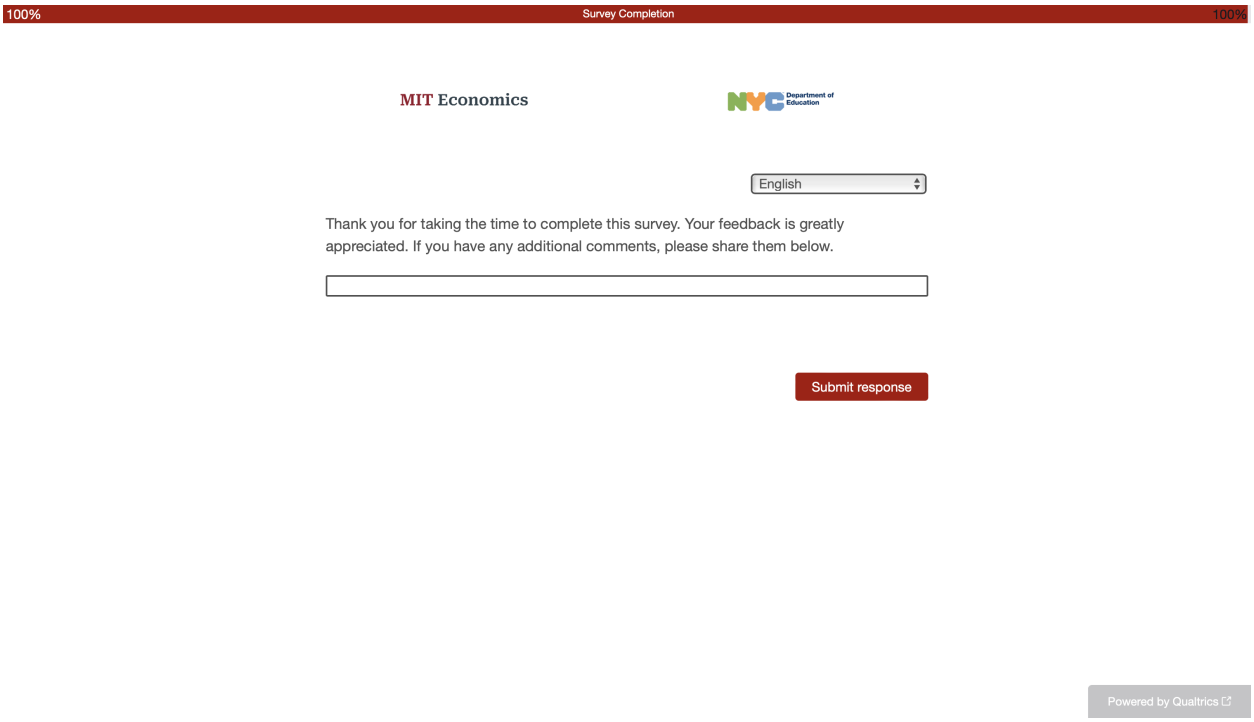


Figure C53: Question 21



Thank you message

0% Survey Completion 100%

MIT Economics

NYC Department of Education

We thank you for your time spent taking this survey.
Your response has been recorded.

