Distributional Impacts of Centralized School Choice*

Jaewon Lee[1] and Suk Joon Son[2]

[1]Yale University
[2]The University of Tokyo

October 9, 2023

Abstract

Informational frictions in centralized school choice can significantly influence its distributional consequences. Recognition of such frictions is also necessary to accurately measure welfare. We build a model of school applications, allowing applicants to consider only a limited set of schools and to have mistaken beliefs about their admission chances. Quasi-experimental variation and rich information in students' rank-ordered lists enable identification. Utilizing this model, we evaluate the impacts of centralized school choice in New York City on racial segregation and equity in welfare, decomposing the contributions of the frictions and the preferences of students and schools. We also quantify matching stability and deviations from truthful reporting. Our results show that while school choice improves welfare across races, limited consideration substantially compromises these gains, particularly for Black and Hispanic students. A counterfactual policy involving personalized school recommendations designed using our model is projected to recover 20–36% of the welfare losses.

Keywords: Centralized School Choice, Informational Frictions, Consideration Sets, School Segregation, Racial Inequality, Subjective Beliefs
JEL Codes: D47, D63, D83, H75, I21, I28, J15

*Son is deeply indebted to Steven Berry, Philip Haile, and Yusuke Narita for their constant guidance and support. We thank Jason Abaluck, Nikhil Agarwal, Victor Aguirregabiria, Joseph Altonji, Ian Ball, Marianne Bernatzky Koehli, Barbara Biasi, Ali Bray, Jin-wook Chang, Yeon-koo Che, Andrew Chesher, Susan Dynarski, Chao Fu, Kenneth Gillingham, Charles Hodgson, John Eric Humphries, Mitsuru Igami, Gerald Jaynes, Daiji Kawaguchi, Alvin Klevorick, Changhyun Kwak, Koohyun Kwon, Soonwoo Kwon, Mariana Laverde, Costas Meghir, Luis Carvalho Monteiro, Christopher Neilson, Cormac O'Dea, Jonathan Pierot-Hawkins, Jeff Qiu, Katja Seim, Nicholas Snashall-Woodhams, Jaehee Song, Jintaek Song, Eduardo Souza-Rodrigues, Allen Vong, Stephanie Weber, and Seth Zimmerman for helpful discussions, advice, and encouragement. We are very grateful of the support from the NYC Department of Education, particularly from Derek Li, Joshua Smith, and Stewart Burns Wade. This research is financially supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number 22K13412. It is also financially supported by the Departmental Research Grant of the University of Tokyo.

1 Introduction

School choice policies aim to alleviate the effects of residential locations on educational opportunities by enabling broader access to schools. A notable approach to school choice involves the use of centralized assignment mechanisms. These mechanisms are often motivated by theoretical results that guarantee stability, efficiency, or strategy-proofness.[1] However, their real-world impacts remain contested. The theoretical results depend on the assumptions that the applicants make well-informed and rational choices,[2] and that the implemented mechanisms adhere to the theoretical ideal.[3] Moreover, policymakers often prioritize distributional goals such as racial integration and equity,[4] which are aspects not directly addressed by the theoretical results. Furthermore, frictions in information or rationality, which may vary across demographic groups, can exacerbate the distributional outcomes. Therefore, the impact of centralized school choice on various outcomes is an empirical question, necessitating careful consideration of these frictions.

In this paper, we examine the effects of centralized high school choice in New York City (NYC), utilizing a school application model that emphasizes two key optimization frictions: applicants may consider only a subset of the 763 school programs and may have incorrect beliefs about their assignment chances. Quasi-experimental variation, such as the positioning of schools in the school directory, along with rich information embedded in students' rank-ordered lists of schools, enables the identification of the model using observational data. The ideas behind identification are formalized with results on nonparametric identification. Using the estimated model, we analyze the impact of school choice on integration and equity of welfare across different demographic groups. We further measure the contributions of several factors, including students' preferences, limited consideration sets, deviations from truthful reporting, as well as schools' admissions priorities and screening policies. After discovering that informational frictions substantially suppress student welfare, and especially so for Black and Hispanic students, we investigate the design of counterfactual interventions that make personalized school recommendations based on the estimated preferences and consideration probabilities. Additionally, we examine the stability of the matching and the truthfulness of the rank-orderings, the two main theoretical targets of NYC's assignment mechanism.

NYC's high school assignment procedure allocates students to schools through a version of the Deferred Acceptance (DA) mechanism, motivated by theoretical results that predict matching stability and strategy-proofness under idealized assumptions of informed and rational behavior. However, these theoretical results do not directly target important dis-

[1]See, e.g., Gale and Shapley (1962), Shapley and Scarf (1974), Ergin (2002), and Abdulkadiroğlu and Sönmez (2003). Such centralized mechanisms are used in, for example, New York City, Chicago, Boston, New Orleans, Paris, Spain, and Romania (Abdulkadiroğlu et al., 2020; Fack et al., 2019).

[2]See, e.g., Hassidim et al. (2017), Li (2017), and Fack et al. (2019).

[3]See, e.g., Abdulkadiroğlu et al. (2005), Haeringer and Klijn (2009), and Calsamiglia et al. (2010).

[4]See, e.g., NYC DOE (2020).

tributional outcomes such as racial diversity and equity. Furthermore, the presence of more than 700 school programs in NYC suggests potentially large informational frictions. NYC's implementation of DA also deviates from the canonical version; students can list at most twelve school programs, and there is an aftermarket. Consequently, there may be instances where a student's optimal strategy does not coincide with truthful reporting. In such situations, students must assess their admission chances, which requires a correct understanding of the nuances of NYC's specific DA algorithm. Disparities in information about NYC's school programs[5] or in grasping the intricacies of the algorithm[6] can have distributional consequences. Informational frictions and the divergence from truthful reporting can also compromise matching stability.

Our paper begins by presenting descriptive evidence suggesting frictions in optimization and the presence of racial disparities. Our findings indicate that applicants take admission chances into account even under situations where such behavior is weakly dominated. We further document that students are significantly less likely to apply to the schools appearing on the later pages of NYC's school directory—even though schools are alphabetically ordered—suggesting substantial informational frictions. Such patterns are more pronounced for Black and Hispanic students, whose neighborhood schools tend to be lower-performing and less selective.

To accommodate these observations, our model of students' application behavior incorporates elements of optimization friction. Such a model is particularly important in our context. A model without such frictions would force the researcher to interpret any observed behavior under school choice as optimal, potentially biasing the results in favor of school choice. Furthermore, a frictionless model attributes differences in the choice patterns across demographic groups to differences in preferences when, in fact, they may be caused by differences in frictions. In contrast, a model encompassing frictions enables us to disentangle the contributions of preferences and frictions and to provide guidance on possible policy interventions.

Specifically, our model allows each applicant to consider only a limited set of the school options[7] and have incorrect beliefs about equilibrium admission chances. An applicant may fail to consider a school because she is unaware of the school or feels she can never be admitted. Even if she does consider a school, she may have incorrect beliefs about how her rank-ordering of schools can affect her assignment probabilities.

Rich information in students' rank-ordered lists, combined with exogenous variation in consideration, enables the identification of the model using observational data. As an ex-

[5] See, e.g., Sattin-Bajaj (2016) and Corcoran et al. (2018).

[6] See, e.g., Pathak and Sönmez (2008), Sattin-Bajaj (2016), Basteck and Mantovani (2018), and Rees-Jones (2018).

[7] To be precise, some schools host multiple *programs*, and these programs are the primary units of analysis for most of our results. We will distinguish between schools and their programs when such distinction becomes necessary.

ample, while a lack of consideration may affect *which* schools are listed, it cannot affect *where* a listed school will be ranked. Concerning the exogenous variation, we argue that certain observables, such as the positioning of schools in the NYC directory, can affect the consideration set but not preferences.[8] Another assumption that assists identification is that some students have a set of schools (e.g., noncompetitive schools close to home) that they will surely consider. We formalize our intuitive identification strategy by establishing sufficient conditions for nonparametric identification using the type of rank-ordered choice data typically available from centralized school choice systems, with an appropriate instrument. These conditions clarify the sources of identification and the limited role played by functional form assumptions.

The estimates reveal racial differences in consideration and preference patterns. Black and Hispanic students are more likely to consider less selective schools and prefer them to their outside options. Asian and White students' consideration sets are better aligned with their preferences. Across all races, students' reporting strategies are estimated to be approximately consistent with truthful reporting among the considered programs.

Using our estimated model, we quantify racial integration and equity in school assignments. The results indicate that school choice slightly promotes racial integration, mainly by reducing the isolation index of Black students by approximately 7.7 percentage points. Furthermore, school choice also significantly improves welfare across all racial groups; the proportion of students matched to one of their top five preferred school programs increases from about 3% under neighborhood matching to around 28% under school choice matching. The improvement is larger for Black and Hispanic students. However, limited consideration substantially suppresses the welfare gains. If students considered all schools, students would be about twice as likely to be matched to one of their top five preferred school programs, with the greatest potential gains for Black and Hispanic students. Schools' admissions priorities and screening policies segregate races and tend to place Asian and White students in their preferred schools.

Recognizing the significant welfare losses resulting from limited consideration, we propose using our model to design targeted information interventions.[9] These interventions utilize the estimated preferences and consideration sets to recommend 30 programs to each student. Some of these interventions show significant promise, with the most effective one estimated to address between 20–36% of the welfare losses.

We also measure matching stability by quantifying the prevalence of justified envy, a situation where a student and a school program prefer each other over their current matches.

[8]Relatedly, Martin and Yurukoglu (2017) use local channel positions as exogenous variation that shifts channel viewership but are uncorrelated with the local political inclinations. A number of previous research considered the effects of positioning of items in online settings; see, e.g., Feng et al. (2007), Koulayev (2014), Ursu (2018).

[9]Allende et al. (2019) also use the estimated model to study alternative designs of their information intervention.

On average, our estimates indicate that students view only around three school programs with justified envy, a relatively small number compared to around 750 available programs.

Relation to the Literature Our paper provides the following contributions to the literature. First, we estimate a model of school applications that allows the applicants to have limited awareness about school options in addition to potentially incorrect beliefs about admission chances. We also provide an identification strategy based on observational choice data. To the best of our knowledge, aside from our own, only the work by Ajayi and Sidibe (2022) estimates a school applications model that allows for limited consideration due to a lack of awareness. They also allow for incorrect beliefs about admission chances. In their main model, applicants engage in a sequential search process to expand their consideration sets. While we do not model the sequential aspect of search, we allow the consideration probabilities to be correlated with utilities through a rich set of observables.[10] We also complement their work by providing nonparametric identification results. These results relate to those of Agarwal and Somaini (2022), who examine the identification of preferences and latent choice sets. They consider the case of single-unit demand with the presence of two types of instruments, one that affects preferences but not the choice sets and the other that affects choice sets but not preferences. In contrast, in our empirical setting, while only the latter kind of instruments are present,[11] students can list and rank-order multiple schools, which provide additional identifying variation. Our model also distinguishes between consideration and nondegenerate beliefs given consideration. Our identification strategy also builds upon Agarwal and Somaini (2018), who provide sufficient conditions for nonparametric identification of preferences while assuming full consideration and holding fixed a mode of beliefs in a centralized school choice setting.[12] Allende et al. (2019) estimate a school choice model where students have imperfect information about the school attributes. Our paper also relates to the broader literature on the estimation and identification of discrete choice models with limited consideration.[13]

Other studies have documented the importance of limited information about the school options. Using surveys and informational interventions, Arteaga et al. (2022) show that the search frictions are significant in their school choice setting and that search behavior is affected by their (updated) beliefs about admission chances. Corcoran et al. (2018) provide evidence that information intervention affects application behavior in the NYC high school

[10] In Ajayi and Sidibe (2022)'s model, conditional on deciding to continue searching, the probability of discovering a school is determined by schools' number of seats and, in their directed search model, also by the perceived admission chances.

[11] As a supplementary nonparametric identification result, we discuss the case where both types of instruments are present (Proposition D.1) unlike the empirical setting.

[12] The approaches used in nonparametric identification results are further related to, for example, Thompson (1989), Bresnahan and Reiss (1991), Lewbel (2000), Berry et al. (2013), and Berry and Haile (2020).

[13] See, e.g., Goeree (2008), Conlon and Mortimer (2013), Gaynor et al. (2016), Hortaçsu et al. (2017), Abaluck and Adams-Prassl (2021), Barseghyan et al. (2021a), and Barseghyan et al. (2021b).

application procedure. Narita (2016) shows that students in NYC modify their orderings of schools during the re-application process; many applicants self-report that these changes arise from evolving preferences or updated information. Informational frictions are also important in other environments of school or college applications (e.g., Hastings and Weinstein, 2008; Hoxby and Turner, 2013; Ajayi et al., 2017; Dynarski et al., 2021). We find that limited consideration significantly suppresses the welfare of students and discuss how we may use the estimated model to construct effective school recommendation policies.

Our second contribution is to disentangle the role that limited consideration plays in racial segregation and inequality in school assignments from the role played by students' preferences. Relatedly, Ajayi and Sidibe (2022) estimate the welfare loss due to information frictions in a centralized school choice system in Ghana and that the loss is concentrated on low-ability students. Other studies have empirically examined the contributions of various factors to equity or segregation under centralized school choice procedures (Kessel and Olme, 2018; Laverde, 2020; Oosterbeek et al., 2021; Akbarpour et al., 2022; Hahm and Park, 2022; Sartain and Barrow, 2022; Idoux, 2023; Park and Hahm, 2023). Calsamiglia et al. (2021) theoretically examine the impact of matching algorithms on segregation. There have been studies that examine the distributional impacts of school choice in other contexts (e.g., Epple and Romano, 1998; Hsieh and Urquiola, 2006; Bifulco and Ladd, 2007; Neilson, 2013; Altonji et al., 2015; Hom, 2018; Avery and Pathak, 2021).

We further contribute to a growing literature that allows for subjective beliefs about admission chances in school choice settings. We additionally allow for imperfect awareness of school options. Kapor et al. (2020) estimate a model that allows for subjective beliefs using survey data on perceived admission chances and data on rank-ordered lists. Our model of beliefs is based on theirs, and we complement their work by providing results on identification that use data on observed choices and instruments rather than survey data. Relatedly, Luflade (2018) and Calsamiglia et al. (2020) estimate preferences and potentially incorrect beliefs with observed choice data without surveys.[14] Some studies propose strategies for estimating preferences while allowing for mistaken beliefs under nontruthful mechanism (He, 2017; Hwang, 2017; Agarwal and Somaini, 2018) and while allowing for nontruthful behavior under (approximately) truthful mechanisms (Artemov et al., 2017; Fack et al., 2019; Che et al., 2020; Larroucau and Rios, 2020; Idoux, 2023).[15] Our findings indicate that, while students may drop schools from their submitted reports because of admission chances that are perceived to be negligible (even when the list length constraint is not binding), they rarely place a lower-utility school above a higher-utility school. These findings are consistent with the literature that finds or assumes that nontruthful ordering is less common than

[14]More broadly, Aguirregabiria (2021) studies the identification of firms' preferences and beliefs about the competitors' behavior using data on observed actions.

[15]Abdulkadiroğlu et al. (2017) and Che and Tercieux (2019) assume weak versions of the truthtelling assumption.

dropping an unlikely school (Fack et al., 2019; Fabre et al., 2021; Shorrer and Sóvágó, 2022).

We also measure matching stability and the influences of various factors on student welfare. Luflade (2018) analyzes the value of information about admission chances on welfare. This paper measures the effect of limited consideration sets and the deviations from truthful reporting on welfare.[16] Other studies have investigated student welfare or matching stability (Narita, 2016; Abdulkadiroğlu et al., 2017; He, 2017; Hwang, 2017; Agarwal and Somaini, 2018; Luflade, 2018; Che and Tercieux, 2019; Abdulkadiroğlu et al., 2020; Kapor et al., 2020; Calsamiglia et al., 2020). Our paper ensures that frictions in awareness and in the assessments of admission chances are not conflated with utilities. Thus, our evaluation of welfare and stability reflects preferences net of the influences from the frictions.

2 Overview of New York City's High School Choice

2.1 The Context

The NYC public high school choice system matches approximately 80,000 students to more than 700 public high school programs annually. The system uses the following centralized procedure:[17]

(1) Each applicant submits her rankings over the school programs. She can rank up to 12 school programs.

(2) Each school program ranks applicants using the admissions priority groups, screening policies, and/or lotteries.

(3) NYC runs the student-proposing DA algorithm to assign students to school programs using the rankings of the students and the school programs.

The matching procedure in NYC creates incentives for the applicants to deviate from truthfully reporting their preferences, due to the list length constraint and the presence of the aftermarket. This is discussed in Section 2.2.

Characteristics of the student sample are summarized in Table 1.[18] The district has many minority students and low-income students. Of the students in the sample, 40.5% of the students are Hispanic, 26.9% are Black, 16.1% are Asian, and 15.0% are White.[19] 71.3% of the students are eligible for free or reduced-price lunch. The table also (partially) demonstrates the housing racial segregation in NYC.

[16] As discussed above, Ajayi and Sidibe (2022) measures the welfare loss due to limited search.

[17] We focus on the applications towards traditional public high schools, excluding specialized high schools or charter schools. Approximately 70% of NYC high school students attend traditional public high schools. See Appendix C for details.

[18] For discussions of the data and the sample, refer to Section 3.1.

[19] We use race and ethnicity interchangeably in this paper.

Table 1: Characteristics of Students by Ethnicity

	Asian	Black	Hispanic	White	Total[a]
Proportion in the sample	16.1%	26.9%	40.5%	15.0%	98.4%
Female	47.8%	48.8%	48.4%	48.2%	48.4%
English Language Learner	13.5%	2.7%	18.2%	6.4%	11.3%
Subsidized lunch	69.5%	76.4%	80.5%	40.2%	71.3%
Students with disabilities	7.4%	25.0%	24.8%	17.3%	20.8%
Neighborhood income[b]($)	58553.0	49469.1	47624.1	73686.9	54119.7
Mean distance to schools	9.12	8.99	8.47	10.92	9.09
Home boroughs					
Bronx	6.4%	25.9%	36.2%	6.0%	23.7%
Brooklyn	29.2%	42.5%	20.3%	33.5%	29.8%
Manhattan	7.5%	8.8%	12.7%	12.8%	10.9%
Queens	52.9%	19.4%	26.5%	25.5%	29.0%
Staten Island	4.0%	3.4%	4.2%	22.1%	6.7%
State Reading Category[c]					
High	42.7%	16.7%	16.3%	43.7%	25.1%
Middle	50.6%	68.4%	67.0%	50.4%	62.0%
Low	6.7%	14.9%	16.7%	5.8%	12.8%
Report length	7.2	7.5	7.2	5.6	7.1

Notes: Except for the *proportion in the sample*, all the percentage terms represent the proportions of the relevant categories within each ethnicity.

[a] 1.6% of students are multi-racial or Native American.

[b] Based on the ZIP code of the student's home address. Median household income is from U.S. Census Bureau, 2013–2017 American Community Survey five-year estimates, in 2017 dollars.

[c] Based mostly on NY State English Language Arts test submitted from a student's 7th grade school year. See Appendix E.5 for details.

The school and program characteristics are summarized in Table 2 by borough. Schools vary widely in their characteristics, both within and across boroughs. For example, across boroughs, while the average proportion of Hispanic students is 65% in the Bronx schools, it is only 28% in Staten Island. There is also wide within-borough variability. For instance, the standard deviation of the proportion of Hispanic students is as large as 22 percentage points within Brooklyn. While there are only nine schools in Staten Island, there are roughly 100 schools in each of the other four boroughs.

A school may have multiple programs within it, and each program has its own admission policy and interest area. How the programs rank their applicants may be based on *admissions priority groups*, *screening policies*, and *lotteries*. Priorities groups are lexicographically more important than the rankings based on screening or lotteries.[20] To break ties within a group, 37.2% of the programs use screening. It can be based on various criteria, including grades,

[20]The high school directory writes that "All students in the first priority group will be considered first. If seats are available, students in the second priority group will be considered next, and so on". However, we observe that 4.34% of students experience deviations from this stated lexicographic rule.

Table 2: Characteristics of Schools and Programs by Borough

	Bronx	Brooklyn	Manhattan	Queens	Staten Island	Total
Schools						
Graduation rate	0.68 (0.15)	0.74 (0.14)	0.79 (0.16)	0.79 (0.16)	0.78 (0.10)	0.75 (0.15)
College/career rate	0.49 (0.15)	0.51 (0.17)	0.61 (0.19)	0.64 (0.19)	0.65 (0.15)	0.56 (0.18)
Average grade 8 math (std.)	-0.50 (0.58)	-0.19 (0.81)	0.35 (1.17)	0.50 (1.11)	0.55 (0.64)	0.00 (1.00)
Proportion White	0.03 (0.03)	0.07 (0.12)	0.10 (0.15)	0.11 (0.11)	0.43 (0.21)	0.08 (0.13)
Proportion Black	0.27 (0.12)	0.55 (0.28)	0.26 (0.15)	0.28 (0.26)	0.17 (0.12)	0.35 (0.24)
Proportion Asian	0.03 (0.03)	0.07 (0.10)	0.09 (0.12)	0.22 (0.15)	0.08 (0.03)	0.09 (0.12)
Proportion Hispanic	0.65 (0.13)	0.29 (0.22)	0.52 (0.21)	0.35 (0.21)	0.28 (0.12)	0.45 (0.24)
9th grade school seats	115.51 (71.19)	157.98 (145.20)	133.98 (84.56)	187.08 (141.38)	304.33 (217.18)	149.25 (121.12)
Number of schools	116	122	105	79	9	431
Programs						
9th grade program seats	88.20 (36.60)	84.60 (71.29)	98.38 (58.91)	96.59 (52.50)	62.25 (27.55)	89.34 (57.18)
Number of programs: All	155	240	146	172	50	763
By admission methods						
Uses admissions priority groups	123	154	87	94	42	500
Uses screening	68	139	100	115	37	459
Uses lottery only	2	9	5	3	0	19
By interest area						
Arts	25	47	26	20	7	125
STEM	35	59	27	37	10	168

Notes: The standard deviations in each respective borough or in NYC are given in parentheses. Standardized values are indicated by (std.). *College/career rate* indicates the proportion of students who graduated from high school four years after entering 9th grade and then enrolled in college, a vocational program, or a public service program within six months of graduation. All schools and programs have equal weight regardless of their number of seats. The numbers under *By admission methods* and *By interest area* denote the number of programs. The sample excludes the nine specialized high schools. See Appendix E.4 for our definition of *interest area*. *Uses lottery only* are the programs that use admission lotteries and neither screening nor admission priority groups.

standardized test scores, attendance, punctuality, interview, and auditions. Other programs use lotteries to break ties within a priority group.[21]

2.2 Deferred Acceptance Algorithm: Theory and Practice

The DA algorithm has been gaining popularity,[22] based partly on theoretical results that promise certain desirable properties. One such property is that the mechanism is strategy-proof for the applicants: truthfully reporting their preference rankings weakly dominates any other strategy. Another such property is matching stability. An important feature of matching stability is that the matching does not have any unmatched student-program pair such that each side prefers the other to (one of) the current assignment(s), i.e., the matching does not have any case of justified envy.[23] However, these properties do not directly address

[21]Educational Option programs use both screening priorities and lotteries.