Powered by Qualtrics

Information questions - Version 2

Figure C54: Question Q10 - Version 2

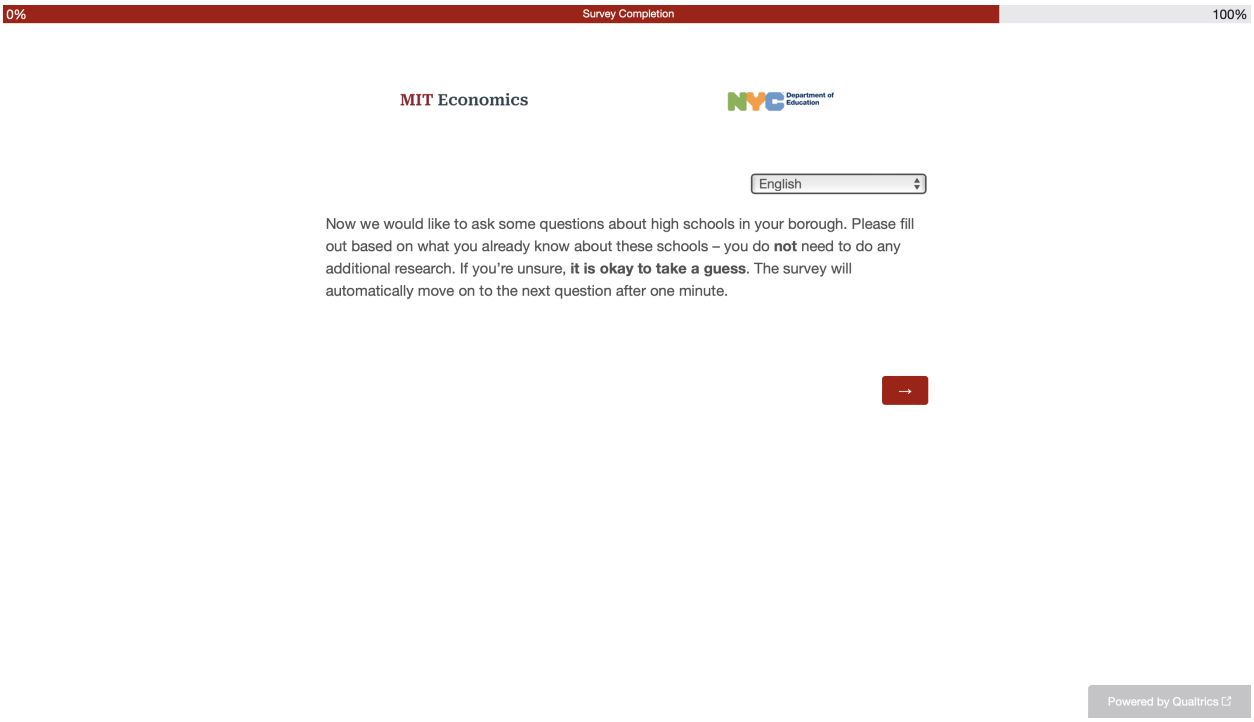


Figure C55: Question Q10a - Version 2

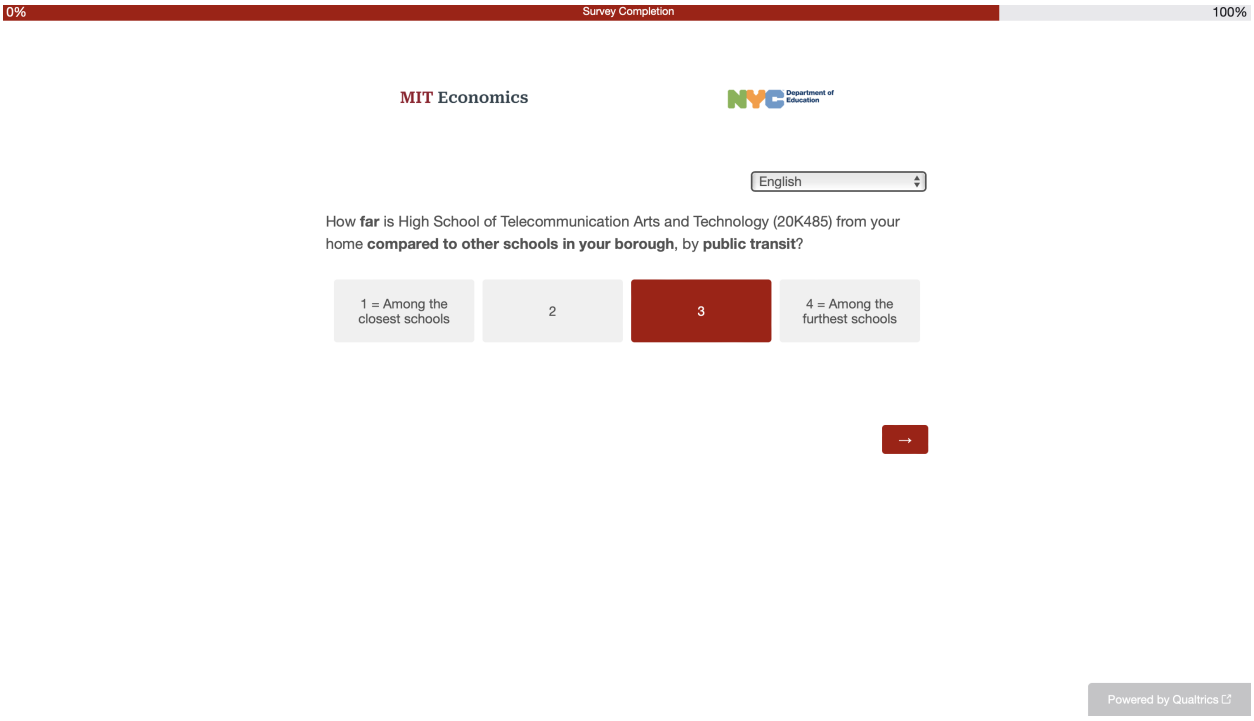


Figure C56: Question Q10b - Version 2

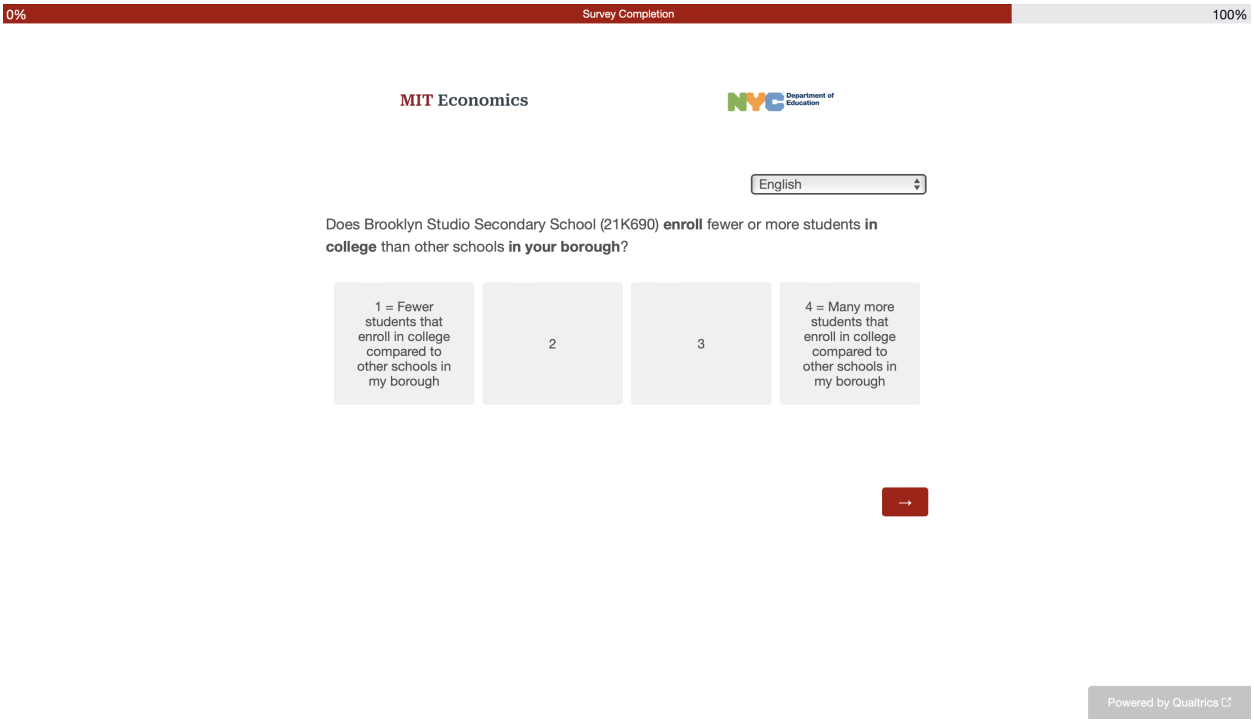