[22]DA is used in, for example, Boston, Chicago, Finland, Ghana, and Taiwan (Fack et al., 2019).

[23]Following the standard definition (e.g., Roth and Sotomayor (1992)), a matching is *stable* if there does not exist: (1) any case of a *blocking pair*, i.e., an unmatched student-school pair where each side prefers the other to [one of] the current assignment[s] (which might be an empty seat or no school assignment), and (2) any case of *individual irrationality*, where a student [school] would prefer to remain unmatched [have one additional empty seat] than to be matched to [one of] the current assignment[s]. It follows that a student has justified envy if he is part of some blocking pair (Abdulkadiroğlu and Sönmez, 2003).

distributional outcomes such as racial integration or the equity of assignments.

Even the two desirable outcomes promised by the theoretical results, namely, stability and truthful reporting, may fail in practice. Survey- and experiment-based evidence shows that a fraction of applicants do not truthfully report even in DA mechanisms (Chen and Sönmez, 2006; Calsamiglia et al., 2010; Hassidim et al., 2017; Rees-Jones, 2018; Hassidim et al., 2021). Complementing these results, Ashlagi and Gonczarowski (2018) theoretically show that, in generic cases, DA is not obviously strategy-proof in the sense of Li (2017); applicants with limited rationality may not understand its strategy-proofness. Stability may also not hold in DA; failure of truthful reporting may undermine stability (Gale and Shapley, 1962; Artemov et al., 2017; Fack et al., 2019) or when students consider and choose from only a limited set of schools. Furthermore, theoretically ideal versions of DA that guarantee strategy-proofness and stability are only occasionally implemented in practice (Abdulkadiroğlu et al., 2009; Haeringer and Klijn, 2009).

The matching procedure in NYC creates incentives for the applicants to deviate from truthfully reporting their preferences. This is because NYC's implementation of DA deviates from its canonical implementation in two respects. First, while the canonical implementation allows applicants to list arbitrarily many school programs, in NYC, applicants can list only up to 12 school programs. Students who wish to apply to more than 12 school programs must then decide which of these programs will be listed, which optimally depends not only on their preferences but also on their admission chances to the schools. Reflecting this, the 2017 NYC High School Directory states that "[i]f you are applying to 'reach' programs, be sure to include 'target' or 'likely-match' programs on your application." Second, while the canonical implementation conceives a single round of applications, in NYC, there is an aftermarket that follows the main round.[24] If a student believes that she can be matched to a school in this aftermarket, she may choose not to apply to this school in the main round.

In addition, given that there are more than 700 school programs in NYC, it is likely that students are not aware of many of them. Corcoran et al. (2018) has found that providing information about the nearest 30 schools with above-median graduation rates altered the students' choices in NYC. Additionally, lower-income families may have differentially less information about high-performing schools due to various reasons (Sattin-Bajaj, 2016).

In the next section, we document descriptive evidence of informational frictions and the influences of admission chances in school applications.

3 Evidence of Frictions, Disparities, and Choice Behavior

Section 3.1 introduces the data used. Evidence in Section 3.2 suggests a substantial lack of awareness of the schools. It also indicates that students take admission chances into account

[24] Until 2019, there was a second round of DA for the schools with remaining seats (see, e.g., Narita, 2016). In 2020, the waitlist system replaced the second-round DA.

when applying to schools. Section 3.3 documents the patterns of racial disparities and usage of school choice.

3.1 Data

Our primary dataset is the administrative records provided by the NYC Department of Education (DOE) for the 2016–2017 academic year. The data include students' rank-ordered choices, final school assignments, admissions priorities, schools' rankings over students, and demographic information. The demographic information includes students' race, English Language Learner status, home address, subsidized lunch status, disability status, and performance on statewide seventh-grade English and math tests. We restrict our sample to eighth graders attending an NYC DOE public school at the time of application mainly due to missing demographic data for other students.[25] We also use publicly available school-level data, including those from NYC's High School Directory and School Quality Reports.

3.2 Evidence of Frictions

Our descriptive analysis indicates that students face substantial frictions in learning about the school options, and that they drop the schools with lower admission chances out of the list even when the list constraint is not binding. We first examine the students' awareness of schools by inspecting whether the page at which a school appears in the NYC High School Directory affects the application rates to the school's programs. According to Sattin-Bajaj et al. (2018), guidance counselors reported that the printed directory, which is about 600 pages long, is the main source of information for the applicants. In the directory, schools are first grouped into the five boroughs of NYC, and within each borough, they are ordered alphabetically by their names. If the alphabetical ordering is independent of unobserved tastes, lower application rates for the schools appearing on later pages would suggest that the students are not considering all the schools.

Table 3 reports estimates from a probit model predicting whether a student applies to a program, focusing on the effect of the positioning of the program's school in the directory. A student is considered to have applied to a program if that program appears anywhere in their rank-ordered list. Page rank denotes the within-borough rank of the schools in terms of the order in which they are listed in NYC's High School Directory. We first focus on the results for the *All eligible* sample, which includes all programs each student is eligible for. Columns (1) and (2) show that the ordering significantly impacts the application rates. Moving a school's position backward by 100 page ranks (typically equating to about 125 pages) is associated with a 24.16% decrease in application rates, even after controlling for a

[25]The sample includes the students who opted out of the school choice process, who constitute 8.06% of the sample. There are some ninth graders who participate in the process, but they constitute 0.01% of the total applicants, and they can apply to only a subset of the schools.

Table 3: Regression of application on page rank and controls

	Dependent variable: Student applies to the program					
	Sample: All eligible					
	(1) All	(2) All	(3) Asian	(4) Black	(5) Hispanic	(6) White
Page rank / 100	−0.144***	−0.123***	−0.124	−0.151***	−0.114***	−0.090
	(0.052)	(0.040)	(0.077)	(0.041)	(0.039)	(0.065)
Mean application rate	1.1248%	1.1155%	1.1499%	1.2307%	1.1180%	0.9114%
ADE (page rank from 1 to 100)	-0.42%p	-0.27%p	-0.24%p	-0.39%p	-0.26%p	-0.14%p
ADE (div. by mean app rate)	-37.59%	-24.16%	-21.17%	-31.57%	-23.32%	-15.60%
Controls	No	Yes	Yes	Yes	Yes	Yes
Observations	11,210,474	9,114,710	1,681,067	2,129,089	3,752,117	1,552,437
Log Likelihood	-690,170.962	-394,223.478	-63,291.173	-108,462.764	-170,079.344	-47,781.523
AIC	1,380,345.924	788,576.955	126,682.345	217,025.528	340,258.687	95,663.046
	Sample: Near					
	(7) All	(8) All	(9) Asian	(10) Black	(11) Hispanic	(12) White
Page rank / 100	−0.123	−0.024	0.159	−0.030	−0.101	0.158
	(0.084)	(0.073)	(0.157)	(0.077)	(0.071)	(0.149)
Mean application rate	12.9080%	13.3124%	15.2884%	11.4853%	12.8271%	16.3963%
ADE (page rank from 1 to 100)	-2.57%p	-0.40%p	2.44%p	-0.48%p	-1.71%p	2.55%p
ADE (div. by mean app rate)	-19.92%	-3.01%	15.99%	-4.16%	-13.30%	15.53%
Controls	No	Yes	Yes	Yes	Yes	Yes
Observations	63,085	51,408	8,683	13,422	23,131	6,172
Log Likelihood	-24,243.188	-15,725.585	-2,440.372	-3,934.720	-7,149.979	-1,806.911
AIC	48,490.376	31,581.170	4,978.743	7,969.441	14,399.957	3,713.822

Notes: An observation is a student-program pair. A random sample of 20,000 students was used. Programs that each student is ineligible for were dropped. *All eligible* indicates the sample of all student-program pairs for which the student is eligible for applying for the program. *Near* sample uses only the student-program pairs for which the program is within a half mile from the student's home or a quarter mile from the student's middle school. The *page rank / 100* variable indicates the within-borough rank of the school in terms of the order in which they are listed in the school directory, divided by 100. Standard errors are clustered at the program level. Appendix E.5 describes the controls used.

rich set of observables (Appendix E.5), as suggested by the average difference effect (ADE) relative to the mean application rate. For the purpose of exposition, we define the ADE as the average increment in the (predicted) inclusion probability when each school's position is first set at the frontmost position and then moved to the 100th position, while holding other covariates fixed.

The results also hint at disparities in information. Separate estimates by ethnicity using the *All eligible* samples suggest that the negative associations are stronger and significant for Black and Hispanic students. Such patterns may emerge if Asian and White students have better information sources or better neighborhood schools and can therefore rely less on the directory. Our main analysis suggests Asian and White students' consideration sets are more aligned with their preferences (e.g., see Table 6). We also find they have better neighborhood schools (Figures 1 and B.1).

We now assess the assumption that page rank is uncorrelated with unobserved preferences. Table B.1 regresses the page rank on observable school characteristics. The F-statistic has a p-value of 0.163, and no coefficients are found to be significant at the 5% level, showing that page rank is largely uncorrelated with preferences as captured by observable school characteristics. The results for the *Near* samples in Table 3 also support the assumption. The samples in these columns consist only of the student-program pairs for which the high school program is within a half mile from the student's home or within a quarter mile from the student's middle school. If students were applying less to later-page schools in the columns for the *All eligible* samples due to unobserved preferences, such negative associations should continue to appear in the *Near* samples. On the other hand, if students are not applying to these schools due to a lack of awareness, the association should tend to disappear in these samples, given that students are likely aware of the schools near their homes or middle schools. Our findings align closely with the latter scenario. Conversely, if we take as given that alphabetical ordering is independent of preferences, the results for the *Near* samples support the assumption that the students are indeed aware of these nearby schools. We utilize this assumption to estimate the model of application behavior in Section 6.

Table 4 summarizes OLS regressions of applications on admissions priority groups. We restrict these regressions to students who did not exhaust their lists, implying that the list length constraint is not binding for them.[26] In such cases, a weakly dominant strategy is truthful reporting in the order of preferences independent of beliefs about admission probabilities, rendering priorities inconsequential apart from potential correlation with preference.

Yet the results demonstrate a substantial influence of priorities on whether the student

[26]The regression uses admissions priority groups we reconstruct, since the observed priority numbers are only available for student-program pairs for which the student applies for the program, as described in Section E.1. However, we find 19.59% of the observed priority numbers are incorrectly predicted (based on a sample of 20,000 students), 97.05% of which involve attendance at information sessions. Because of this, we further exclude 238 programs (out of 743) that employ (as a device to determine priority groups) "attendance at information sessions."

Table 4: Regressions of Application on Priority Group

	Dependent variable: student applies to the program			
Sample	All	All	Near and likely	Near and likely
Priority	-0.002***	-0.026***	-0.055***	-0.091***
	(0.0001)	(0.005)	(0.005)	(0.025)
Mean application rate	0.88%	0.88%	15.07%	15.07%
Controls	Yes	Yes	Yes	Yes
Program fixed effects	No	Yes	No	Yes
Observations	8,669,191	8,669,191	31,623	31,623
R^2	0.0567	0.0494	0.32	0.164
Adjusted R^2	0.0567	0.0494	0.319	0.158

Notes: *p<0.1; **p<0.05; ***p<0.01. An observation is a student-program pair. Students who listed 12 programs, ineligible student-program pairs, and programs with information sessions are dropped. Standard errors are clustered at the program level. See Appendix E.5 for details.

lists a program; lower-priority students (i.e., those with higher values of the *priority* variable) tend not to list the program. This effect holds true irrespective of whether we account for the potential correlation of priorities with unobserved program quality via program dummies. The inclusion of fixed effects intensifies the effect of priorities, suggesting that programs that attract students for reasons unexplained by observables often accommodate more priority groups. Even when we narrow down the analysis to the *Near and likely* sample, consisting of student-program pairs where students presumably perceive high admission chances and the students are likely to know that because the programs are near their home or middle school,[27] the effect persists albeit at a weaker strength relative to the mean application rates, using our preferred estimates with program dummies. Overall, the analyses suggest that students may omit schools they feel are out of reach for them. Our model of school applications therefore allows for such possibility.

Given indications that students consider admission chances when deciding whether to list a program, it is also of interest to explore whether these chances also affect the rankings in the report. Having fixed which programs to list, factoring in admission chances when ranking the programs cannot benefit the students, regardless of whether the list length constraint binds (Haeringer and Klijn, 2009). Table 5 presents OLS regressions of the submitted rank of a program on the student's priority group, using a sample of applicants who listed a given number of programs. The regressions include program dummies and other controls (Appendix E.5). The effects of priorities on rankings are somewhat mixed and are milder than

[27]More specifically, these are the student-program pairs such that the program is within a half mile from the student's home or a quarter mile from their middle school and satisfies one of the following criteria: (1) the program did not fill its seats in the prior year, (2) the student belongs to the program's first priority group and the percent of offers that went to this group in the prior year is less than 90% (as stated in the high school directory), or (3) the student scored higher than 350 in both the NY State Math and ELA tests; 4.18% of students satisfy the last criterion.

Table 5: Regression of Submitted Rank on Priority Group

	Dependent variable: rank in submitted report					
No. of listed programs	2	4	6	8	10	12
Priority	-0.138	-0.048	0.124	0.371***	0.443***	0.477***
	(0.091)	(0.083)	(0.086)	(0.118)	(0.137)	(0.104)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Program fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,750	14,858	29,666	36,794	33,316	130,460
R^2	0.0246	0.0152	0.0122	0.0157	0.0172	0.0154
Adjusted R^2	-0.122	-0.031	-0.0113	-0.00316	-0.00361	0.0101

Notes: *p<0.1; **p<0.05; ***p<0.01. An observation is a student-program pair. Refer to Appendix E.5 for the controls used. Standard errors are clustered at the program level.

their effects on the decision to list a program. For instance, among students listing twelve programs, moving to the next less-preferred priority group (a unit increase in the variable *priority*) results in a drop of 0.477 in rank (0.477/11 = 4.3% of the available variation in rank). However, for those listing only two programs, the same shift in priority tends to raise (albeit insignificantly) the program's rank by 0.138 slots (0.138/1 = 13.8% the of available variation). This seemingly counterintuitive result, especially when viewed alongside Table 4, may stem from selection bias; the sample only includes listed programs. If a lower-priority program makes it to the list despite being unlikely (as seen in Table 4), it suggests a strong preference for that program.

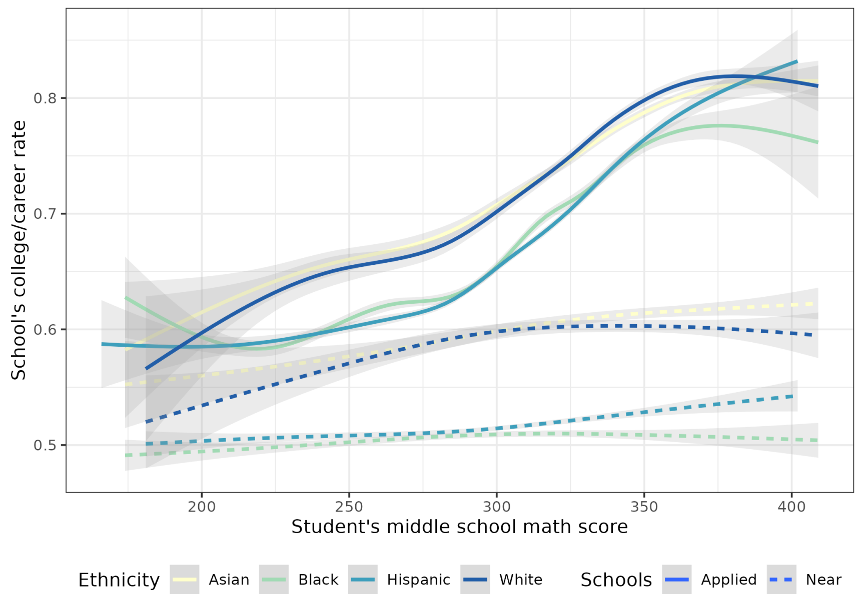
Overall, it is difficult to definitively conclude whether admission chances influence how a student ranks a program based on Table 5. We further explore this possibility using the model of students' application behavior in Section 4. The model also takes into account our descriptive observations that students may not be aware of all the schools and that they seem to factor in chances of admission when choosing which schools to list.

3.3 Patterns of Disparities and Choice

Figure 1 documents some patterns of racial disparities in neighborhood schools and the usage of choice. First, focusing on the dashed curves, we observe substantial racial disparities in the applicants' neighborhood schools (within a mile from home). These disparities do not disappear even after controlling for applicants' performance in the middle school mathematics tests; the neighborhood schools of even the best-performing Black and Hispanic students have lower college/career rates than those of the lowest-performing Asian and White students.

With the solid curves, we gather several patterns regarding how students utilize school choice. Applicants do take advantage of school choice and apply to higher-performing

Figure 1: Nearby and Applied Schools, by Ethnicity



Notes: *College/career rate* denotes the school's proportion of students who enrolled in college, a vocational program, or a public service program within six months of graduation. *Student's middle school math score* is the applicant's performance in the New York State Math test during seventh grade. The lines represent smoothed conditional means, using cubic regression spline with shrinkage. The shaded regions represent 95% confidence intervals. The dashed lines are drawn using the schools within one mile from the applicant's home address. The solid lines are drawn using the schools that the applicant has listed on the submitted rank-order report. A sample of 20,000 students is used.

schools.[28] In terms of the schools that applicants apply to, the racial disparities appear reduced. High-performing applicants are more likely to apply more aggressively to high-performing schools. These patterns could be explained by differences in preferences, in awareness, or in assessments about admission chances.

4 Model of Students' Application Behavior

This section lays out our main model: how students apply to schools. In our model, students maximize expected utility subject to two types of optimization frictions. First, they may consider only a limited set of school programs due to a lack of awareness or the perception that they have no chance of being admitted. Second, even when they consider the programs, they may still have incorrect beliefs about the equilibrium assignment probabilities. In particular, these incorrect beliefs may reflect students' misunderstandings of the properties of DA.

Specifically, a school program is *considered* by an applicant if (1) he is aware of that

[28]Figure B.1 shows that students are typically *matched* to schools whose characteristics fall between those of the neighborhood schools and those of the schools they apply to.

school program, and (2) he feels the school is reachable, i.e., that he has a positive chance of assignment to that school program upon listing it.[29] The *consideration set* of applicant i, which is the set of school programs considered by applicant i, is denoted by $\mathcal{C}_i$. Consideration of school program j by applicant i is determined by a latent variable $c_{ij} \in (-\infty, \infty]$.[30] A school program is considered if and only if $c_{ij} > 0$. We assume that students do not consider ineligible schools.

Formally, each applicant i solves

$$\max_{r \in \mathcal{R}(\mathcal{C}_i)} \sum_{j=0}^{J} p_{ij}^r v_{ij} \tag{4.1}$$

where r denotes a report, $j \in \{1, \cdots, J\} \equiv \mathcal{J}$ denotes a school program that is potentially matched through the application procedure, $j = 0$ is the outside option,[31] $p_{ij}^r \in [0, 1]$ denotes i's subjective assessment of the probability of being assigned to j upon submitting report r, and v_{ij} is the utility that i derives from being assigned to j. The applicant, with the consideration set $\mathcal{C}_i$, chooses a report r from $\mathcal{R}(\mathcal{C}_i)$, the set of all ordered lists of school programs in $\mathcal{C}_i$ with length at most 12, including an empty list denoted by $r = \emptyset$; formally, $\mathcal{R}(\mathcal{C}_i) \equiv \{\emptyset\} \cup \bigcup_{k=1}^{12} \left\{(j_1, \ldots, j_k) \in \mathcal{C}_i^k | j_m \neq j_n \text{ for } m \neq n\right\}$. The empty list represents non-participation in the main (first) round of the application process. Although r is an ordered list, we occasionally abuse notation and treat r as if it were an (unordered) set; for instance, we write $j \in r$ to denote that j is written somewhere in r, regardless of its position in r. The solution to the maximization problem in Equation 4.1 is denoted by r_i. Multiple solutions occur with probability zero under our assumptions and are ignored.

We model nondegenerate[32] beliefs about assignment probabilities similarly to Kapor et al. (2020), which is motivated by the cutoff and score representation of the matching algorithms (Agarwal and Somaini, 2018; Azevedo and Leshno, 2016). The representation uses two quantities: score_{ij} and cutoff_{ij}. Being a function of admissions priority groups, screening

[29]This definition differs from the typical definition of consideration in the discrete choice literature in that we impose (2) in addition to (1). However, the imposition of (2) is natural in this setting where assignments are stochastic at the time of reporting. Furthermore, (the lack of) consideration may be interpreted to additionally capture some factors other than awareness and zero admission chances: fear of rejection, risk aversion, or the (psychological) cost of application. In other words, the model of consideration intends to capture any reason other than preferences that might prevent a student from listing a school program. We focus on awareness and degenerate assignment probabilities as the main reasons why students may drop the school program from the list, as evidence suggests these channels are significant.

[30]In Sections 5 and 7, we will assume that there are certain school programs that are surely considered by an applicant; such a school program is denoted by $c_{ij} = \infty$ for notational convenience.

[31]The outside option is interpreted as the inclusive value of remaining unassigned in the main round of the application process.

[32]Zero admission chances are modeled through consideration. Upon consideration, the students have nonzero admission chances. In the paper, *beliefs* refer to the beliefs about admission chances upon consideration, implying positive subjective admission chances.

rankings, and lotteries, score_{ij} represents program j's evaluation of applicant i, with a lower score denoting higher preference. One important aspect of DA is that score_{ij} is *not* a function of the student's submitted ranking of the school program.

On the other hand, the student-type-specific $\mathsf{cutoff}_{ij} \equiv \mathsf{cutoff}_j(\mathsf{type}_i)$ determines how many students of type_i are admitted by program j. In NYC, type_i indicates whether the student has disabilities. Separate capacities are set for each type.[33]

Under the cutoff-score representation, each student is matched to his first school program in the list for which score_{ij} is below cutoff_{ij}. That is,

$$\begin{gathered} i \text{ is matched to } j \\ \Leftrightarrow j \text{ is the highest-ranked school program in } r_i \text{ for which } \mathsf{cutoff}_{ij} - \mathsf{score}_{ij} > 0. \end{gathered}$$

We model beliefs about the assignment probabilities based on this representation. Each student forms subjective assessments of his $\mathsf{cutoff}_{ij} - \mathsf{score}_{ij}$ for each school program j. For student i, his assessment of $\mathsf{diff}_{ij} := \mathsf{cutoff}_{ij} - \mathsf{score}_{ij}$ is represented by the student-specific random variable $\widetilde{\mathsf{diff}}_{ij}(k) := \widetilde{\mathsf{cutoff}}_{ij} - \widetilde{\mathsf{score}}_{ij}(k)$, where k denotes the rank of j in i's report. The randomness in $\widetilde{\mathsf{diff}}_{ij}(k)$ represents the student's perceived uncertainty about the scores and the cutoffs. Note that the distribution of $\widetilde{\mathsf{score}}_{ij}(k)$ can depend on the rank k; although programs' ranking in students' reports cannot affect the scores in DA, we allow for the possibility that students may not understand this property.[34] On the other hand, we do assume that applicants are monotone in their misunderstanding; while they might mistakenly believe that ranking a school higher can improve their scores, they correctly understand that ranking a school program lower cannot. Formally, we assume $k < k'$ implies $\widetilde{\mathsf{score}}_{ij}(k) \leq \widetilde{\mathsf{score}}_{ij}(k')$ for all (i, j) in any realization.