Figure C57: Question Q10c - Version 2

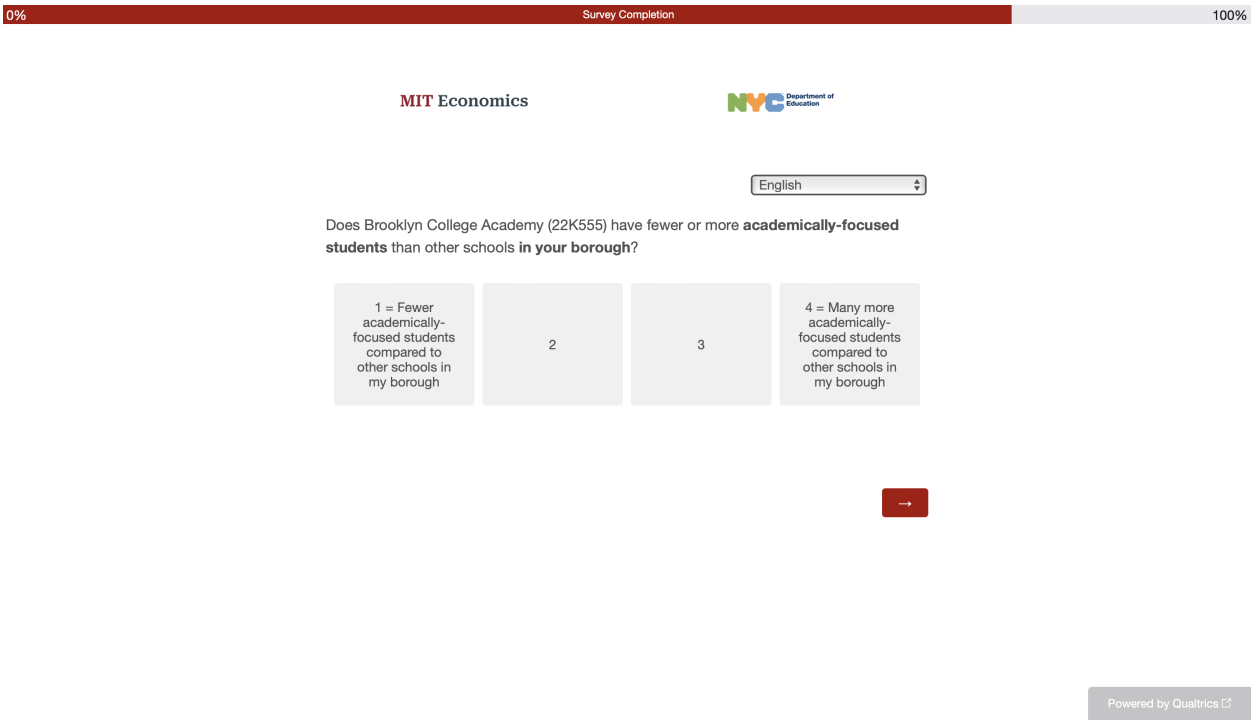


Figure C58: Question Q10d - Version 2

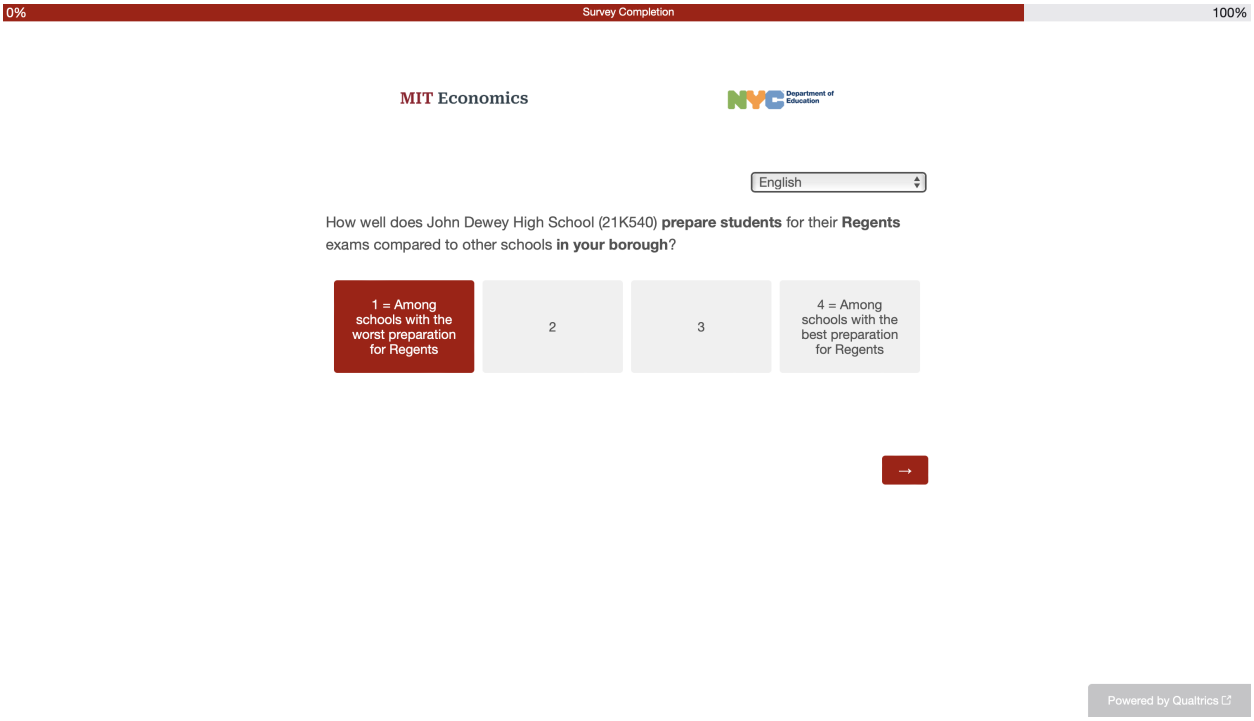


Figure C59: Question Q10f - Version 2

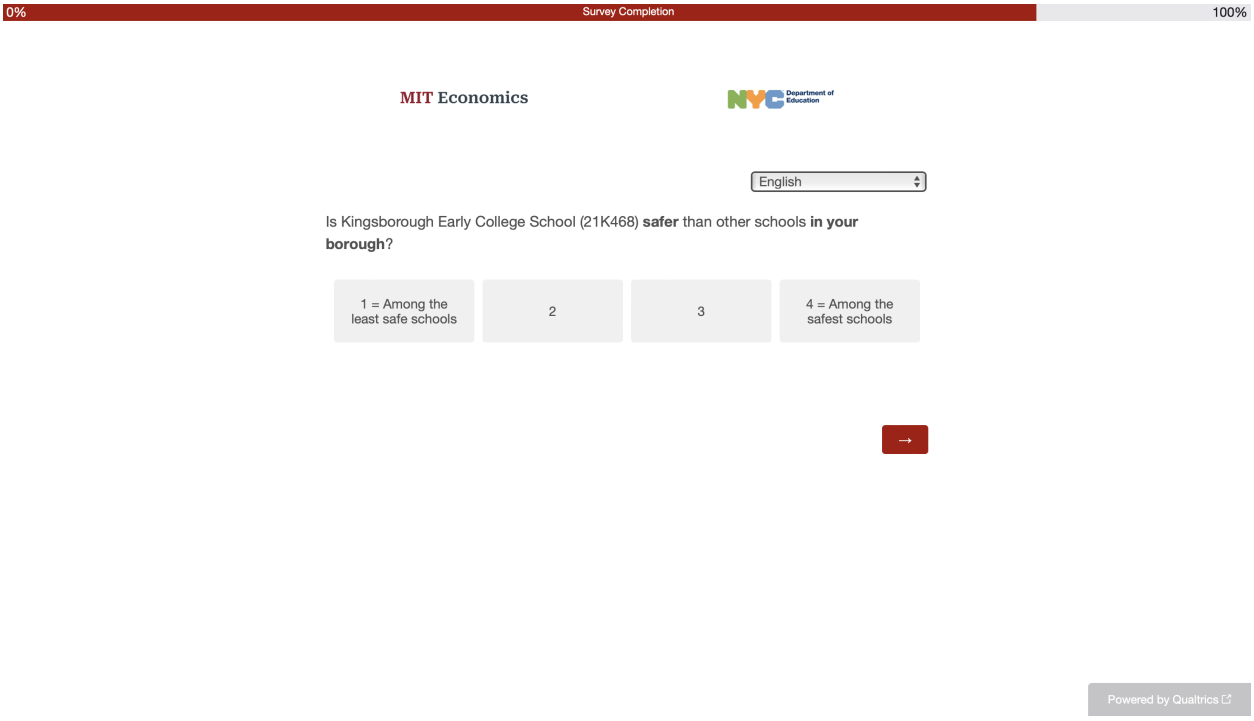


Figure C60: Question Q10g - Version 2