Using the scores-and-cutoffs representation, we model nondegenerate subjective beliefs as follows. For program j listed in report r,

$$p^r_{ij} = \mathbb{P}_i(\widetilde{\mathsf{diff}}_{ij'}(k^r_{j'}) < 0 \text{ for all } j' \text{ listed before } j)\mathbb{P}_i(\widetilde{\mathsf{diff}}_{ij}(k^r_j) > 0) = \prod_{l=1}^{k-1}(1 - q_{ij_{r_l}l})q_{ijk} \quad (4.2)$$

where k^r_j denotes the rank of j in report r, q_{ijk} denotes $\mathbb{P}_i(\widetilde{\mathsf{diff}}_{ij}(k) > 0)$, and j_{r_l} denotes the school program listed at the lth spot in r. The formulation implicitly assumes that students regard admissions into programs as independent events conditional on the observables to the students, as the (subjective) joint distribution of $\left(\widetilde{\mathsf{diff}}_{ij}(k^r_j)\right)_{j\in r}$ is the product of marginal distributions. This assumes away the possibility that students may believe that the screening

[33] For school programs that employ the *educational option* admission method, type also depends on the applicant's reading category as determined by the English Language Arts (ELA) score in the middle school.

[34] If students correctly understood that the rank cannot affect the scores, they would always truthfully rank the school programs among those listed. Even in such a case, however, it is still possible that some unlisted program is considered and preferred to a listed program if the list length constraint binds.

outcomes of the programs can be unobservably correlated or that they correctly understand that a single lottery is used to break ties for all lottery-based programs. This simplifying assumption allows us to reduce dimensionality in representing the optimal report choice problem in Equation 4.1 as a "dynamic" problem solvable through backward induction as in Calsamiglia et al. (2020) (Appendix G.1). This makes the computation feasible even with a vast choice set of rank-ordered reports.

For j not listed in the report r, $p^r_{ij} = 0$; that is, the student correctly believes that he cannot be matched to j in the main round unless he lists it in the report.

5 Identifying Preferences, Consideration, and Beliefs

Here, we outline an intuitive overview of the identification strategy, demonstrating how the three channels in our model—preferences, consideration, and nondegenerate beliefs—can be separated out. These ideas are formalized in Appendix A, where we develop sufficient conditions for nonparametric identification.

We first demonstrate that there is variation in the data that is affected only by preferences and consideration, and not by nondegenerate beliefs: (1) the number of school programs in an applicant's list and (2) whether a program is written on an applicant's list, provided that the applicant's list contains strictly fewer than 12 programs.

Observation 1 (Variation reflecting only preferences and consideration)**.** *Suppose applicant i's list r_i has strictly fewer than 12 school programs. Then, $j \in r_i$ if and only if both $c_{ij} > 0$ and $v_{ij} > 0$. Furthermore, r_i has strictly fewer than 12 programs if and only if $\{j \in \mathcal{J} | v_{ij} > 0, c_{ij} > 0\}$ has strictly fewer than 12 programs.*

The proof is given in Appendix D.3. Intuitively, if you are not constrained by the length constraint and you consider a program (thus aware of the program and perceive it as reachable), you have no reason to drop it from your report, as long as you prefer it to the outside option. Conversely, if you do not prefer it to the outside option or do not consider it, you will not list it.

Given that Observation 1 shows that there is data variation that is strictly affected by preferences and consideration, a natural question is whether there is also variation that can be used to disentangle preferences from consideration. Intuitively, such separation may be possible if (1) there were some school programs that are "surely" considered by an applicant or if (2) there were shifters of consideration that were excluded from utilities. We define the *surely considered set* of applicant i as the set of programs assumed to be (surely) considered by applicant i. It is denoted by $\mathcal{S}_i$, and $\mathcal{S}_i \subseteq \mathcal{C}_i$ with probability 1. The following observation, which follows as a corollary of Observation 1, is helpful in separating preferences and consideration using the surely considered sets.

Observation 2 (Variation only reflecting preferences)**.** *Suppose applicant i's list r_i has strictly fewer than 12 school programs and that $j \in \mathcal{S}_i$. Then, $j \in r_i$ if and only if $v_{ij} > 0$.*

Combined, Observations 1 and 2 provide the basis for separately identifying preferences and consideration. Intuitively, one may first identify preferences using Observation 2 and then identify consideration using Observation 1. Propositions A.1 and A.2 in Appendix A formalize the intuition here by providing sufficient conditions under which the distributions of preferences and consideration sets are nonparametrically identified. These results also clarify how the consideration instruments that are excluded from preferences, which we did not utilize in Observations 1 and 2, aid in identification.

We discuss how the potential selection issues—Observations 1 and 2 only utilize the students who do not exhaust all slots in the report in the two Observations—are resolved by an independence assumption in Section 7.1. Propositions A.1 and D.1 show the conditions under which the selection issues regarding the exhaustion of the slots do not arise without the independence assumption.[35]

To identify nondegenerate beliefs, we may use two kinds of remaining variation in the data. First, in Observations 1 and 2, we did not utilize the information in how the applicants *ordered* the programs; we used only the information of whether programs were *listed*. Second, we have not yet utilized the variation in the portfolio choices of applicants who had more than 12 considered programs that they preferred to the outside option. These aspects of data variation are affected by beliefs in addition to preferences and consideration.

Observation 3 (Variation reflecting nondegenerate beliefs)**.**

(i) Suppose that the applicant has more than 12 programs that are considered and preferred to the outside option. Then, the identities of the programs in r_i are determined as a function of $\big(v_{ij}, c_{ij}, (p^r_{ij})_{r \in \mathcal{R}(\mathcal{J})}\big)_{j \in \mathcal{J}}$. In particular, the function is not constant in $(p^r_{ij})_{j \in \mathcal{J}, r \in \mathcal{R}(\mathcal{J})}$.

(ii) Suppose that r_i contains at least two programs. Then, r_i is determined as a function of $\big(v_{ij}, c_{ij}, (p^r_{ij})_{r \in \mathcal{R}(\mathcal{J})}\big)_{j \in \mathcal{J}}$. In particular, the function is not constant in $(p^r_{ij})_{j \in \mathcal{J}, r \in \mathcal{R}(\mathcal{J})}$.[36]

In a restricted setting, Proposition A.3 outlines the conditions for nonparametric identification of beliefs. As nonparametric identifiability in a general setting is ambiguous, our

[35]The key is that these results utilize the presence of a shifter of consideration (excluded from utilities) in addition to the surely considered schools, unlike in the Observations. Proposition D.1 further assumes the presence of a utility shifter that is excluded from consideration. On the other hand, case (ii) of Proposition D.2, which does not utilize the excluded shifters (and rather only utilize surely considered sets), still allows us to bound the joint cumulative distribution of the utilities among the surely considered programs within an interval per each student. The average length of the intervals (across students) is approximately 0.16.

[36]From the construction of the maximization problem in Equation 4.1, report r_i and the identities in the report is a function of $(p^r_{ij})_{j \in \mathcal{J}, r \in \mathcal{R}(\mathcal{J})}$. To see examples of nonconstancy of the functions with respect to $(p^r_{ij})_{j \in \mathcal{J}, r \in \mathcal{R}(\mathcal{J})}$ under a simplified setting, see the cases in Proposition A.3 and the corresponding proof.

parametric specification of beliefs only intends to gauge the degree of truthtelling, separately for when the list length constraint binds and when it does not. Intuitively, the variation in Observation 1 and 2 identifies the distribution of preferences and consideration conditional on observables. This then determines the distribution of truthful reports among considered programs. In particular, these reports should exhibit a declining trend in predicted utilities (which can be constructed from the identified preferences) as we descend the list. By contrasting this with the diminishing rate of predicted utilities in actual reports, we can assess the degree of truthfulness in these reports (Figure B.5).

6 Empirical Specification

Student Preferences The utility v_{ij} in our empirical analysis is specified as

$$v_{ij} = x_j^v \beta^v_{\mathsf{eth}_i} + z_{ij}^v \alpha^v_{\mathsf{eth}_i} + \epsilon^v_{ij},$$

where x_j^v denotes the vector of observed program characteristics and z_{ij}^v denotes the vector of observable variables that vary across i or (i, j). The idiosyncratic taste shock is represented by $\epsilon^v_{ij} \sim_{i.i.d} N(0, 1)$, and we assume that it is independent of (x_j^v, z_{ij}^v).[37] The utilities v_{ij} are normalized in scale and location. The scale is normalized by setting the standard deviation of ϵ^v_{ij} equal to 1. The location is normalized by setting the value of the outside option to zero, i.e., $v_{i0} = 0$. Thus, v_{ij} is interpreted as the utility of j relative to 0. As we allow i-specific terms in z_{ij}, the value of the outside option relative to all the inside options can vary across these student-level observables. The parameters are specified separately according to the four ethnicities.[38] The vector x_j includes, for example, college/career rate, average middle school math achievement, ethnic composition, and program interest area dummies. The vector z_{ij} includes subsidized lunch status, distance to school, and home borough dummies.

Consideration We specify the latent variable c_{ij} as

$$c_{ij} = \begin{cases} x_j^c \beta^c_{\mathsf{eth}_i} + z_{ij}^c \alpha^c_{\mathsf{eth}_i} + \epsilon^c_{ij} & \text{if } j \notin \mathcal{S}_i \\ +\infty & \text{if } j \in \mathcal{S}_i \end{cases}$$

where $\mathcal{S}_i$ denotes the surely considered set for applicant i, which we specify below. The vector x_j^c includes observed program characteristics, and z_{ij}^c denotes the vector of observable variables that vary across i or (i, j). The idiosyncratic shock is represented by $\epsilon^c_{ij} \sim_{i.i.d}$

[37] The assumption may be mild in the sense that we do not need to regard the coefficients on (x_j^v, z_{ij}^v) as causal in the counterfactual analyses.

[38] Native American and Multi-racial students, who make up 1.6% of the sample, were grouped with White students, comprising 15% of the sample. This decision was based on the similarity in observable characteristics between these groups and the White student population.

$N(0,1)$,[39] and we assume that $(\epsilon^c_{ij})_j$ is independent of $(x^c_j, z^c_{ij}, \epsilon^v_{ij})_j$, implying that dependence of v_{ij} and c_{ij} is modeled only through observables.[40] The scale is normalized by setting the variance of ϵ^c_{ij} equal to 1. The parameters are specified separately according to each ethnicity. The parameters encapsulate the association of each characteristic with the likelihood of a student being aware of a program and perceiving it as reachable.

In our specification, the observables (x^c_j, z^c_{ij}) contain all the observables that enter utility, i.e., (x^v_j, z^v_{ij}), with a trivial exception.[41] On the other hand, there are variables that only enter (x^c_j, z^c_{ij}) but not (x^v_j, z^v_{ij}). These variables reflect the order in which the school program appears in the school directory within its borough, whether the program is located in the borough where the student lives, an indicator for the program being close to the applicant's middle school, and a proxy for applicants' admission probabilities at the program.

Specifically, the page rank variable records the order in which the program's school appears in the NYC High School Directory (ranked within its borough), which the students use as the main reference for the application process and is about 600 pages long. The schools are ordered alphabetically within their respective boroughs in this directory. Because applicants may overlook the schools that are listed later, the page rank may shift consideration. However, because the schools are ordered alphabetically, we argue that page rank is excluded from the preferences. Section 3.2 discussed how Tables 3 and B.1 are consistent with the hypothesis. We allow a program's distance from an applicant's middle school to affect consideration, as the applicant may be more aware of the schools close (within one mile) to her middle school. A student's (objective) admission probability to the program likely influences her assessment of having a positive chance of admission, and therefore we include its proxy in the consideration equation. The proxy is calculated as the difference between her objective expected scores and cutoffs; see Appendix E.2 for details. Whether a student resides in the same borough as a school program could influence the student's awareness, in part because schools are categorized by borough in the directory. It can also influence the student's subjective assessment of whether the program is reachable, as priority groups often depend on whether the student's home borough matches the program's borough.

The surely considered set $\mathcal{S}_i$ is the intersection of the two sets: (1) the programs that are within a half mile from her home or a quarter mile from her middle school, and (2) the eligible programs that are *likely* for the student *and* the student is in their first priority

[39]With the independence assumption, the model becomes an alternative-specific consideration model (Swait and Ben-Akiva, 1987). For more discussion, see Abaluck and Adams-Prassl (2021).

[40]Proposition D.1 and Agarwal and Somaini (2022) suggest joint distribution of $(\epsilon^v_{ij}, \epsilon^c_{ij})_j$ can be nonparametrically identified if there is a special regressor that shifts utility but is excluded from consideration (in addition to a shifter of consideration excluded from preferences). While we have a variable that enters only utility and not consideration—an indicator for high school being the same as the applicant's middle school—it is far from being a special regressor.

[41]An indicator for the program being in the same school as the student's middle school is in z^v_{ij} but is not in z^c_{ij}. Such a school is assumed to be surely considered (as we explain below) and therefore excluded from z^c_{ij}, which only affects those not surely considered.

group. Consistent with the usage of the term in Table 4, a program is *likely* for the student if (a) the program did not fill its seats in the prior year, (b) the student is in the program's first priority group, and fewer than 90% of students admitted in the prior year belong to this group, or (c) the student scored higher than 350 in both the NY State Math and ELA tests; 4.18% of students satisfy the last criterion. Despite the strength of being a *likely* program as a criterion for ensuring that a student feels the program is reachable, evidence in Table 4 indicates that priority groups still influence application rates. Consequently, we require further that the student be within the first priority group. The requirement that the program must be proximate to the student's home or middle school serves to ensure the student's awareness of the program and of their high (and therefore nonzero) chances of admission. This specification results in 2.16 surely considered programs per applicant on average. Note that the *surely* considered sets are entirely determined by observables, while consideration sets are jointly determined by observables and unobservables.

Beliefs Once a student considers a program, his subjective assessments of assignment probabilities are derived from his beliefs about the actual cutoffs and scores. As explained in Section 4, student's anticipation regarding the actual $\mathsf{diff}_{ij} \equiv \mathsf{cutoff}_{ij} - \mathsf{score}_{ij} \equiv \mathsf{cutoff}_j(\mathsf{type}_i) - \mathsf{score}_{ij}$ is represented by the random variable $\widetilde{\mathsf{diff}}_{ij}(k)$, where k is the rank at which the student places the program within his report (Kapor et al., 2020). We parametrize the distribution of $\widetilde{\mathsf{diff}}_{ij}(k)$ as

$$\begin{aligned}
\widetilde{\mathsf{diff}}_{ij}(k) &:= \widetilde{\mathsf{cutoff}}_{ij} - \widetilde{\mathsf{score}}_{ij}(k) \\
&= \mathsf{cutoff}_{ij} - E_{\text{obj}}[\mathsf{score}_{ij}] + \epsilon^b_{ijk} \\
&\equiv \underbrace{\mathsf{cutoff}_{ij} - E_{\text{obj}}[\mathsf{score}_{ij}] + \beta^{\mathsf{eth}_i}_{\mathsf{rank}} \log(k)}_{:=\delta^{\mathsf{diff}}_{ijk}} + \nu_{ij}
\end{aligned}$$

where $\delta^{\mathsf{diff}}_{ijk}$ represents the student's subjective (mean) prediction of $\mathsf{cutoff}_j(\mathsf{type}_i) - \mathsf{score}_{ij}$. The objective part $\mathsf{cutoff}_{ij} - E_{\text{obj}}[\mathsf{score}_{ij}]$ is calculated based on the data of the admission decisions by the programs as outlined in Appendix E.2. Roughly, we construct the expected scores based on the written rules about admissions priority groups and on the data about how the students were ranked by the programs that use screening policies. The cutoff is determined by the score of the least preferred applicant among those accepted. The subjective part—prediction bias—arises when $\beta^{\mathsf{eth}_i}_{\mathsf{rank}} \neq 0$, which implies that the students incorrectly believe that how they rank the program influences their scores in DA.

The last term $\nu_{ij} \sim_{iid} \text{Logistic}(0, \sigma^{\mathsf{eth}_i}_{\nu})$ encapsulates the student's assessment of his own prediction error; larger $\sigma^{\mathsf{eth}_i}_{\nu}$ implies more doubt about his own assessment. Students understand that prediction errors can arise for two reasons: their predictions may be biased, and there are uncertainties, such as admission lotteries, that are inherently impossible to resolve. From the perspective of the student, his subjective assessment $\widetilde{\mathsf{diff}}_{ij}(k)$ follows Logistic($\delta^{\mathsf{diff}}_{ijk}, \sigma_\nu^{\mathsf{eth}_i}$), implying

$$q_{ijk} \equiv \mathbb{P}_i(\widetilde{\mathsf{diff}}_{ij}(k) > 0) = \mathbb{P}_i\left(\frac{\nu_{ij}}{\sigma_\nu^{\mathsf{eth}_i}} > \frac{-\delta^{\mathsf{diff}}_{ijk}}{\sigma_\nu^{\mathsf{eth}_i}}\right) = \left(1 + \exp\left(\frac{-\delta^{\mathsf{diff}}_{ijk}}{\sigma_\nu^{\mathsf{eth}_i}}\right)\right)^{-1}.$$

For each ethnicity, the two parameters that govern belief are $(\beta^{\mathsf{eth}_i}_{\mathsf{rank}}, \sigma_\nu^{\mathsf{eth}_i})$. Together, they determine the degree of truthful ranking behavior and to which such behavior is affected by the list length constraint. When $\beta^{\mathsf{eth}_i}_{\mathsf{rank}} = 0$, subjectively optimal lists are truthfully ordered in terms of utilities *among the listed programs* (Haeringer and Klijn, 2009). A student may still prefer some unranked program over certain ranked programs for two reasons: (1) the student did not consider the program because he believed he had de-facto zero admission chance or was unaware, or (2) the student did consider the program, but his chances or utilities were low enough that he decided to exclude it from his twelve slots to list another program; the latter case only arises when the length constraint binds. On the other hand, if $\beta^{\mathsf{eth}_i}_{\mathsf{rank}} < 0$, the submitted rankings may not be truthfully ordered in terms of utilities even among the listed programs.[42] The level of $\sigma_\nu^{\mathsf{eth}_i}$, which governs the level of doubt the student has about his prediction, can also affect the degree of truthtelling. If $\sigma_\nu^{\mathsf{eth}_i} = \infty$, which may be understood as "giving up" on trying to predict the admission chances, then students rank the programs truthfully *among the considered programs* that are preferred to the outside option until they run out of such programs or exhaust all the 12 slots. This implies that ranked programs are always preferred to any considered but unranked program.

It follows that, when the list constraint binds, naive beliefs ($\sigma_\nu^{\mathsf{eth}_i} = \infty$) imply truthful reporting among considered programs, while deviations may occur with sophisticated beliefs ($\beta^{\mathsf{eth}_i}_{\mathsf{rank}} = 0$ and small $\sigma_\nu^{\mathsf{eth}_i}$). When it does not bind, the subjectively optimal strategies given sophisticated and naive beliefs coincide. This demonstrates a core strength of DA that there is no benefit in strategizing unless the list constraint binds.

7 Estimated Preferences, Consideration, and Beliefs

Section 7.1 outlines the estimation procedure. Section 7.2 presents a summary of the estimated model.

7.1 Estimation

Estimation proceeds in two stages. In the first stage, we estimate preference and consideration parameters using the partial likelihood of inclusion of programs in applicants' reports.

[42] We assume students understand that ranking a program lower cannot improve their scores. This rules out $\beta^{\mathsf{eth}_i}_{\mathsf{rank}} > 0$.

The second stage estimates the belief parameters using moment conditions comparing the actual and the simulated reports, taking as given the estimates from the first stage.

The first-stage partial likelihood, guided by Observations 1 and 2 (or more formally, Propositions A.1 and A.2), is a function only of preference and consideration parameters, and not the belief parameters $(\beta^{\mathsf{eth}}_{\mathsf{rank}}, \sigma^{\mathsf{eth}}_{\nu})$.[43] While Observations 1 and 2 are statements regarding the pairing of programs with all students who do not exhaust all the slots ($|r_i| < 12$), the sample in the likelihood consists only of the student-program pairs for which the j-excluded r_i has fewer than eleven programs ($|r_i \backslash \{j\}| < 11$). This condition, being stronger than $|r_i| < 12$, is to resolve the selection problem. Under our specification that $(\epsilon^v_{ij}, \epsilon^c_{ij})_{j \in \mathcal{J}}$ is $i.i.d$ across j,[44] we can show that our selection criterion $|r_i \backslash \{j\}| < 11$ is independent of $(\epsilon^v_{ij}, \epsilon^c_{ij})_{j \in \mathcal{J}}$ (Lemma F.1), so that the distribution of the unobservables are unaffected by such selection. The independence does not hold for $|r_i| < 12$. Appendix F.1 delineates the partial likelihood and shows that the true parameter vector maximizes it. We randomly sample 4,000 students per ethnicity to facilitate estimation.[45]

In the second stage, belief parameters are estimated using the Generalized Method of Moments, taking as given the first-stage estimates for preference and consideration.[46] Contrary to the first-stage likelihood, the moment conditions incorporate the students who exhausted all the twelve slots and (not only the inclusion but also) the ordering of programs in reports. The moments compare simulated and actual reports in terms of the characteristics of the programs being listed in the first top $k \in \{1, \cdots, 12\}$ slots, separately depending on whether the applicant exhausts the twelve slots. They also capture the within-list variation in the characteristics, intending to capture the degree to which applicants diversify their portfolios. These moment conditions use the identifying information in Observation 3 or that in Proposition A.3. The exact moment conditions are provided in Appendix F.2.

7.2 Estimates

Preference and Consideration In Table 6, we provide a summary of the key features of the estimated parameters (raw parameter estimates are shown in Table 7). Students

[43] The partial likelihood further has a part that depends solely on preference parameters, to take advantage of the variation suggested by Observation 2 (Appendix F).

[44] Such assumption precludes the use of a random coefficient model. However, the assumption may not be overly restrictive. Pathak and Shi (2020) find little gains in performance from allowing for random coefficients in the utilities, given the allowed heterogeneity in coefficients along the students' observed characteristics.

[45] We also weight (i, j) pairs for which i surely considers j, so that such pairs have a combined weight of 5% in the sample. In the current specification, such (i, j) pairs constitute only approximately 0.24% of the sample; we amplify their importance by weighting. Without the weighting, the level of consideration probabilities for the Asian and White students were not robust to different specifications. We hypothesize that this might be due to page rank instrument being weaker for the Asian and White students, and therefore having to rely more on other instruments or surely considered programs.

[46] The moment conditions also contain information about preference and consideration parameters. Hence, joint estimation of all the parameters using the scores of the likelihood in the first stage on top of the moments here would be more efficient. We proceed in two stages to keep the computation tractable.

Table 6: Summary of Preference and Consideration

	Asian	Black	Hispanic	White
% Programs considered	8.97%	13.83%	10.48%	6.24%
% Programs surely considered	0.22%	0.29%	0.29%	0.19%
% Programs considered among those preferred to outside option	16.84%	11.66%	13.84%	11.53%
% Programs preferred to outside option	5.56%	6.12%	6.19%	6.43%
% Programs preferred to outside option among surely considered	12.93%	10.98%	12.25%	15.86%
% Programs preferred to outside option among considered	13.23%	6.62%	9.27%	13.26%
% Programs both considered and preferred to outside option	0.96%	0.99%	0.97%	0.73%

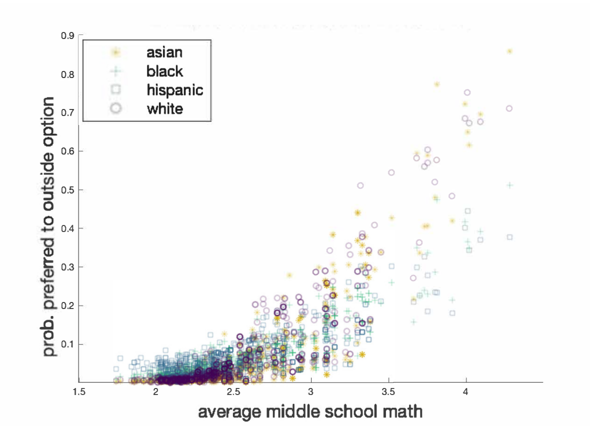
across all ethnicities are estimated to consider approximately 10.6% of programs on average. White students consider the smallest proportion of schools, potentially since their average distance to schools is the farthest (Table 1). The correlations between preference and consideration appear stronger for Asian and White students. For Black and Hispanic students, the proportions of considered programs $\Pr(c_{ij} > 0)$ are roughly equal to the proportions of considered programs among those preferred to outside option $\Pr(c_{ij} > 0 | v_{ij} > 0)$, suggesting near independence of the two events $c_{ij} > 0$ and $v_{ij} > 0$. On the other hand, for Asian and White students, the latter is roughly twice the former, indicating a positive alignment between preference and consideration. The results also show that White students are the most likely to prefer their surely considered programs.

Figure 2 summarizes preference and consideration estimates by race, illustrating significant racial differences in both channels. A point in the scatter plots corresponds to a program-race pair. Figures 2a and 2b depict the within-race average probability of a school being preferred to the outside option or being considered. Figures 2c and 2d present the within-race average predicted latent values of v_{ij} and $\tilde{c}_{ij}$, where $\tilde{c}_{ij}$ adjusts c_{ij} for the fact that sure consideration implies $c_{ij} = \infty$ by construction.[47]

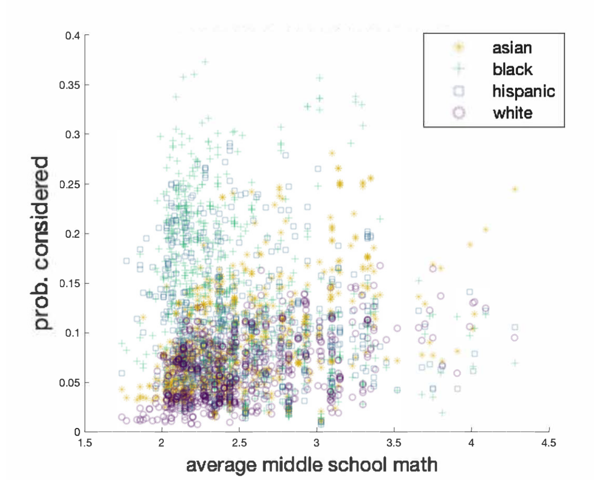
Our findings reveal that Asian and White students have stronger preferences for more selective programs, represented by the average middle school math proficiency of incoming students, compared to Black and Hispanic students. Although this trend might simply be mirroring the geographical distribution of less selective schools, which tend to be located further from Asian and White students' homes (Figure B.1), results are similar even after nullifying the effect of distance by re-calculating the latent values after setting the distance-to-school variable to 0, as we do in Figure 2e. The patterns remain substantially consistent

[47]Specifically, we use extrapolated values of $x_j^c \beta^c_{\mathsf{eth}_i} + z^c_{ij} \alpha^c_{\mathsf{eth}_i} + \epsilon^c_{ij}$ (which equals c_{ij} only for the not surely considered programs) even for the surely considered programs. Across all students and ethnicity, predicted latent variables are comparable across students and ethnicity, in the sense that they are one-to-one with probability of consideration $\Phi(\hat{\tilde{c}}_{ij})$ and of being preferred to the outside option $\Phi(\hat{v}_{ij})$, where $\hat{\tilde{c}}_{ij} = \tilde{c}_{ij} - \epsilon^c_{ij}$ and $\hat{v}_{ij} = \tilde{v}_{ij} - \epsilon^v_{ij}$.

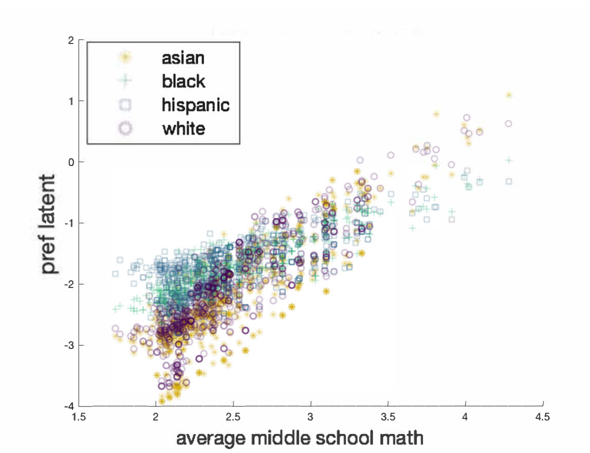
Figure 2: Probability and Latent Values for Preference and Consideration



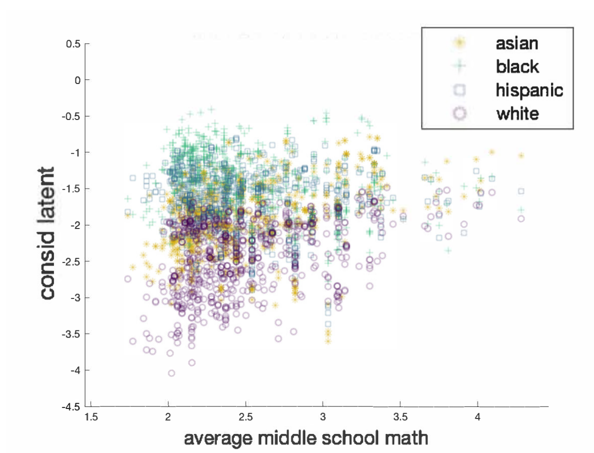
(a) Probability of Being Preferred to the Outside Option



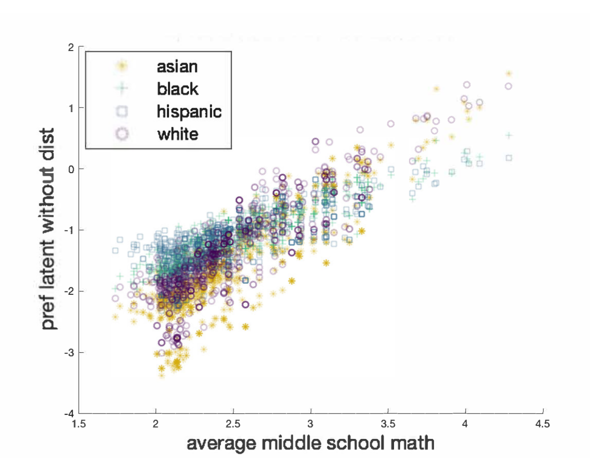
(b) Probability of Being Considered



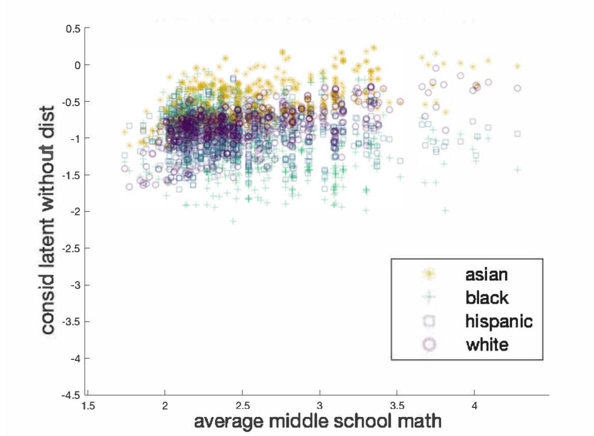
(c) Mean Latent Values for Preference



(d) Mean Latent Values for Consideration



(e) Mean Latent Values for Preference (Distance=0)



(f) Mean Latent Values for Consideration (Distance=0)

Notes: For each ethnicity, each point in the scatter plot denotes a program.

Table 7: Estimated Parameters by Race

Parameter	Asian		Black		Hispanic		White	
Preference								
Subsidized lunch	0.13	(0.01)	0.17	(0.01)	-0.01	(0.01)	0.20	(0.01)
Lives in Brooklyn	-0.51	(0.02)	-0.29	(0.02)	0.00	(0.02)	-0.52	(0.02)
Lives in Manhattan	-0.80	(0.02)	-0.19	(0.02)	-0.00	(0.02)	-1.03	(0.02)
Lives in Queens	-0.75	(0.02)	-0.30	(0.02)	-0.16	(0.02)	-0.37	(0.03)
Lives in Staten Island	0.31	(0.05)	0.01	(0.04)	0.14	(0.04)	-0.33	(0.03)
High is middle	1.80	(0.12)	1.71	(0.06)	1.87	(0.06)	1.51	(0.09)
Distance to school	-0.07	(0.01)	-0.06	(0.00)	-0.06	(0.00)	-0.07	(0.00)
Arts	-0.95	(0.04)	-0.21	(0.02)	-0.59	(0.02)	-0.79	(0.02)
STEM	-0.19	(0.02)	-0.07	(0.02)	-0.14	(0.02)	-0.13	(0.03)
College/career rate	0.70	(0.11)	0.90	(0.08)	0.74	(0.08)	0.75	(0.12)
Avg. grade 8 math proficiency (std.)	0.38	(0.02)	0.15	(0.02)	0.16	(0.02)	0.36	(0.02)
Proportion Asian	0.15	(0.08)	-0.61	(0.10)	-1.38	(0.09)	-0.63	(0.10)
Proportion Black	-1.90	(0.08)	-1.65	(0.06)	-1.95	(0.06)	-1.84	(0.09)
Proportion Hispanic	-1.45	(0.08)	-1.92	(0.05)	-1.24	(0.06)	-1.32	(0.08)
Proportion White	-0.72	(0.09)	-1.22	(0.10)	-0.75	(0.10)	0.07	(0.09)
Standard deviation of ϵ_{ij}^{v}	1		1		1		1	
Consideration								
Subsidized lunch	-0.14	(0.01)	-0.25	(0.02)	0.02	(0.02)	-0.26	(0.01)
Lives in Brooklyn	-0.41	(0.02)	-0.10	(0.02)	-0.21	(0.02)	-0.09	(0.02)
Lives in Manhattan	-0.26	(0.03)	-0.22	(0.03)	-0.23	(0.02)	0.16	(0.04)
Lives in Queens	-0.48	(0.02)	0.04	(0.03)	-0.13	(0.02)	-0.51	(0.02)
Lives in Staten Island	-0.54	(0.04)	-0.11	(0.04)	0.02	(0.05)	-0.22	(0.03)
Borough match	0.73	(0.02)	1.15	(0.02)	0.99	(0.02)	0.76	(0.02)
High is near middle	0.62	(0.03)	1.49	(0.09)	1.40	(0.07)	0.72	(0.03)
Distance to school	-0.15	(0.00)	-0.04	(0.00)	-0.08	(0.00)	-0.15	(0.00)
Arts	0.05	(0.07)	-0.22	(0.03)	0.21	(0.04)	0.13	(0.04)
STEM	0.42	(0.03)	0.03	(0.03)	0.10	(0.03)	0.13	(0.03)
College/career rate	-0.19	(0.15)	0.24	(0.13)	1.13	(0.12)	-0.31	(0.13)
Avg. grade 8 math proficiency (std.)	0.17	(0.02)	0.15	(0.02)	-0.03	(0.02)	0.22	(0.02)
Page rank in borough (std.)	-0.02	(0.01)	-0.08	(0.01)	-0.08	(0.01)	0.02	(0.01)
Proxy of objective admission probability	0.08	(0.01)	0.21	(0.01)	0.13	(0.01)	0.18	(0.01)
Proportion Hispanic	-0.07	(0.12)	-0.88	(0.10)	-1.51	(0.09)	-0.54	(0.10)
Proportion Black	-0.29	(0.13)	-0.83	(0.10)	-2.07	(0.09)	-0.63	(0.10)
Proportion Asian	-0.67	(0.10)	-2.84	(0.12)	-2.08	(0.11)	-1.20	(0.09)
Proportion White	-0.17	(0.11)	-1.81	(0.12)	-2.40	(0.11)	-0.15	(0.10)
Standard deviation of ϵ_{ij}^{c}	1		1		1		1	
Beliefs								
$\sigma_{\nu}^{\mathsf{eth}_i}$	10.68	(1.81)	12.00	(0.34)	8.03	(2.51)	9.36	(1.07)
$\beta_{\mathsf{rank}}^{\mathsf{eth}_i}$	-0.56	(0.23)	-3.89	(0.16)	-0.00	(0.12)	-0.00	(0.16)
No. student-program pairs	2,216,059		1,999,245		2,129,397		2,564,323	
No. surely considered student-program pairs	5,046		6,845		7,093		1,966	
No. students	4,000		4,000		4,000		4,000	

Notes: High is middle is an indicator of whether the student's middle school is the same as the high school. *College/career rate* indicates the proportion of students who graduated from high school four years after they entered 9th grade and then enrolled in college, a vocational program, or a public service program within six months of graduation. *High is near middle* is an indictor of a high school program being within one mile from the student's middle school. Standardized values are indicated by (std.). Intercepts are omitted for preference and consideration parameters as the ethnic compositions approximately sum to one. A random sample of 4,000 students was used for each race. The counts of (surely considered) student-program pairs include only those with $|r_i \backslash \{j\}| < 11$ for the reasons explained in 7.1.

when school selectivity is replaced with school performance gauged through college/career rates (Figure B.2) or when we use only a partial likelihood[48] that only utilizes the surely considered programs (Figure B.3).

We also note racial disparities in consideration. Black and Hispanic students are more likely to consider less selective programs than Asian and White students, while such a pattern disappears for selective programs. This pattern emerges partly because Asian and White students typically live farther from less selective programs and because distance to schools is an important determinant of consideration, especially for these groups (Table 7). The larger impact of distance on consideration for Asian and White students may also be influenced by the quality of schools in their neighborhoods. They tend to have better local schools (Figure B.1 and Figure 1), potentially reducing incentives to explore distant schools, for instance, through the school directory. Consistent with this hypothesis and mirroring the descriptive evidence in Table 3, we find that page rank affects consideration more for Black and Hispanic applicants (Table 7). White students live farther from schools in general (Table 1); after removing the effect of distance, they are the second most likely group to consider highly selective programs after Asian students (Figure 2f). If we assumed that students would be aware of highly selective programs if the distance were negligible, the racial disparities in consideration probabilities for these programs in Figure 2f would only reflect the varying perceptions among races about the reachability of these programs, and not awareness. Under the assumption, the figure suggests that Black and Hispanic students feel the highly selective programs are less reachable compared to Asian and White students.

Beliefs Our two belief parameters (per ethnicity) determine the extent of truthtelling behavior when the list length constraint binds and when it does not. Table 8 indicates that students tend to truthtell in both cases. The fractions represent how many of the simulated subjectively optimal reports from our estimated model exactly match the simulated truthful-among-considered reports.[49] We define a report as *truthful among considered* if the considered programs are ranked truthfully according to the utilities until either no more program is preferred to the outside option or all 12 slots are filled. Note that such a report may "skip" some programs that the student finds unreachable or is unaware of. The results show that, regardless of whether the constraint binds, the subjectively optimal reports approximate the truthful-among-considered reports. Section 8.3 discusses the implication of our findings about the truthfulness of reports among all eligible (not just considered) programs. As a diagnostic analysis, Figure B.5 overlays the plots of mean utilities against the rank at which the program is listed. The downward trend in utility levels along the subjectively optimal reports well approximates that of the observed reports from the data, suggesting that our parameter estimates are in a reasonable scope. The subjectively optimal reports,

[48]See Appendix F.

[49]We use the same simulated utilities and consideration sets for both types of reports.

Table 8: Fraction Truthful

List length	Asian	Black	Hispanic	White	All
	Truthful among considered				
All reports	99.7%	98.2%	99.8%	100.0%	99.4%
Full (12 programs)	97.5%	93.4%	96.8%	98.8%	95.8%
Not full	99.8%	98.5%	100%	100%	99.6%
	Truthful ordering among listed				
All reports	99.8%	98.4%	100%	100%	99.5%
Full (12 programs)	99.6%	96.8%	100%	100%	98.8%
Not full	99.8%	98.5%	100%	100%	99.6%
	Truthful inclusion among considered				
All reports	99.9%	99.8%	99.8%	100.0%	99.8%
Full (12 programs)	97.7%	96.5%	96.8%	98.8%	96.9%
Not full	100%	100%	100%	100%	100%

Notes: In the top panel, the numbers represent the fraction of (simulated) subjectively optimal reports that are indeed truthful among considered programs. The middle panel tabulates subjectively optimal reports in which the *listed* programs are written in the order of decreasing utility. The bottom panel tabulates subjectively optimal reports that list the same set of programs as truthful-among-considered reports, ignoring the ordering.

in turn, are almost indistinguishable from the truthful-among-considered reports, reflecting that the belief parameters are in line with truthtelling among considered programs.

Since we do not accommodate individual heterogeneity in truthtelling attitude within race,[50] the findings here should not be literally interpreted to imply that almost no student deviates from truthtelling. Instead, the results suggest that a representative student for each race may be viewed as essentially truthtelling.

8 Impacts of School Choice and Counterfactual Analyses

8.1 Distributional Outcomes and Decomposition

We analyze the impact of school choice on (1) racial integration and (2) the proportion of students matched to their top preferred programs by each race. We further quantify the contributions of different factors by turning off the influences of each factor one at a time. Table 9 summarizes the matchings used in these exercises.

Effects on Racial Integration We find that NYC's school choice slightly promotes racial integration, with the largest impact for Black students. Our analyses also reveal that student preferences contribute to integration, relative to neighborhood matching, net of the effects

[50] The error term ν_{ij} in the belief model is integrated out in the calculation of q_{ijk} and therefore individuals with the same observables from each race have the same q_{ijk}. This contrasts with our model of preference and consideration.

Table 9: Matching Definitions

A. Matchings without school choice						
Matching	**Matching method**					
Random	Random allocation of students to the programs with capacity constraints					
Neighborhood	Minimize total distance traveled by the students to the programs with capacity constraints					

B. Matchings with school choice						
Matching	**Simulated?**	**Preferences**	**Beliefs**	**Consideration Sets**	**Screening**	**Priority Groups**
a. Baseline matchings						
Actual	No	–	–	–	–	–
Estimated	Yes	Estimated	Estimated	Estimated	Estimated	Approximated
b. Decomposition matchings						
Change to Truthful among Considered	Yes	Estimated	**Truthful among Considered**	Estimated	Estimated	Approximated
Change to Full Consideration	Yes	Estimated	Truthful among Considered	**All eligible**	Estimated	Approximated
Change to Random Screening	Yes	Estimated	Truthful among Considered	All eligible	**Random**	Approximated
Change to No Admissions Priorities (Student-Preferences-Only Choice)	Yes	Estimated	Truthful among Considered	All eligible	Random	**None**

Notes: Approximation solution was used in the minimization for *Neighborhood* matching (Appendix G.2). *Actual* matching refers to the actual school choice matching in 2017 from the main round of DA. See Appendix E.5 for other details about the implementation of the matchings in the table.

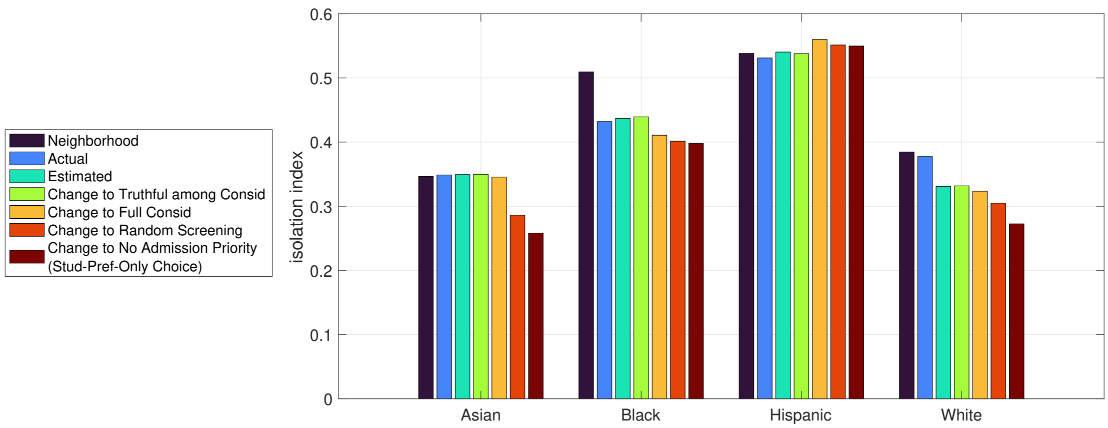
from limited information and potential nontruthful behavior. Schools' admission priorities and screening policies tend to exacerbate segregation.

In Figure 3, we measure racial segregation by the isolation index, which is the average proportion of students of the same ethnicity within each student's matched program. We see that isolation indices are similar or lower with the matchings representing school choice—Actual or Estimated—compared to Neighborhood matching. We then sequentially shut off each channel as described in Table 9. Changing the estimated beliefs to Truthful among Considered does not lead to significant changes, which is natural as the beliefs are estimated to be close to Truthful among Considered. Limited consideration is estimated to have mixed impacts across races. Schools' preferences—reflecting its screening policies and admissions priority groups—act together to segregate races.

For the decomposition exercises, note that we first deactivated the two student channels (regarding beliefs and consideration) before turning off the school channels. We chose this approach to avoid making assertions about how changes in admission policies will alter consideration sets and subjective beliefs.

While we observed both similarities and differences in preferences across races (e.g., in Figure 2), overall, student preferences work to integrate races. This is evident when we compare Student-Preference-Only School Choice allocation—resulting from DA with fully-informed students making truthful reports and programs randomly ranking the students—with Neighborhood allocation. Figure B.6 compares the density of the proportions of same-ethnicity students under Random, Neighborhood, and Actual allocation.

Figure 3: Isolation Indices by Matching—Decomposition



Notes: Each bar represents isolation index of an ethnic group in a matching. See Table 9 for the definitions of the matchings.

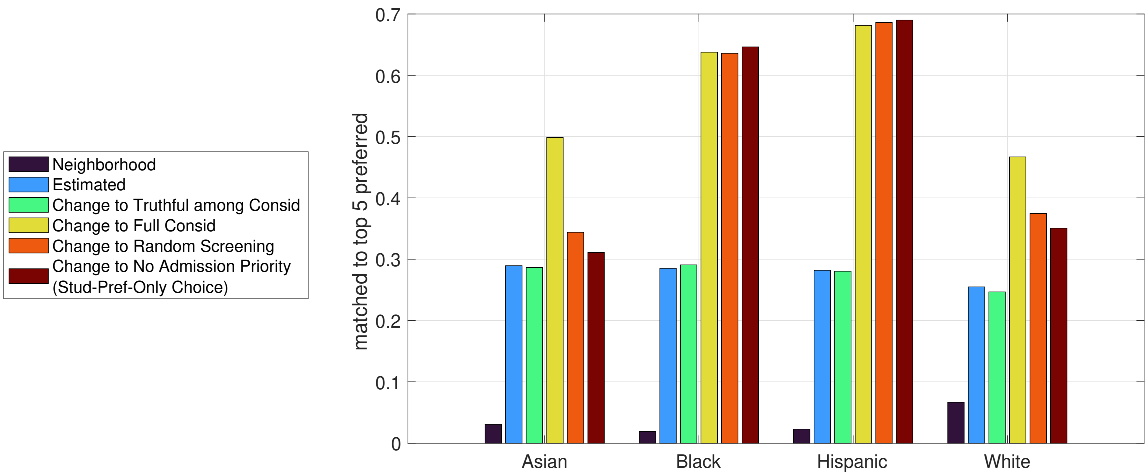
Effects on Assignment to Preferred Programs by Race We find that school choice increases the likelihood that students are matched to one of their top preferred programs, regardless of ethnicity. However, the gains are mitigated by limitations in consideration. Such limitations are more consequential for Black and Hispanic students.

Each bar in Figure 4 depicts the fraction of students who are matched to one of their top five[51] preferred programs based on their utilities. These are the top five programs they would have preferred the most if they considered all eligible programs. First, focusing on Neighborhood matching, we see that only a small fraction of the students are placed in their top five preferred programs. White students are the most likely to be matched to one of their preferred programs in this matching. We also observe that school choice—represented by Estimated—tends to increase the proportion of students placed in their top five preferred programs compared to Neighborhood matching, regardless of students' ethnicity. The improvement is large: it increases such proportion from about 2.3%–6.7% to 25.5%–28.9% on average.

Regarding the influences of different factors, we see that limited consideration substantially suppresses the proportion of students matched to one of their preferred programs. Such effect is larger for the Hispanic and Black students, which in part is because Asian and White students are more likely to consider their preferred programs (Table 6 and Figure 2). We further find that programs' screening policies tend to match Asian and White students to their preferred programs. This partially reflects the fact that Asian and White students tend to have better performance in middle school (Table 1) so that they tend to be more likely to have higher admission scores for the programs that can screen students. We also observe

[51] Figure B.7 presents the equivalent figure for top ten preferred programs.

Figure 4: Proportion Matched to Top Five Preferred Programs—Decomposition



Notes: Each bar represents the fraction of the students matched to their top five preferred programs. The sample includes both the programs that are considered and those that are not. See Table 9 and the discussions for the definitions of the matchings.

programs' admissions priorities act to place Asian and White students in their preferred programs. This may reflect that a large proportion of the admissions priorities are based on geographic proximity. Since Asian and White students live closer to higher-performing programs, they tend to be prioritized for admissions to these programs.

8.2 Designing Personalized School Recommendations

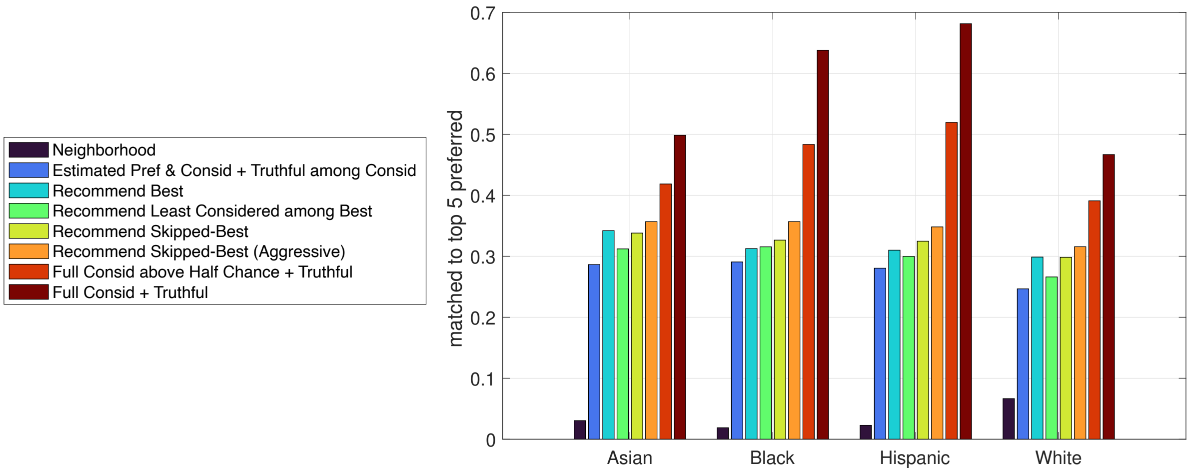
Although the preceding section showed that substantial welfare gains can arise when students consider all eligible programs, it is almost impossible to achieve it in practice. As such, this section assesses various feasible information interventions: personalized school recommendations. A judicious use of both the preference and consideration estimates turns out to be useful for designing effective interventions.

We assume that, after an intervention, students will surely consider the recommended programs in addition to those they would have already considered based on our estimated model. Based on the empirical findings which indicate that students are essentially truthfully reporting among the considered programs, we impose such truthful reporting in this section to facilitate the computation of optimal reports.[52]

We explore the simulated impacts of the following interventions, all of which recommend 30 eligible programs per student. The first three interventions only recommend programs

[52] The computation of optimal reports with subjective beliefs takes more time post-intervention due to the enlarged consideration sets.

Figure 5: Proportion Matched to Top Five Preferred Programs—Interventions



Notes: Each bar represents the fraction of the students matched to their top five preferred programs. The sample includes both the programs that are considered and those that are not. See Table 9 and the discussions for the definitions of the matchings.

with objective admission chances exceeding 50%,[53] while the last intervention (Aggressive Skipped Best) relaxes this requirement. *Best* intervention proposes the top 30 programs per student in terms of the highest predicted utilities based on the estimated parameters and the student's observable characteristics. *Least Considered among Best* intervention first curates a list of top 60 programs based on each student's predicted utilities. From this list, it recommends the 30 programs with the lowest student-specific consideration probabilities as estimated by our model. *Skipped Best* intervention is akin to Best intervention but skips the programs that are already likely to be considered (those with estimated student-specific consideration probabilities greater than 0.5) from recommendations. *Aggressive Skipped Best* intervention parallels Skipped Best intervention except that, unlike the three other interventions, it also recommends programs with objective admission chance below 50%.

Figure 5 summarizes the results.[54] The findings suggest substantial gains from some interventions. For instance, Aggressive Skipped Best recommendation is estimated to capture around 20%–36% of the welfare differences between the status quo represented by Estimated Pref & Consid + Truthful among Consid matching[55] and Full Consideration + Truthful matching.[56] This is an encouraging result, recognizing that we recommended only 30 out of

[53]This is computed as $\mathsf{cutoff}_{ij} > E_{obj}[\mathsf{score}_{ij}]$. The admission chance here corresponds to the assignment chance when the student ranks the program first in the report.

[54]Figure B.8 discusses the impacts on the isolation indices and the proportion matched to the top ten preferred programs. Aggressive Skipped Best intervention slightly increases isolation indices. The results on the matching to the top ten preferred programs are qualitatively similar to the results presented here.

[55]This is the Change to Truthful Among Considered matching in Table 9.

[56]This is the Change to Full Consideration matching in Table 9. In comparison, Full Consideration above

Table 10: Cases of Justified Envy

	Asian	Black	Hispanic	White	All
Number of school programs viewed with justified envy	3.10	2.57	3.02	3.70	3.02
% students with any justified envy	76.0%	71.3%	72.9%	74.9%	73.3%
% students who view ≥ 5 programs with justified envy	25.0%	20.6%	26.1%	31.8%	25.4%

Notes: The number of school programs viewed with justified envy is the average across students. Estimated matching (defined in Table 9) was used.

approximately 750 programs.

Notably, both preference and consideration estimates are useful for designing interventions. The highest-performing intervention, Aggressive Skipped Best, employs both preference and consideration estimates. Nevertheless, the results suggest that consideration estimates should be employed judiciously. Least Considered Among Best intervention also uses the consideration estimates in addition to preferences but performs worse than Best intervention, which solely uses preference estimates.

Aggressive Skipped Best performs better than its non-aggressive counterpart, but the result presumes that students will actually consider schools with objectively low admission chances. In practice, the actual impact of the aggressive recommendation may be better or worse than the results presented here, depending on how optimally students respond to the aggressively recommended programs.

The discussions here shed light on the potential role that economic models like ours can play in shaping information intervention strategies. There is potential for enhancing these interventions beyond our current findings. For instance, by incorporating a richer set of explanatory variables in the model of preference and consideration, the model could predict individual student preferences or considerations more accurately and thus capture a larger portion of the potential welfare gains.[57]

8.3 Empirical Assessments of the Theory-Targeted Outcomes

Matching Stability and Justified Envy To quantify matching stability, we count the cases of justified envy; a stable matching must not have any cases of justified envy.[58] We say that a student *views* an eligible program with *justified envy* if the student and the program

Half Chance + Truthful matching represents the counterfactual scenario where the students are considering all eligible programs with objective admission chances exceeding 50%.

[57]Overfitting would typically not pose a significant issue as the set of observables and their interactions would remain negligible compared to the number of student-program pairings.

[58]This is true if each program has a responsive preference over students, i.e., if there is a ranking over individual students with which it wants to fill its seats (Roth, 1985). Programs have responsive preferences in our setting; their rankings are determined by score_{ij}. Furthermore, as individual irrationality cannot arise in our model, the matching is stable if and only if there are no cases of justified envy and therefore has no blocking pair. Footnote 23 defines *individual irrationality* and *blocking pairs*.

are not matched to each other, but the student prefers the program to the current assignment and the program also prefers the student to at least one of its currently assigned student of the same type or has an empty seat for the same type. Students' preferences here are determined by the simulated utilities, which are defined regardless of whether the programs were considered or not.[59]

Table 10 shows that the students are 73% likely to have some school program viewed with justified envy, thus becoming a part of a blocking pair. However, the average number of school programs viewed with justified envy is only around three per student. Considering the presence of over 700 programs in NYC, this number is small.

Truthful Reporting Our results indicate that, among the considered programs, the students' reporting strategies approximate truthful reporting. Regarding truthful reporting among all eligible—not just considered—programs, our findings indicate that the students rarely rank a higher-utility school lower in their reports. On the other hand, students are dropping a significant number of *un*considered programs from their reports, which may be due to unawareness or perceived degenerate admission chances.

As we have seen in the top panel of Table 8, students typically rank their *considered* programs truthfully based on their utilities, until no more programs are preferred to the outside option or all 12 slots have been filled. Now we take a closer look at this by separately examining two potential deviations (that our model allows) from the truthful-among-considered reporting: (1) ranking a lower-utility program above another with higher utility, which is a weakly dominated strategy (Haeringer and Klijn, 2009); and (2) the exclusion of considered school programs from the list due to a subjectively low (albeit nondegenerate) probability of admission—the latter can only arise when length constraint binds.

The middle panel shows that students seldom play the first type of deviation: reversals of the true preference ordering in the submitted report. Note that such reversals can occur only among considered programs by design; unconsidered programs are never listed. Hence, the middle panel's figures can be interpreted as the proportions of non-reversals (truthful ordering) not only among the considered programs but among all eligible programs.

The bottom panel indicates that students rarely drop their *considered* programs due to low admission chances *out of fear of wasting their finite number of slots* in their reports. This suggests that the current 12-slot length constraint may not be overly restrictive for students. However, it is important to note that our model implies that *un*considered programs—those that students are unaware of or feel out of reach—will always be dropped from reports. As shown in Table 6, many programs are unconsidered by students. While we do not claim that

[59] Programs' preferences are determined by the expected scores $E_{\text{obj}}[\mathsf{score}_{ij}]$ of the applicants, which is a function of the admissions priority groups and, if the program can screen the applicants, the expected screening ranking given by the program for the applicant. Note that the expected value of the lottery draw is the same for each applicant and therefore does not affect the expected scores. See Appendix E.2 for the definition of expected scores.

we can definitively distinguish between the two reasons for not considering a program, our findings indicate that the latter reason is also important. For instance, we observed that admissions priority is a key determinant in predicting whether the program is included in the submitted report (Table 4) and that the proxy of objective admission probability positively affects consideration chances (Table 7). In Section 7.2, we have suggested that Black and Hispanic students seem to be perceiving the highest selectivity programs as out of reach and therefore not considering them (see the discussions about Figure 2f), which may lead them to drop such programs from their reports.

9 Conclusion

In this paper, we use data on school applications and admissions from the NYC DOE to examine the impacts of its centralized public high school choice procedure for the 2016–17 academic year. We develop and estimate a model of student application behavior that allows for two types of optimization frictions: applicants may consider only a limited set of school options and may have incorrect beliefs about admission chances. Latent preferences, consideration, and beliefs are revealed through observational data. Sources of identification include the instruments that shift consideration but are excluded from preferences, whether and where a school program is ranked in the submitted reports, and the assumption that students must consider highly likely high school programs near their home or middle school. We have also developed nonparametric identification results further clarify the sources of identification.

The empirical results show that, compared to neighborhood allocation, school choice slightly improves racial integration and markedly boosts the number of students matched to their preferred schools across all races. We delve deeper to discern the contributions of different factors. We find that admissions priorities and screening policies tend to segregate races. They make it more likely for the Asian and White students to be matched to their preferred schools.

We also find that limited consideration results in substantial negative welfare costs, especially for Black and Hispanic students. To counter the welfare loss, we investigate the potential impacts of personalized school recommendations based on the utilities and consideration probabilities predicted through our model. We find that certain recommendation policies can significantly counteract the negative welfare effects of limited consideration. Our analysis further suggests that the students rank their considered programs in a largely truthful manner.

Some key aspects highlighted in our paper align with the NYC DOE's recent policy initiatives after our analysis of the academic year 2016–17. For instance, some NYC DOE schools later adopted "Diversity in Admissions" policies, which prioritize admissions for students of lower socioeconomic status and English Language Learner students.[60] The NYC DOE also transitioned from a physical high school directory to an online version, aiming to facilitate better navigation for applicants and provide more timely and accurate information.[61]

This paper's limitations suggest avenues for future research. First, this paper currently models the students as taking the school characteristics as given from the year before the applications. Therefore, the counterfactual results presented here are best understood as either decompositions or short-run impacts. Second, this paper treats the supply of schools as given. However, the current variety of highly differentiated schools in NYC likely depends on the presence of a large-scale school choice program. Exploring how school supply is affected by the presence of school choice, and its implications for welfare and distributional outcomes, presents a compelling research opportunity.

Appendix

A Nonparametric Identification

In this section, we provide sufficient conditions for the nonparametric identification of the model. The main results are provided here, and Appendix D.1 provides additional results under stronger and weaker sets of assumptions. Proofs are in Appendix D.3.

In stating the nonparametric identification results, we do not make any parametric assumption about utilities, latent consideration variables, and beliefs $(v_i, c_i, p_i) \equiv \big((v_{ij})_{j\in\mathcal{J}}, (c_{ij})_{j\in\mathcal{J}}, (p^r_{ij})_{r\in\mathcal{R}(\mathcal{J}), j\in\mathcal{J}}\big)$ as made in Section 6. Furthermore, we do not assume that the maximum allowed list length, denoted L, has to equal 12.

On the other hand, we do assume the following for every result. First, we assume that beliefs are generated by students making anticipations about differences in their scores and cutoffs, in the sense that Equation 4.2 holds. Second, we assume that perceived scores are increasing in submitted rank as in Section 4. Third, we assume that the distribution of $v_i|z_i$ is continuous for every $z_i \in \text{supp}(z_i)$ and that $q_{ijk} \equiv \mathbb{P}_i(\widetilde{\mathsf{diff}}_{ij}(k) > 0) \in (0,1)$ for every considered schools.

To discuss the results, we define two concepts: an extreme consideration shifter excluded from preferences and a special regressor with large support (Thompson, 1989; Lewbel, 2000).

Definition 1. Let $z_i \equiv (a_i, z_i^-)$. A $J-$dimensional random vector a_i is called an **extreme consideration shifter excluded from preferences** if $v_i \perp\!\!\!\perp a_i$ conditional on z_i^- and, for

[60] See https://www.schools.nyc.gov/docs/default-source/default-document-library/diversity-in-new-york-city-public-schools-english or https://www.schools.nyc.gov/enrollment/enrollment-help/meeting-student-needs/diversity-in-admissions.

[61] See https://www.thecity.nyc/education/2019/5/22/21211050/city-public-high-school-directory-takes-virtual-turn.

all z_i^- in its support, there exist some known $\bar{a}(z_i^-) \in \text{supp}(a_i|z_i^-)$ such that $\mathbb{P}\big(c_{ij} > 0|a_{ij} = \bar{a}_j(z_i^-)\big) = 1$.

In the empirical setting, the role of an extreme consideration shifter excluded from preferences is jointly played by surely considered sets and the excluded consideration shifters, such as page rank and distance from middle school. However, they each play an imperfect role; surely considered sets only move certain schools' consideration probabilities for each student, and the excluded consideration shifters do not move consideration probabilities to 1, i.e., to the extreme.[62]

Definition 2. A random vector z_i^y is called a **special regressor for** y_i **with large support conditional on** x_i if $y_i = \tilde{y}_i - z_i^y$ with $\tilde{y}_i \perp\!\!\!\perp z_i^y$ conditional on x_i and $\text{supp}\left(z_i^y|x_i\right) = \mathbb{R}^K$ for all x_i in its support, where K is the dimension of y_i.

In the empirical setting, the role of a special regressor is played jointly[63] by any exogenous $(i, j)-$level observables, including distance to school, and the interactions between school characteristics and the student-level observables.[64]

We first establish the nonparametric identifiability of preference. Proposition A.1 shows that the joint distribution of utilities is nonparametrically identified with a large-support special regressor for the utilities and an extreme consideration shifter.

Proposition A.1 (Identification of preferences)**.** *Suppose that we observe the following:*

(a) an extreme consideration shifter excluded from preferences, named a_i, and

(b) a special regressor for v_i, named z_i^v, with large support conditional on $z_i \backslash (z_i^v, a_i)$.

Then, the joint distribution of utilities conditional on observables, $\mathbb{P}\big(v_i \leq v\big|z_i\big)$, is identified for almost all $(v, z_i) \in \text{supp}(v_i, z_i)$.[65]

All proofs are in Appendix D.3. Intuitively, one can use the extreme consideration shifter to push the consideration probability of every school to 1, in which case the probability of listing schools becomes a sole function of the utilities. One can then use the special regressor to "trace out" the distribution of the utilities (Agarwal and Somaini, 2018). This distribution of the utilities is not conditioned on the value of the extreme consideration shifter, as it was

[62]To complement our main result, Proposition D.2 only assumes the presence of surely considered sets.

[63]We conjecture that results in Berry and Haile (2020) may be used to formally show how different variables can form an index that mimics the role of a special regressor.

[64]Note that most results—except case (ii) of Proposition D.1, which uses identification-at-infinity argument—can be extended to allow for limited support on the special regressor at the cost of identifying the distribution of the utilities or the latent variables for consideration on limited support.

[65]If the large support assumptions on the special regressors are weakened, then $\mathbb{P}\big(v_i \leq v\big|z_i\big)$ is also identified on a limited support.

assumed to be conditionally independent of the utilities. Note that no assumption was made about allowed list length.

Now we turn to the identification of consideration. Proposition A.2 states that the distribution of consideration indicators $c^*_{ij} := \mathbb{1}(c_{ij} > 0)$ can be nonparametrically identified with a special regressor with large support, given that the distribution of utilities are already identified (potentially through Proposition A.1). It also assumes that the allowed list length L equals the number of schools J, i.e., an applicant can list arbitrarily many schools.[66] The joint distribution of consideration indicators is point-identified if the utilities v_i are independent of latent consideration variables c_i conditional on observables. It is partially identified if the conditional independence fails.

Proposition A.2 (Identification of consideration)**.** *Suppose that $\mathbb{P}\big(v_i \leq v | z_i = z\big)$ is identified for almost all $(v, z) \in \operatorname{supp}(v_i, z_i)$. Suppose that we observe a special regressor for c_i, named z^c_i, with large support conditional on $z_i \backslash z^c_i$. Suppose also that $L = J$. Then,*

(i) if c_i is independent of v_i conditional on z_i, the joint distribution of consideration indicators conditional on observables, $\mathbb{P}\big(c^*_i \leq c^* \big| z_i\big)$, is identified for almost all $(c^*, z_i) \in$ $\operatorname{supp}(c^*_i, z_i)$.[67]

(ii) if c_i is not independent of v_i conditional on z_i, $\mathbb{P}\big((c^*_{ij})_{j \in \mathcal{A}} \leq c^* \big| (v_{ij})_{j \in \mathcal{A}} > 0, z_i\big)$ is identified for almost all $(c^*, z_i) \in \operatorname{supp}\big((c^*_{ij})_{j \in \mathcal{A}}, z_i\big)$ and for all $\mathcal{A} \subseteq \mathcal{J}$.

Remark. In relation to Proposition A.1, it is allowed that $a_i = z^c_i$ or $z^c_i = z^v_i$.[68]

The intuition for part (i) is as follows. Given that an applicant can write an arbitrarily long list, whether to list a school is a function of only utilities and consideration. However, knowing the distribution of the utilities already, the probability of schools being listed is informative only about consideration. The special regressor then traces out the distribution of c_i, the latent consideration variable, and therefore the distribution of $c^*_i = \mathbb{1}(c_{ij} > 0)$.

Now we turn to the identification of the beliefs about assignment probabilities. To present this result, we first define equivalent classes of beliefs. Two beliefs are behaviorally equivalent if they lead to the same reporting behavior conditional on any realization of the utilities and the consideration sets:

Definition 3. Two beliefs $\{p^r_j\}_{j \in \mathcal{J}, r \in \mathcal{R}(\mathcal{J})}$ and $\{p'^r_j\}_{j \in \mathcal{J}, r \in \mathcal{R}(\mathcal{J})}$ are **behaviorally equivalent** if for all $v \in \mathbb{R}^J$ and $\mathcal{C}_i \subseteq \mathcal{J}$, $\arg\max_{r \in \mathcal{R}(\mathcal{C}_i)} v \cdot p^r = \arg\max_{r \in \mathcal{R}(\mathcal{C}_i)} v \cdot p'^r$.

where $(p^r) = (p^r_j)_{j \in \mathcal{J}}$ and similar for (p'^r). The notion of behavioral equivalence relates to the notion of normalization and is distinct from observational equivalence.

Here we state the identification result on beliefs, which holds under a restricted setting.

[66] Proposition D.1 presents a result with length constraints with stronger data requirements.

[67] If the large support assumptions on the special regressors are weakened, then $\mathbb{P}\big(c^*_i \leq c^* \big| z_i\big)$ are also identified on a limited support.

[68] On the other hand, it is not possible that $a_i = z^c_i = z^v_i$.

Proposition A.3 (Identification of beliefs)**.** *Suppose that* $\mathbb{P}\big(v_i \leq v, c_i^* \leq c^* | z_i = z\big)$ *is identified for every* $(v, c^*, z) \in \operatorname{supp}(v_i, c_i^*, z_i)$*. Suppose that either (1)* $L = J = 2$*, or (2)* $L = 1$*. Suppose also that beliefs are constant given observables, i.e.* $p_{ij}^r = p_j^r(z_i)\ \forall (i, j, r)$*. Then, beliefs* $\{p_j^r(z_i)\}_{j,r}$ *are identified up to behaviorally equivalent classes.*

References

Abaluck, J. and Adams-Prassl, A. (2021). What do consumers consider before they choose? identification from asymmetric demand responses. *The Quarterly Journal of Economics*, 136(3):1611–1663.

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

Abdulkadiroğlu, A., Che, Y.-K., Pathak, P. A., Roth, A. E., and Tercieux, O. (2020). Efficiency, justified envy, and incentives in priority-based matching. *American Economic Review: Insights*, 2(4):425–442.

Abdulkadiroğlu, A., Pathak, P. A., and Roth, A. E. (2005). The New York City High School Match. *American Economic Review*, 95(2):364–367.

Abdulkadiroğlu, A., Pathak, P. A., and Roth, A. E. (2009). Strategy-Proofness Versus Efficiency in Matching With Indifferences: Redesigning the NYC High School Match. *American Economic Review*, 99(5):1954–1978.

Abdulkadiroğlu, A. and Sönmez, T. (2003). School Choice: A Mechanism Design Approach. *American Economic Review*, 93(3):729–747.

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

Agarwal, N. and Somaini, P. J. (2022). Demand analysis under latent choice constraints. *NBER Working Paper*, number 29993.

Aguirregabiria, V. (2021). Identification of firms' beliefs in structural models of market competition. *Canadian Journal of Economics/Revue canadienne d'économique*, 54(1):5–33.

Ajayi, K. and Sidibe, M. (2022). School choice under imperfect information. *Working Paper*.

Ajayi, K. F., Friedman, W. H., and Lucas, A. M. (2017). The Importance of Information Targeting for School Choice. *American Economic Review*, 107(5):638–643.

Akbarpour, M., Kapor, A., Neilson, C., Van Dijk, W., and Zimmerman, S. (2022). Centralized school choice with unequal outside options. *Journal of Public Economics*, 210:104644.

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

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

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

Artemov, G., Che, Y.-K., and He, Y. (2017). Strategic 'Mistakes': Implications for Market Design Research. *Working Paper*.

Ashlagi, I. and Gonczarowski, Y. A. (2018). Stable Matching Mechanisms Are Not Obviously Strategy-Proof. *Journal of Economic Theory*, 177:405–425.

Avery, C. and Pathak, P. A. (2021). The distributional consequences of public school choice. *American Economic Review*, 111(1):129–52.

Azevedo, E. M. and Leshno, J. D. (2016). A Supply and Demand Framework for Two-Sided Matching Markets. *Journal of Political Economy*, 124(5):1235–1268.

Barseghyan, L., Coughlin, M., Molinari, F., and Teitelbaum, J. C. (2021a). Heterogeneous choice sets and preferences. *Econometrica*, 89(5):2015–2048.

Barseghyan, L., Molinari, F., and Thirkettle, M. (2021b). Discrete choice under risk with limited consideration. *American Economic Review*, 111(6):1972–2006.

Basteck, C. and Mantovani, M. (2018). Cognitive Ability and Games of School Choice. *Games and Economic Behavior*, 109:156–183.

Berry, S., Gandhi, A., and Haile, P. (2013). Connected Substitutes and Invertibility of Demand. *Econometrica*, 81(5):2087–2111.

Berry, S. and Haile, P. (2020). Nonparametric Identification of Differentiated Products Demand Using Micro Data. *NBER Working Paper*, number 27704.

Bifulco, R. and Ladd, H. F. (2007). School Choice, Racial Segregation, and Test-Score Gaps: Evidence From North Carolina's Charter School Program. *Journal of Policy Analysis and Management*, 26(1):31–56.

Bresnahan, T. F. and Reiss, P. C. (1991). Empirical Models of Discrete Games. *Journal of Econometrics*, 48(1-2):57–81.

Calsamiglia, C., Fu, C., and Güell, M. (2020). Structural Estimation of a Model of School Choices: The Boston Mechanism Versus Its Alternatives. *Journal of Political Economy*, 128(2):642–680.

Calsamiglia, C., Haeringer, G., and Klijn, F. (2010). Constrained School Choice: An Experimental Study. *The American Economic Review*, 100(4):1860–1874.

Calsamiglia, C., Martínez-Mora, F., and Miralles, A. (2021). School choice design, risk aversion and cardinal segregation. *The Economic Journal*, 131(635):1081–1104.

Che, Y.-K., Hahm, D. W., and He, Y. (2020). Leveraging uncertainties to infer preferences: Robust analysis of school choice. *Working Paper*.

Che, Y.-K. and Tercieux, O. (2019). Efficiency and Stability in Large Matching Markets. *Journal of Political Economy*, 127(5):2301–2342.

Chen, Y. and Sönmez, T. (2006). School Choice: An Experimental Study. *Journal of Economic Theory*, 127(1):202–231.

Conlon, C. T. and Mortimer, J. H. (2013). Demand Estimation Under Incomplete Product Availability. *American Economic Journal: Microeconomics*, 5(4):1–30.

Corcoran, S. P., Jennings, J. L., Cohodes, S. R., and Sattin-Bajaj, C. (2018). Leveling the Playing Field for High School Choice: Results From a Field Experiment of Informational Interventions. *NBER Working Paper*, number 24471.

Dynarski, S., Libassi, C., Michelmore, K., and Owen, S. (2021). Closing the gap: The effect of reducing complexity and uncertainty in college pricing on the choices of low-income students. *American Economic Review*, 111(6):1721–56.

Epple, D. and Romano, R. E. (1998). Competition Between Private and Public Schools, Vouchers, and Peer-Group Effects. *American Economic Review*, pages 33–62.

Ergin, H. I. (2002). Efficient Resource Allocation on the Basis of Priorities. *Econometrica*, 70(6):2489–2497.

Fabre, A., Larroucau, T., Martınez, M., Neilson, C., and Rios, I. (2021). Application mistakes and information frictions in college admissions. *Working Paper*.

Fack, G., Grenet, J., and He, Y. (2019). Beyond Truth-Telling: Preference Estimation With Centralized School Choice and College Admissions. *American Economic Review*, 109(4):1486–1529.

Feng, J., Bhargava, H. K., and Pennock, D. M. (2007). Implementing sponsored search in web search engines: Computational evaluation of alternative mechanisms. *INFORMS Journal on Computing*, 19(1):137–148.

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

Gaynor, M., Propper, C., and Seiler, S. (2016). Free to Choose? Reform, Choice, and Consideration Sets in the English National Health Service. *American Economic Review*, 106(11):3521–57.

Goeree, M. S. (2008). Limited Information and Advertising in the U.S. Personal Computer Industry. *Econometrica*, 76(5):1017–1074.

Haeringer, G. and Klijn, F. (2009). Constrained School Choice. *Journal of Economic Theory*, 144(5):1921–1947.

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

Hassidim, A., Marciano, D., Romm, A., and Shorrer, R. I. (2017). The Mechanism Is Truthful, Why Aren't You? *American Economic Review*, 107(5):220–224.

Hassidim, A., Romm, A., and Shorrer, R. I. (2021). The limits of incentives in economic matching procedures. *Management Science*, 67(2):951–963.

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.

He, Y. (2017). Gaming the Boston School Choice Mechanism in Beijing. *Working Paper*.

Hom, M. O. (2018). School Choice, Segregation and Access to Quality Schools: Evidence From Arizona. Technical report, Yale University.

Hortaçsu, A., Madanizadeh, S. A., and Puller, S. L. (2017). Power to Choose? An Analysis of Consumer Inertia in the Residential Electricity Market. *American Economic Journal: Economic Policy*, 9(4):192–226.

Hoxby, C. and Turner, S. (2013). Expanding college opportunities for high-achieving, low income students. *Stanford Institute for Economic Policy Research Discussion Paper*, 12:014.

Hsieh, C.-T. and Urquiola, M. (2006). The Effects of Generalized School Choice on Achievement and Stratification: Evidence From Chile's Voucher Program. *Journal of Public Economics*, 90(8-9):1477–1503.

Hwang, S. I. M. (2017). How Does Heterogeneity in Beliefs Affect Students in the Boston Mechanism. *Working paper*.

Idoux, C. (2023). Integrating new york city schools: The role of admission criteria and family preferences. *Working Paper*.

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

Kessel, D. and Olme, E. (2018). School Choice Priority Structures and School Segregation. *Working paper*.

Koulayev, S. (2014). Search for differentiated products: identification and estimation. *The RAND Journal of Economics*, 45(3):553–575.

Larroucau, T. and Rios, I. (2020). Do short-list students report truthfully? strategic behavior in the chilean college admissions problem. *Working Paper*.

Laverde, M. (2020). Unequal assignments to public schools and the limits of school choice. *Working paper*.

Lewbel, A. (2000). Semiparametric Qualitative Response Model Estimation With Unknown Heteroscedasticity or Instrumental Variables. *Journal of Econometrics*, 97(1):145–177.

Li, S. (2017). Obviously Strategy-Proof Mechanisms. *American Economic Review*, 107(11):3257–3287.

Luflade, M. (2018). The value of information in centralized school choice systems. *Working Paper*.

Martin, G. J. and Yurukoglu, A. (2017). Bias in cable news: Persuasion and polarization. *American Economic Review*, 107(9):2565–99.

McFadden, D. (1989). A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration. *Econometrica*, 57(5):995.

Narita, Y. (2016). Match or mismatch: Learning and inertia in school choice. *Working Paper*.

Neilson, C. (2013). Targeted vouchers, competition among schools, and the academic achievement of poor students. *Working Paper*.

NYC DOE (2020). https://www.schools.nyc.gov/about-us/vision-and-mission.

Oosterbeek, H., Sóvágó, S., and Van Der Klaauw, B. (2021). Preference heterogeneity and school segregation. *Journal of Public Economics*, 197:104400.

Park, M. and Hahm, D. W. (2023). Location choice, commuting, and school choice. *Working Paper*.

Pathak, P. A. and Shi, P. (2020). How Well Do Structural Demand Models Work? Counterfactual Predictions in School Choice. *Journal of Econometrics*.

Pathak, P. A. and Sönmez, T. (2008). Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism. *American Economic Review*, 98(4):1636–1652.

Rees-Jones, A. (2018). Suboptimal Behavior in Strategy-Proof Mechanisms: Evidence From the Residency Match. *Games and Economic Behavior*, 108:317–330.

Roth, A. E. (1985). The college admissions problem is not equivalent to the marriage problem. *Journal of Economic Theory*, 36(2):277–288.

Roth, A. E. and Sotomayor, M. (1992). Two-Sided Matching. *Handbook of Game Theory with Economic Applications*, 1:485–541.

Sartain, L. and Barrow, L. (2022). The Pathway to Enrolling in a High-Performance High School: Understanding Barriers to Access. *Education Finance and Policy*, 17(3):379–407.

Sattin-Bajaj, C. (2016). *Unaccompanied Minors: Immigrant Youth, School Choice, and the Pursuit of Equity*. Harvard Education Press. Publication Title: Harvard Education Press.

Sattin-Bajaj, C., Jennings, J. L., Corcoran, S. P., Baker-Smith, E. C., and Hailey, C. (2018). Surviving at the Street Level: How Counselors' Implementation of School Choice Policy Shapes Students' High School Destinations. *Sociology of Education*, 91(1):46–71.

Shapley, L. and Scarf, H. (1974). On Cores and Indivisibility. *Journal of Mathematical Economics*, 1(1):23–37.

Shorrer, R. I. and Sóvágó, S. (2022). Dominated choices in a strategically simple college admissions environment. *Working Paper, Accepted for Journal of Political Economy Microeconomics*.

Swait, J. and Ben-Akiva, M. (1987). Incorporating Random Constraints in Discrete Models of Choice Set Generation. *Transportation Research Part B: Methodological*, 21(2):91–102.

Thompson, T. S. (1989). Identification of semiparametric discrete choice models. *Discussion Paper, Center for Economic Research, Department of Economics, University of Minnesota*, 249.

Ursu, R. M. (2018). The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science*, 37(4):530–552.

Online Appendix to
Distributional Impacts of Centralized School Choice

Jaewon Lee   Suk Joon Son

October 9, 2023

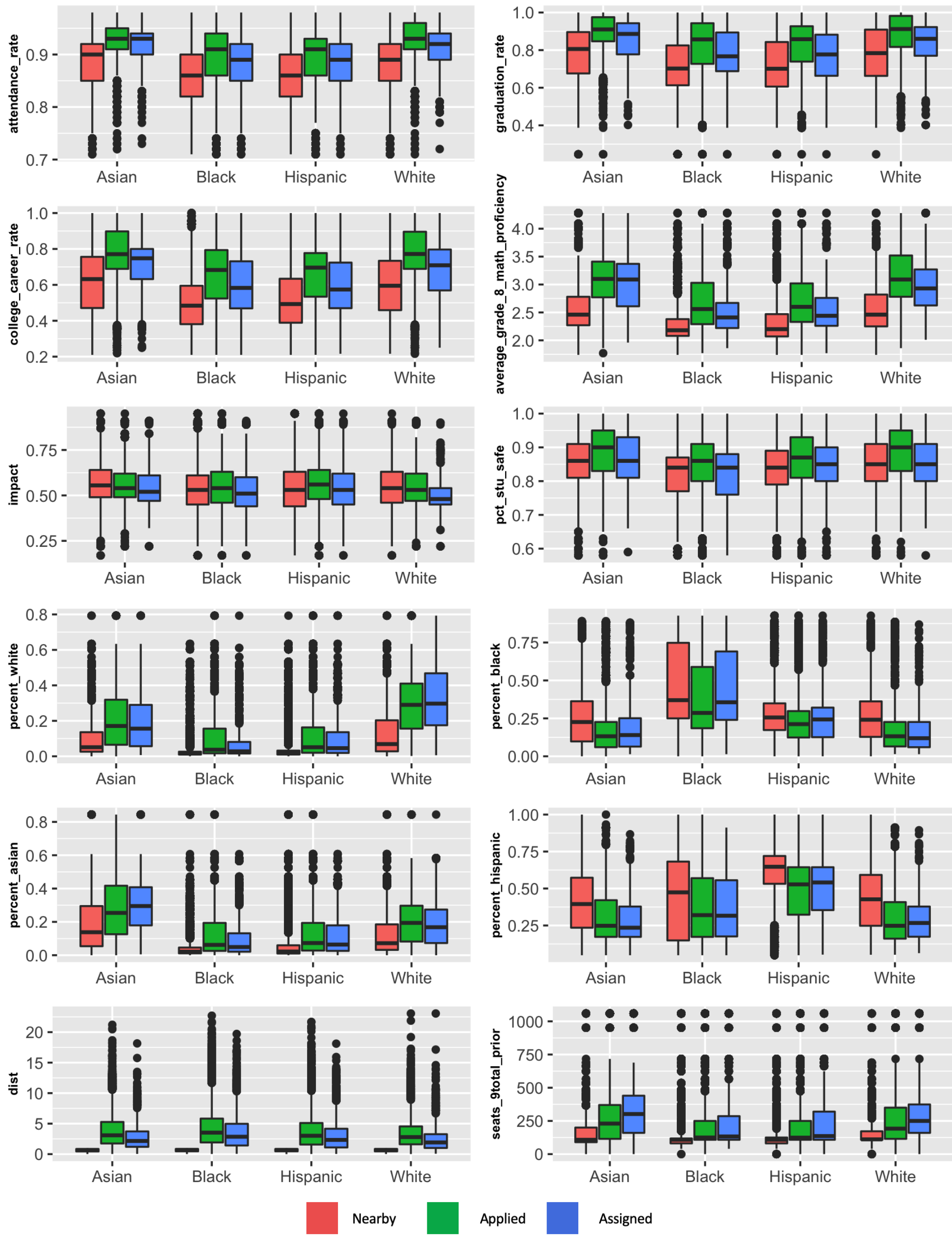
B Additional Tables and Figures

Table B.1: Regression of Page Rank on School Characteristics

	Dependent variable:
	Page rank
Constant	52.965 (50.340)
Average grade 8 math proficiency (std.)	−5.129 (3.738)
Graduation rate	41.384* (23.927)
Attendance rate	−54.318 (60.453)
College/career rate	8.453 (21.425)
Percent of students who feel safe	16.884 (29.427)
9th grade seats	−0.007 (0.015)
Percent Asian	−3.557 (19.134)
Percent Black	1.555 (8.762)
Percent White	−19.579 (18.452)
Observations	352
R^2	0.037
F Statistic	1.456 (df = 9; 342)

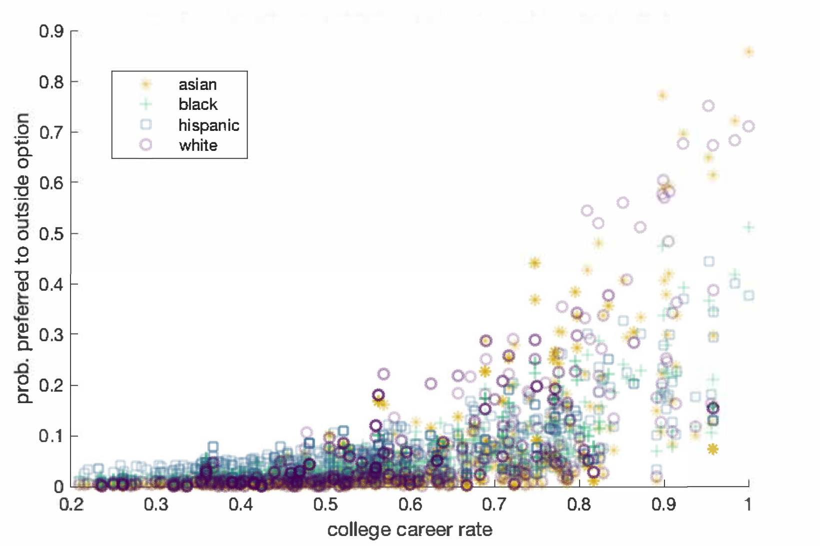
Notes: *p<0.1; **p<0.05; ***p<0.01. Standard errors in parentheses. Standardized values are indicated by (std.). *College/career rate* indicates the proportion of students who graduated from high school four years after they entered 9th grade and then enrolled in college, a vocational program, or a public service program within six months of graduation. Each school has equal weight regardless of class size. The sample excludes the nine specialized high schools and schools with missing data.

Figure B.1: Schools Nearby, Applied to, and Matched, by Ethnicity

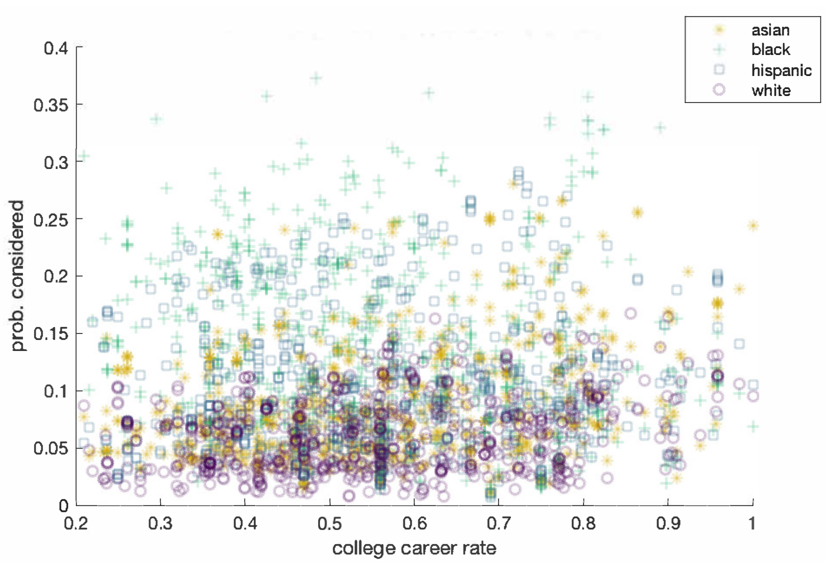


Notes: Nearby schools are the schools within one mile from student's home. The applied and assigned schools are from the main round of applications. *Pct_stu_safe* denotes the proportion of students who have reported that they feel safe in the school. *College_career_rate* indicates the proportion of students who graduated from high school four years after they entered 9th grade and then enrolled in college, a vocational program, or a public service program within six months of graduation.

Figure B.2: Preference and Consideration and School Performance



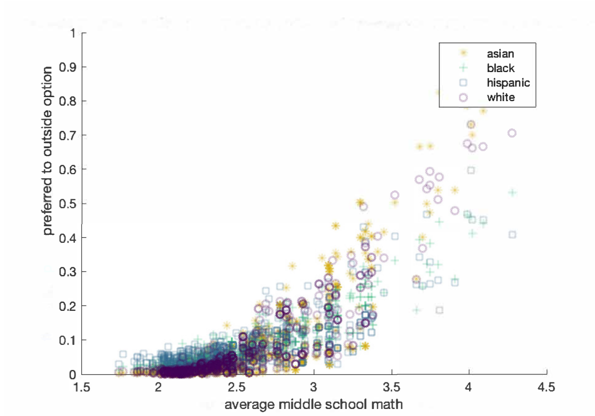
(a) Probability of Being Preferred to the Outside Option



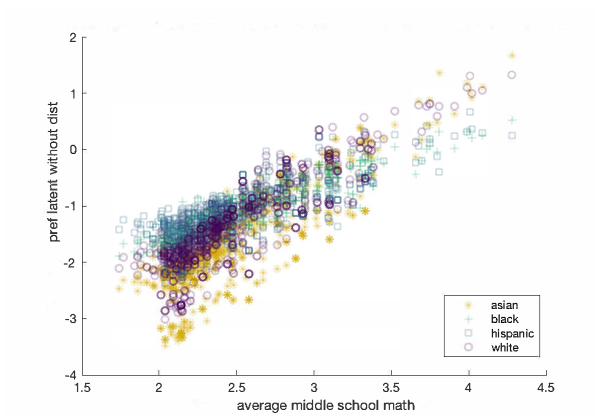
(b) Probability of Being Considered

Notes: For each ethnicity, each point in the scatter plot denotes a school. Each line represents a cubic polynomial fit. *College/career rate* indicates the school's proportion of students who graduated from high school four years after they entered 9th grade and then enrolled in college, a vocational program, or a public service program within six months of graduation.

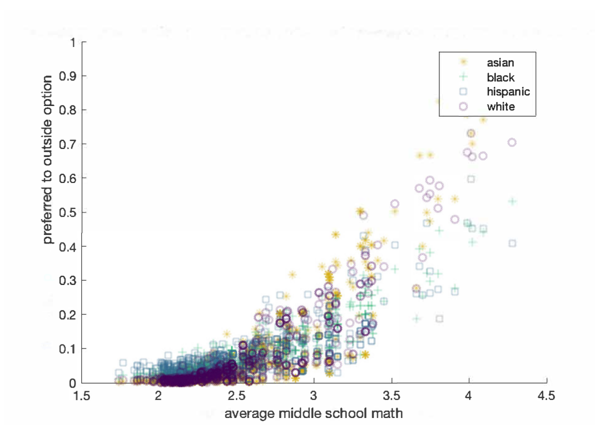
Figure B.3: Probability and Latent Values for Preference and Consideration



(a) Probability of Being Preferred to the Outside Option



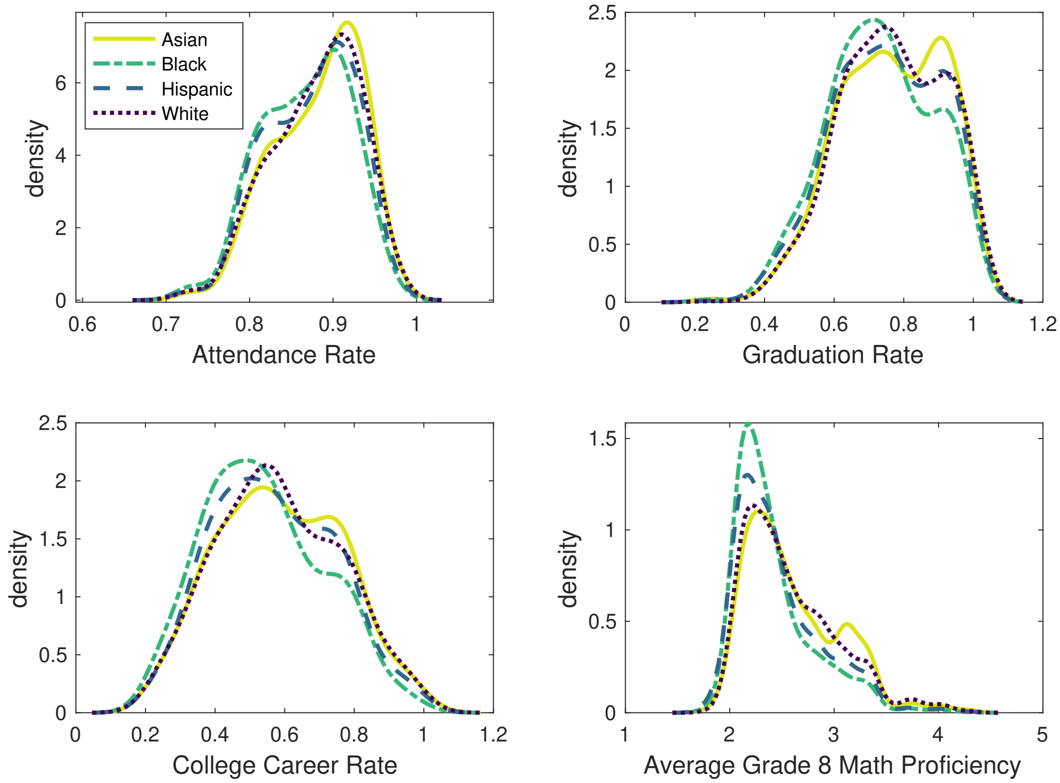
(b) Mean Latent Values for Preference



(c) Mean Latent Values for Preference (Distance = 0)

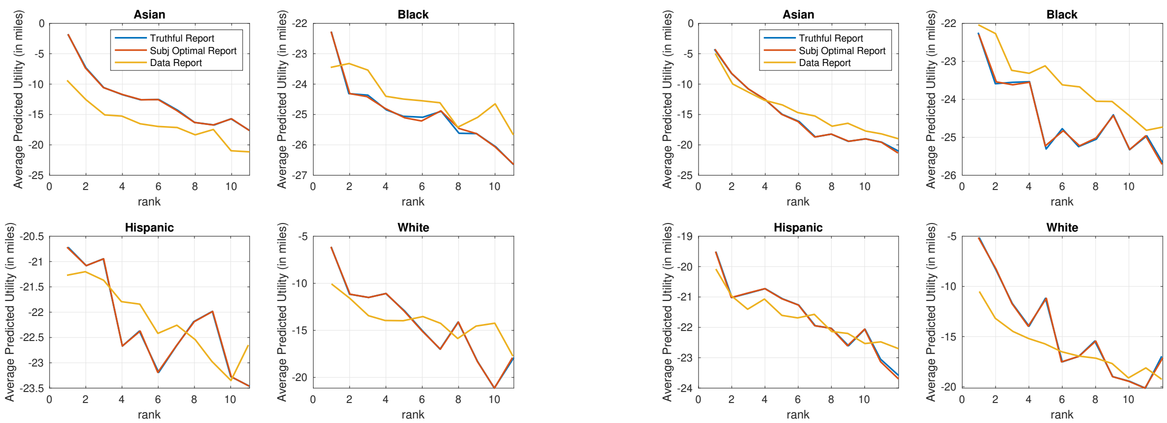
Notes: For each ethnicity, each point in the scatter plot denotes a school.

Figure B.4: Characteristics of Considered Schools by Ethnicity



Notes: College/career rate indicates the proportion of students who graduated from high school four years after they entered 9th grade and then enrolled in college, a vocational program, or a public service program within six months of graduation.

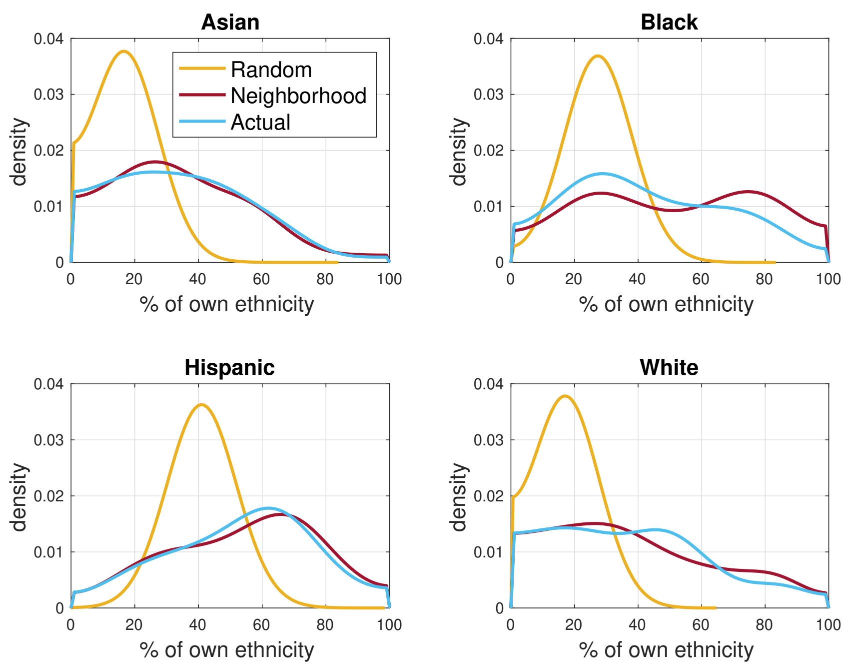
Figure B.5: Slope of Predicted Utilities against Rank in Report



(a) Among 11-program reports (b) Among 12-program (full) reports

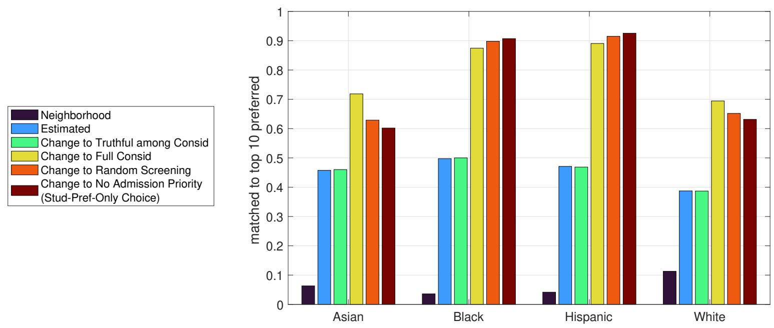
Notes: Average predicted utility at each rank is the within-race average of predicted utilities (i.e., net of ϵ_{ij}^{v}) normalized by the coefficient on the race-specific coefficient on distance.

Figure B.6: Percent of Own Ethnicity by Matching, Model-Free



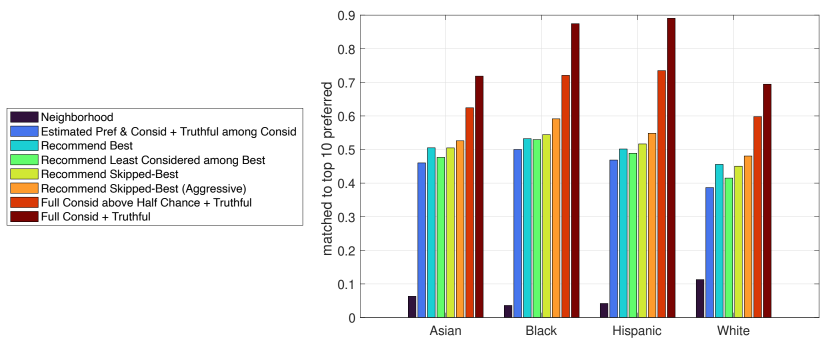
Notes: For each ethnicity and matching, the plot represents the kernel-smoothed density of the proportion of students with the same ethnicity in the students' assigned programs. The kernel density estimation uses Gaussian kernel with bandwidth 10, and is boundary corrected. See Table 9 for the definitions of the matchings.

Figure B.7: Proportion Matched to Top Ten Preferred Programs—Decomposition

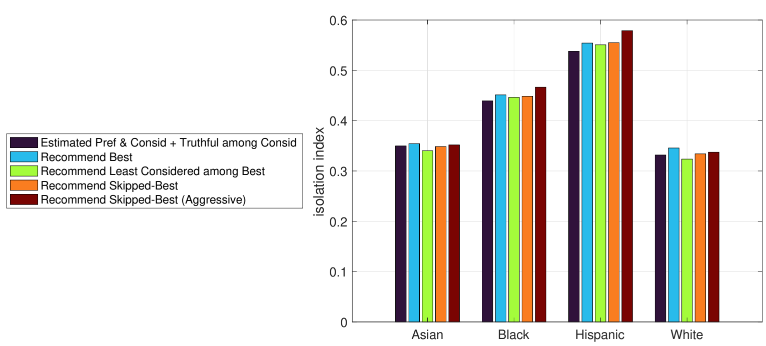


Notes: Each bar represents the fraction of the students matched to their top ten preferred programs. The sample includes both the programs that are considered and those that are not. See Table 9 and the discussions for the definitions of the matchings.

Figure B.8: Impacts of Information Interventions—Top Ten Preferred and Isolation Indices



(a) Proportion Matched to Top Ten Preferred



(b) Isolation Indices

Notes: Each bar in (a) represents the fraction of the students matched to their top ten preferred programs. Each bar in (b) represents isolation index of an ethnic group in a matching.

C Deferred Acceptance Mechanism in NYC

In the 2016–2017 school year, the DOE ran two rounds of DA assignments for the traditional (non-specialized) high schools and one round of DA assignment for the nine specialized high schools. In our analysis, we focus on the first (main) round for the non-specialized high schools. This is the "main" round in the sense that approximately 85% of the final matches coincide with the match in this round.

Using the students' submitted rankings over the school programs and the programs' rankings over the students, the DA algorithm (Gale and Shapley, 1962; Abdulkadiroğlu and Sönmez, 2003) matches the students to the school programs according to the following procedure.

- *Step* 1*:* Each applicant proposes to his first-ranked school program, if any. Each school program sorts the proposers according to its rankings and tentatively accepts all the highest-ranking proposers up to its capacity. It rejects any other proposers.

  Then, for each $k \geq 2$,

- *Step k:* Each applicant who was not tentatively accepted by any program in Step $(k-1)$ proposes to his highest-ranked school program that has not previously rejected him, if any. Each school program sorts the new proposers and the applicants tentatively accepted previously according to its rankings and tentatively accepts all the highest-ranking applicants up to its capacity. All the other proposers are rejected.

The algorithm stops when there are no proposing students. Each student is assigned his final tentative assignment. In NYC high school match, the school programs have separate seats (capacities) for students with and without disabilities. Therefore, DA algorithms are run separately for the two student groups defined by their disabilities type.

D Identification: Details

D.1 Supplementary Results on Nonparametric Identification

Proposition D.1 (Identification of preferences and consideration with ideal data)**.** *Suppose that we observe $z_i \equiv (z_i^v, z_i^c, z_i^-)$ where (z_i^v, z_i^c) is a special regressor for (v_i, c_i) with large support conditional on z_i^-. Then,*

(i) if $L = J$, $\mathbb{P}\big(v_i \leq v, c_i^* \leq c^* | z_i = z\big)$ is identified for every $(v, c^*, z) \in \operatorname{supp}(v_i, c_i^*, z_i)$.

(ii) if $L < J$, $\mathbb{P}\big(c_i^* \leq c^* | z_i = z\big)$ is identified for every $(c^*, z) \in \operatorname{supp}(c_i^*, z_i)$ and $\mathbb{P}\big(v_i \leq v | z_i = z\big)$ is identified for every $(v, z) \in \operatorname{supp}(v_i, z_i)$.[69]

[69] In the case of $L < J$, a stronger result as in the case for $L = J$ is available following a proof similar to Lemma 1 of Agarwal and Somaini (2022).

Proposition D.2 (Identification of preferences with surely considered sets)**.** *Suppose that we observe a special regressor for* v_i*, named* z_i^v*, with a large support conditional on* z_i^-*. Suppose also that* $\mathcal{S}_i \equiv \mathcal{S}(z_i)$ *is constant with respect to* z_i^v*. Then,*

(i) if $L = J$*,* $\mathbb{P}\big((v_{ij})_{j\in\mathcal{S}(z_i)} \le v|z_i\big)$ *is identified for all* (v, z_i) *in its support.*

(ii) if $L < J$*,* $\mathbb{P}\big((v_{ij})_{j\in\mathcal{A}} \le x|z\big)$ *is bounded within an interval of width* $\mathbb{P}(|r_i| = L, r_i \cap \mathcal{A} = \emptyset|z_i^v = x, z^-)$ *for all* $(x, z, \mathcal{A})$ *such that* $\mathcal{A} \subseteq \mathcal{S}(z)$ *with* $|\mathcal{A}| \le L$*.*

D.2 Lemmas

These lemmas are used in the proofs of the observations and the propositions. We define a school to be *acceptable* if $v_{ij} > 0$ and *unacceptable* if $v_{ij} < 0$.

Lemma D.1. *Consider a list* r *that contains an unacceptable school before an acceptable school, and the lowest-ranked school is an acceptable school. Then, in any realization,* r *gives a weakly less payoff than an alternative list that switches the lowest-ranked school unacceptable school with the school that gives the maximum utility among the schools that follow this lowest-ranked unacceptable school.*[70]

Lemma D.2 (Never write an unacceptable school)**.** *For any list* r *that contains a considered but unacceptable school, there is an alternative list that contains no unacceptable school and gives strictly higher expected utility.*

D.3 Proofs

Proof of Lemma D.1. By assumption, the list r has an unacceptable school before an acceptable school. Let j_- denote the lowest-ranked unacceptable school in the list. Then, by construction, (1) there are some schools that follow j_- and (2) these schools are all acceptable. Let the utility-maximum of these school be indicated by $j_{\max}$ (and there is always such a school). Then, the report r reads:

$$r = (\underbrace{\cdots}_{A}, j_-, \underbrace{\cdots}_{B}, j_{\max}, \underbrace{\cdots}_{C})$$

where A, B, and C denote the set of the schools in each respective position. Each of A, B, and C may or may not be empty.

Consider an alternative list r' that switches $j_{\max}$ with j_-, as in the statement:

$$r' = (\underbrace{\cdots}_{A}, j_{\max}, \underbrace{\cdots}_{B}, j_-, \underbrace{\cdots}_{C})$$

where the schools and the ordering within each A, B, and C is unaltered.

[70]The lemma is similar to what appears in the proof of Proposition 3 (ii) in He (2017).

Representing an outcome in the relevant probability space by ω, we want to show that r' weakly dominates r for every ω, i.e., $v_{i\mu(i;r)}(\omega) \leq v_{i\mu(i;r')}(\omega)$ for all ω, where $\mu(i;r)$ is the assignment of i in the case that i reports r. To see this, suppose not: there is ω such that $v_{i\mu(i;r)}(r;\omega) > v_{i\mu(i;r')}(r';\omega)$. Then, it must be that the student get rejected at all the A schools under this ω regardless of submitting r or r', i.e.,

$$\tilde{\pi}_j(\omega) < \widetilde{\mathsf{score}}_{ij}(r(j);\omega) \equiv \widetilde{\mathsf{score}}_{ij}\big(r'(j);\omega\big) \; \forall j \in A$$

where $r(j)$ and $r'(j)$ denote the ranks of school j in r and r', respectively. This is because otherwise, he gets into the same school regardless of reporting r or r' and obtains the same utility. Note that it is impossible that he gets rejected in one report but not in the other report—his scores for any $j \in A$ under the two reports are exactly the same in the two reports as the submitted rank of any $j \in A$ in the two reports are the same. This is because score is restricted to depend only depends on certain aspects of the report—i.e., the rank.

Also, it must be that he gets rejected by j_- under r. Otherwise, conditioning on that the student is reject by all schools in A, this is the worst that can happen to him under r or r' because B and C can never have an unacceptable school by construction. Therefore, there is no way that j_- will strictly beat allocation under r'. Also, it must be that he gets rejected by $j_{\max}$ under r'; otherwise, this is the best that can happen to him under r or r' and so there is no way that allocation under r will strictly beat $j_{\max}$. Thus,

$$\begin{aligned}
&\tilde{\pi}_{j_-}(\omega) < \widetilde{\mathsf{score}}_{ij_-}\big(r(j_-);\omega\big)\\
&\tilde{\pi}_{j_{\max}}(\omega) < \widetilde{\mathsf{score}}_{ij_{\max}}\big(r'(j_{\max});\omega\big)
\end{aligned}$$

Similarly, it must be that he fails to make the cutoffs (in either reports) by all schools in B. Otherwise, he gets same utility under the two reports. Note that he makes the cutoff in any of these schools in B by submitting r iff he does so in r'; the score for the school is the same under the two reports.

Further, it must be that he is rejected by $j_{\max}$ under r and j_- under r'. This follows from the assumption that perceived scores are monotonic in the submitted rank and the second step:

$$\begin{aligned}
&\tilde{\pi}_{j_-}(\omega) < \widetilde{\mathsf{score}}_{ij_-}\big(r(j_-);\omega\big) \leq \widetilde{\mathsf{score}}_{ij_-}\big(r'(j_-);\omega\big)\\
&\tilde{\pi}_{j_{\max}}(\omega) < \widetilde{\mathsf{score}}_{ij_{\max}}\big(r'(j_{\max});\omega\big) \leq \widetilde{\mathsf{score}}_{ij_{\max}}\big(r(j_{\max});\omega\big)
\end{aligned}$$

By the same reasoning, it must be that he fails to make the cutoffs (in either reports) by all schools in C. Otherwise, he gets same utility under the two reports. Note that he makes the cutoffs in all of these schools in B by submitting r iff he does so in r'; the scores are the same under the two reports.

Then, they get rejected by all schools in either of the two reports, and is placed into outside option, in which they derive the same utility. This contradicts $v_{i\mu(i;r)}(r;\omega) > v_{i\mu(i;r')}(r';\omega)$ we started with. □

Proof of Lemma D.2. We first show that, for any r that contains an unacceptable school, there is an alternative list without any unacceptable school that gives weakly higher expected utility.

Suppose that r has an unacceptable school at the very end. Then, it is straightforward to verify that dropping this school weakly increases expected utility. Repeat this process until the last school is an acceptable school. If the list is now composed of only the acceptable schools (or is empty), then such a list is an alternative list that we wanted to find.

If there are still some unacceptable schools in the list, then Lemma D.1 can be applied as there is some acceptable school after any unacceptable school. We further know that the new report found by the lemma must give weakly higher expected utility, as we've claimed that for any outcome ω, the new report must give utility weakly higher than the old report.

Apply the lemma to switch the lowest-ranked unacceptable school to a lower spot in the list. If this schools is now in the last spot, then drop this. If not, the schools that appear after this lowest-ranked unacceptable school are all acceptable, so that we can apply the lemma again. Continue to apply this lemma, this unacceptable the school gets moved to the last spot, in which case we can drop the unacceptable school and obtain even (weakly) higher expected utility.

If the resulting report is now filled with only acceptable schools (or is empty), we have found an alternative list that we wanted to find. If not, repeat the aforementioned process of moving the lowest-ranked unacceptable school down the list and then dropping it, until there is no unacceptable schools in the list. Every such process gives weakly higher expected utility, and therefore the resulting list gives weakly higher expected utility.

Note that the process above now has at least one occasion where an unacceptable but considered school is dropped from the last slot. By assumption, a student believes he has positive chance of matching to a considered school upon listing. Therefore, this drop strictly increases his expected utility. □

Proof of Observation 1. Let L denote the maximum allowed length of the list. We show that the first statement holds.

To show that $j \in r_i$ implies both $j \in \mathcal{C}_i$ and $v_{ij} > 0$, we show the contrapositive. First, if $j \notin \mathcal{C}_i$, j cannot be on r_i by definition of consideration. Second, suppose that $v_{ij} < 0$ and $j \in \mathcal{C}_i$. By Lemma D.2, such a list with an unacceptable but considered school cannot be (subjectively) optimal.

Suppose now that $v_{ij} > 0$ and $j \in \mathcal{C}_i$, but $j \notin r_i$. Then one can strictly gain by adding j on the bottom of the list, which contradicts subjective optimality of r_i. The strict relation comes from $j \in \mathcal{C}_i$; a considered school has (subjectively) positive admission chance upon listing. Addition of a school is possible since r_i has not exhausted all the available slots.

We now show that the second statement holds. The second statement is equivalent to the following statement: r_i has exactly L schools if and only if $\{j \in \mathcal{J} | v_{ij} > 0, j \in \mathcal{C}_i\}$ has L schools or more.

Suppose first that $|r_i| = L$ but $|\{j \in \mathcal{J} | v_{ij} > 0, j \in \mathcal{C}_i\}| < L$. Because all schools in r_i must be considered by definition, there must be some schools in r_i that is subjectively reachable but is unacceptable. By Lemma D.2, such a list cannot be subjectively optimal.

Suppose now that $|\{j \in \mathcal{J} | v_{ij} > 0, j \in \mathcal{C}_i\}| \geq L$ but $|r_i| < L$. Then, there must be some school $j \notin r_i$ such that $v_{ij} > 0$ and $j \in \mathcal{C}_i$. Adding j at the bottom of the list gives strictly higher payoff, contradicting that r_i is subjectively optimal. □

Proof of Proposition A.1. I implicitly condition everything on $z_i \backslash (z_i^v, a_i)$. Take any $z^v \in \text{supp}(z_i^v)$ and the according $\bar{a} \equiv (\bar{a}_1(z^v), \cdots, \bar{a}_J(z^v))$. Note that $\mathbb{P}(c_i > 0 | \bar{a}) = 1$ implies $\mathbb{P}(c_i > 0 | z_i^v, \bar{a}) = 1$ almost surely. Then, almost surely,

$$
\begin{aligned}
&\mathbb{P}(j \notin r_i \ \forall j = 1, \cdots, J | z_i^v = z^v, a_i = \bar{a}) \\
&= \mathbb{P}(c_{ij} < 0 \text{ or } v_{ij} < 0 \ \forall j = 1, \cdots, J | z^v, \bar{a}) && \text{by proof of Observation 1 with generalized } L \\
&= \mathbb{P}(v_{ij} < 0 \ \forall j = 1, \cdots, J | z^v, \bar{a}) && \text{by } \mathbb{P}(c_i > 0 | z^v, \bar{a}) = 1 \\
&= \mathbb{P}(v_i < 0 | z^v, \bar{a}) \\
&= \mathbb{P}(v_i < 0 | z^v) && \text{by } v_i \perp\!\!\!\perp a_i | z_i^v \\
&= \mathbb{P}(\tilde{v}_i < z^v). && \text{by } \tilde{v}_i \perp\!\!\!\perp z_i^v
\end{aligned}
$$

As the first line is observed, the last line is identified almost surely for $z^v \in \mathbb{R}^J$ by the large support assumption on z_i^v. Then, by the independence assumptions on a_i and z_i^v, $\mathbb{P}(v_i > x | z^v, a) = \mathbb{P}(v_i > x | z^v) = \mathbb{P}(\tilde{v}_i > x + z^v)$. Therefore, $\mathbb{P}(v_i > x | z^v, a) = \mathbb{P}(v_i > x | z)$ is identified for almost every $(x, z) \in \text{supp}(v_i, z_i)$. □

Proof of Proposition A.2. I will implicitly condition everything on $z_i \backslash z_i^c$. I first prove (i). Take any $z^c \in \text{supp}(z_i^c)$. Note that

$$
\begin{aligned}
&\mathbb{P}(j \in r_i \ \forall j = 1, \cdots, J | z_i^c = z^c) \\
&= \mathbb{P}(c_i > 0, v_i > 0 | z^c) && \text{by proof of Observation 1 with generalized } L \\
&= \mathbb{P}(c_i > 0 | z^c) \mathbb{P}(v_i > 0 | z^c) && \text{by } c_i \perp\!\!\!\perp v_i | z_i^c \\
&= \mathbb{P}(\tilde{c}_i > z^c) \mathbb{P}(v_i > 0 | z^c) && \text{by } \tilde{c}_i \perp\!\!\!\perp z_i^c
\end{aligned}
$$

but the first line is observed and $\mathbb{P}(v_i > 0 | z^c)$ is known on almost all $z^c \in \text{supp}(z_i^c)$ by assumption. Thus, $\mathbb{P}(\tilde{c}_i > z^c)$ is identified almost surely. By the assumptions on z_i^c, $\mathbb{P}(c_i > x | z^c) = \mathbb{P}(\tilde{c}_i > x + z^c)$ and thus $\mathbb{P}(c_i > x | z^c)$ is identified for almost all $(x, z^c) \in \text{supp}(c_i, z_i^c)$. The result follows from the definition of c_i^*, i.e. $c_{ij}^* = \mathbb{1}(c_{ij} > 0)$ for all (i, j).

The proof of (ii) follows analogously by noting that $\mathbb{P}(j \in r_i \ \forall j \in \mathcal{A} | z_i^c = z^c) = \mathbb{P}((\tilde{c}_{ij})_{j \in \mathcal{A}} > (z_j)_{j \in \mathcal{A}}^c) \mathbb{P}((v_{ij})_{j \in \mathcal{A}} > 0 | z^c)$ and that $\mathbb{P}(j \in r_i \ \forall j \in \mathcal{A} | z_i^c = z^c)$ is observed while $\mathbb{P}((v_{ij})_{j \in \mathcal{A}} > 0 | z^c)$ is assumed identified. □

Proof of Proposition A.3. Define $v_{ij}^* = v_{ij}\big(2 \cdot \mathbb{1}(v_{ij} > 0, c_{ij}^* = 1) - 1\big)$. Note first that the assumptions imply the distribution of $v_i^* \equiv (v_{ij})_{j \in \mathcal{J}}$ is known. Note also that $\arg\max_{r \in \mathcal{R}(\mathcal{C}_i)} v \cdot p^r = \arg\max_{r \in \mathcal{R}(\mathcal{J})} v^* \cdot p^r$. Therefore, two beliefs $p \equiv \{p_j^r\}_{j \in \mathcal{J}, r \in \mathcal{R}(\mathcal{J})}$ and $p' \equiv \{p_j^{\prime r}\}_{j \in \mathcal{J}, r \in \mathcal{R}(\mathcal{J})}$ are behaviorally equivalent if and only if for all $v \in \mathcal{R}^J$, $\arg\max_{r \in \mathcal{R}(\mathcal{J})} v \cdot p^r = \arg\max_{r \in \mathcal{R}(\mathcal{J})} v \cdot p^{\prime r}$.

Let $C^r(p) \equiv \{v \in \mathbb{R}^J | r = \arg\max_{r\in\mathcal{R}(\mathcal{J})} v \cdot p^r\}$ for each $r \in \mathcal{R}(\mathcal{J})$. Then, two beliefs p and p' are behaviorally equivalent if and only if $C^r(p) = C^r(p')$ for all $r \in \mathcal{R}(\mathcal{J})$.

`Proof under assumption (1):` $L = J = 2$.

Implicitly condition on everything on z_i. From Observation 1, it is straightforward to verify that $\left(C^r(p)\right)_{r\in\mathcal{R}(\mathcal{J})}$ is pinned down by a single number $\delta \equiv \frac{p_1^{(1)}-p_1^{(2,1)}}{p_2^{(2)}-p_2^{(1,2)}}$.This can be checked by noting that $C^{\emptyset}(p) = \{(v_1, v_2) \in \mathbb{R}^2 | v_1, v_2 \leq 0\}$, $C^{(1)}(p) = \{(v_1, v_2) \in \mathbb{R}^2 | v_1 \geq 0, v_2 \leq 0\}$, $C^{(2)}(p) = \{(v_1, v_2) \in \mathbb{R}^2 | v_1 \leq 0, v_2 \geq 0\}$, $C^{(1,2)}(p) = \{(v_1, v_2) \geq 0 | v_2/v_1 \leq \delta, \}$, and $C^{(2,1)}(p) = \{(v_1, v_2) \geq 0 | v_2/v_1 \geq \delta\}$. By assumption, everyone (in the subgroup defined by the observables) shares the common belief $p = \{p_j^r\}_{j\in\mathcal{J},r\in\mathcal{R}(\mathcal{J})}$ and therefore $\mathbb{P}(\{v_{i2}/v_{i1} \geq \delta\} \cap \{v_i \geq 0\}) = \mathbb{P}(v_i \in C^{(2,1)}(p)) = \mathbb{P}(r_i = (2,1))$. As $\mathbb{P}(v_i \leq v)$ is known, the left-hand side of the equation is calculable as a function of δ. On the other hand, the right-hand side is observable. Thus, belief is identified.

`Proof under assumption (2):` $L = 1$.

By assumption, everyone has the same belief, which I denote by p. Note that $C^{(j)}(p) = \{v \in \mathbb{R}^J | j = \arg\max_{k\in 0,1,\ldots,J} p_k^{(k)} v_k\} = \{v \in \mathbb{R}^J | j = \arg\max_{k\in 0,1,\ldots,J} \frac{p_k^{(k)}}{p_1^{(1)}} v_k\}$ for $j = 1, \ldots, J$ and $C^{\emptyset}(p) = \{(v_1, v_2) \in \mathbb{R}^2 | v_1, v_2 \leq 0\}$. Thus, the $C_r(p)'s$ are completely characterized by the vector $\tilde{p} \equiv (\tilde{p}_2, \cdots, \tilde{p}_J) \equiv (\frac{p_2}{p_1}, \ldots, \frac{p_J}{p_1})$. Therefore, belief is identified if $\tilde{p}$ is identified.

I now claim that one can use Corollary 1 of Berry et al. (2013), denoted BGH. In their notation, $x = \tilde{p}$, $\mathcal{X}^* = \mathcal{X} = \mathbb{R}_{++}^{J-1}$, and $\sigma(\tilde{p}) = (\sigma_2(\tilde{p}), \cdots, \sigma_J(\tilde{p})) : \mathcal{X} \subseteq \mathbb{R}^{J-1} \to \mathbb{R}^{J-1}$ where $\sigma_j(\tilde{p}) = \frac{\mathbb{P}(v_i \in C^{(j)}(\tilde{p}))}{\sum_{k=1}^J \mathbb{P}(v_i \in C^{(k)}(\tilde{p}))}$ for $j = 1, \ldots, J$. Note that the school $j = 1$ now plays the role of BGH's "outside option" (which is denoted $j = 0$ in their notation).[71] To see that the corollary applies, note first that $\mathcal{X}$ is a Cartesian product. Moreover, $\sigma_j(\tilde{p})$ is strictly decreasing in $\tilde{p}_k$ for all $j = \{1, \ldots, J\}$ and for all $k \neq 1, j$, as (1) $\sum_{k=1}^J \mathbb{P}(v_i \in C^{(k)}(\tilde{p}))$ is constant over $\tilde{p}$, and (2) $\mathbb{P}(v_i \in C^{(j)}(\tilde{p}))$ is strictly decreasing because v_i has full support. Thus, BGH's Corollary 1 applies and $\sigma(\tilde{p})$ is injective. □

Proof of Proposition D.1 . I first prove case (i). Take any $z \equiv (z^v, z^c, z^-)$ such that $z^v \in \mathbb{R}^J$, $z^c \in \mathbb{R}^J$, and $z^- \in \text{supp}(z_i^-)$. Then,

$$\begin{aligned}
&\mathbb{P}\big(j \in r_i \ \forall j = 1, \cdots, J | z_i^v = z^v, z_i^c = z^c, z_i^- = z^-\big) \\
&= \mathbb{P}(\tilde{v}_i - z^v > 0, \tilde{c}_i - z^c > 0 | z^v, z^c, z^-) \\
&= \mathbb{P}(\tilde{v}_i > z^v, \tilde{c}_i > z^c | z^-) \\
&= \mathbb{P}(-\tilde{v}_i < -z^v, -\tilde{c}_i < -z^c | z^-)
\end{aligned}$$

and since the first expression is observed for any $z^v \in \mathbb{R}^J$, $z^c \in \mathbb{R}^J$, and $z^- \in \text{supp}\left(z_i^-\right)$, the last expression is identified for any such (z^v, z^c, z^-). Thus, the joint distribution of $(-\tilde{v}_i, -\tilde{c}_i)$ conditional on z_i^-, and therefore the joint distribution of $(\tilde{v}_i, \tilde{c}_i)$ conditional on z_i^-, is identified on the support of z_i^-. As $v_i = \tilde{v}_i - z_i^v$ and $c_i = \tilde{c}_i - z_i^c$ with $(\tilde{v}_i, \tilde{c}_i) \perp\!\!\!\perp (z_i^v, z_i^c) | z_i^-$ and $z_i \equiv (z_i^v, z_i^c, z_i^-)$ is observed, the joint distribution of (v_i, c_i) conditional on z_i is identified for every z_i in its support.

[71] The outside option $j = 0$ as considered in my model is left out of the discussion here because their choice probability does not change according to p.

To show the first part of case (ii), note that

$$\begin{aligned}
&\mathbb{P}\big(r_i = \emptyset | z_i^v = z^v, z_i^c = z^c, z_i^- = z^-\big) \\
&= \mathbb{P}(v_{ij} \leq 0 \text{ or } c_{ij} \leq 0 \ \forall j \in \mathcal{J} | z^v, z^c, z^-) \\
&= \mathbb{P}(\tilde{v}_{ij} \leq z_{ij}^v \text{ or } \tilde{c}_{ij} < z_{ij}^c \ \forall j \in \mathcal{J} | z^-)
\end{aligned}$$

Now, send all of the elements in z^c to negative infinity. By the dominated convergence theorem, the last expression converges to $\mathbb{P}(\tilde{v}_{ij} \leq z_{ij}^v \ \forall j \in \mathcal{J} | z^-)$. Note that z_i^v is a special regressor for v_i with a large support. Use the special regressor similarly as before to identify the distribution of v_i. The second part of case (ii) follows similarly by sending all of the elements in z^v to negative infinity. □

Proof of Proposition D.2. Proof of part (i) follows by noting that

$$\begin{aligned}
&\mathbb{P}(r_i \text{ includes no school among } \mathcal{S}(z_i) \ | z_i = z) \\
&= \mathbb{P}((v_{ij})_{j \in \mathcal{S}(z_i)} \leq 0 | z_i = z) \\
&= \mathbb{P}(\tilde{v}_{ij} \leq z_{ij}^v \ \forall j \in \mathcal{S}(z_i) | z_i^v = z^v, z_i^- = z^-) \\
&= \mathbb{P}((\tilde{v}_{ij})_{j \in \mathcal{S}(z_i)} \leq (z_j^v)_{j \in \mathcal{S}(z_i)} | z_i^- = z^-)
\end{aligned}$$

and using the independence of the special regressor to recover the distribution of $(v_{ij})_{j \in \mathcal{S}(z_i)} | z_i$.

I now show part (ii). Take $z_i = z$ and $\mathcal{A} \subseteq \mathcal{S}(z)$ with $|\mathcal{A}| \leq L$. Implicitly condition everything on z. Note that for any two events A and B, $\mathbb{P}(A|B)\mathbb{P}(B) \leq \mathbb{P}(A) \leq \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(B^c)$. Consider the events $A = \{v_{ij} \leq 0 \ \forall j \in \mathcal{A}\}$ and $B = \{|r_i| = L, r_i \cap \mathcal{A} = \emptyset\}^c$. One can verify that $\mathbb{P}(A|B) = \mathbb{P}(j \notin r_i \ \forall j \in \mathcal{A}|B)$ using Observation 2. Further, note that $\mathbb{P}(j \notin r_i \ \forall j \in \mathcal{A}|B)$ and $\mathbb{P}(B)$ is observable. Thus, $\mathbb{P}(A) \equiv \mathbb{P}(v_{ij} \leq 0 \ \forall j \in \mathcal{A}) = \mathbb{P}(\tilde{v}_{ij} \leq z_{ij}^v \ \forall j \in \mathcal{A})$ is bounded within an interval of length $\mathbb{P}(B^c)$. One can then use the special regressor similarly as before to bound $\mathbb{P}\big((v_{ij})_{j \in \mathcal{A}} \leq x | z\big)$. □

E Data Appendix

E.1 Eligibility and Priority Groups

Eligibility and priority groups for a program are recorded only for the students who have written down the program in their reports. For consistency, we use the constructed eligibilities and priorities even for those student-school pairs whose actual eligibilities and priorities are observed. While the high school directory offers explicit explanations of criteria for eligibility and priority groups, there are instances where determining if a student meets these criteria based on available data is not feasible. In such cases, approximations are made.

There are several priority and eligibility criteria that we ignore and assume that every applicant satisfies them. These criteria are whether a student attended an information session, whether a student lived in the US for a certain period of time, or whether a student knows or is interested in learning American Sign Language.

There are also criteria that we seek to approximate. Some criteria assess whether a student attended specific middle school *programs*, which is not observed in the data; on the other hand, we observe the middle school (which may contain multiple programs) that each student attends. In these cases, we code the student as satisfying the criteria if the student attends the middle school that contains the program. Such criteria involves either Dual Language Spanish middle school programs or Transitional Bilingual Education Spanish middle school programs.

Some criteria concern granting eligibility or priority to students living in "geographical catchment areas." We approximate these catchment areas based on the addresses (specifically, addresses grouped into school zones) of the students who have applied to these programs and were determined by NYC to be eligible (or ineligible). We apply a similar approach for criteria involving "Brooklyn Area A" and "Brooklyn Area B".

There are also criteria that pertain to students' proficiency in English. While some of these criteria, such as requiring the students to be English Language Learners, are both well-defined and is clearly determinable from the dataset, there are other proficiency criteria that are not directly determinable from data. We use the English Language Learner status to approximate the satisfaction of such criteria.

E.2 Scores and Cutoffs

As in the main text, we model student i's belief regarding program[72] j written at the k-th slot of his report as $q_{ijk} = \mathbb{P}_i\big(\mathsf{cutoff}_j - E_{\mathrm{obj}}[\mathsf{score}_{ij}] + \epsilon^b_{ijk} > 0\big)$ where $\mathbb{P}_i$ is the probability measure of ϵ^b_{ijk}. The belief consists of two parts, namely the objective difference $\mathsf{cutoff}_j - E_{\mathrm{obj}}[\mathsf{score}_{ij}]$ and the subjective assessment ϵ^b_{ijk}. In this section we explain the empirical specification of the objective difference, starting with cutoff_j.

We call a priority group the *threshold priority group* if the subsequent priority groups have no accepted students. We say student i is *contemplated* by program j if i is not assigned to a program listed strictly above j in his report r_i. With these two definitions, we set cutoff_j as the summation of the threshold priority group number and the proportion of accepted students among those who are contemplated by j, within the threshold priority group.[73]

Now we turn to $E_{\mathrm{obj}}[\mathsf{score}_{ij}]$. First, because admissions priority groups are lexicographically more important than the screening outcomes and lotteries (both of which we call tiebreakers), we model $\mathsf{score}_{ij} = \mathsf{priorityGroup}_{ij} + \mathsf{quantile}_{ij}$ where $\mathsf{priorityGroup}_{ij} \in \{1, \ldots, 6\}$ is the admissions priority groups and $\mathsf{quantile}_{ij} \in [0, 1]$ is the quantile of the tiebreaker among the applicants who were *contemplated* by program j. The second term $\mathsf{quantile}_{ij}$ is inherently unobservable (e.g., due to a tiebreaking lottery) from the student's perspective, so he forms an expectation to build

[72]The subscript j denotes pairs of disability type and program, but here we simply call them programs for the sake of simplicity.

[73]By this definition, the cutoff is set as (the last priority group +1) if a program is matched to fewer students than its capacity.

his belief. Therefore we specify the (objectively) expected score $E_{\text{obj}}[\mathsf{score}_{ij}]$ as

$$E_{\text{obj}}[\mathsf{score}_{ij}] = \mathsf{priorityGroup}_{ij} + E_{\text{obj}}[\mathsf{quantile}_{ij}].$$

We detail the construction of $E_{\text{obj}}[\mathsf{quantile}_{ij}]$ momentarily. The uncertainty from the discrepancy between the true score and the objective expectation thereof is subsumed into ν_{ij}.

We do not observe $\mathsf{priorityGroup}_{ij}$ for all (i, j) pairs, and hence we impute their values. As explained in Section E.1, $\mathsf{priorityGroup}_{ij}$ is not observed directly from the dataset if i does not apply to j. We construct $\mathsf{priorityGroup}_{ij}$ based on the priority criteria stated in the school directory which is publicly available. For example, if j states that the program assigns priority group 1 to any students living in Manhattan, and i indeed lives in Manhattan, then we let $\mathsf{priorityGroup}_{ij} = 1$.

Neither is $E_{\text{obj}}[\mathsf{quantile}_{ij}]$ observed for every (i, j) pairs. In this regard, programs can be divided into three categories based on their tie-breaking methods: lottery-based programs, screen-based programs, and Educational Option programs. For lottery-based programs,[74] the tiebreaker is a single lottery, which we do not observe. For these programs, we assign $E_{\text{obj}}[\mathsf{quantile}_{ij}] = 0.5$, the mean of the within-priority group quantile generated by a lottery. For screen-based programs,[75] the tiebreaker is the screening priority and we observe how programs ranked a subset of the applicants by their screening policies.[76] Educational Option programs use both the lottery and screening priority which we detail later.

In order to evaluate $E_{\text{obj}}[\mathsf{quantile}_{ij}]$ for each possible (i, j) pair when j is a screen-based or an Educational Option program, an ideal data would be one in which we observe how a program ranks all the students. However, this is not the case for our data in two senses. First, if a student is not contemplated by a program, the program does not rank the student. Second, even if they are contemplated by the program, they still may not be ranked.

To address this, we predict the counterfactual screening priority ranking as follows. We first run the following OLS regression using (i, j) pairs for which i is ranked by j:

$$\mathsf{rawRank}_{ij} = \beta_j X_i + \delta_{j,\mathsf{priorityGroup}_{ij}} + \epsilon_{ij}$$

where $\mathsf{rawRank}_{ij}$ is the ranking of i evaluated by j in the data, and $\delta_{j,\mathsf{priorityGroup}_{ij}}$ are program and priority group fixed effects. The covariates X_i include English and math test scores, the number of days i has been absent, and the number of days i has been late.

We then use the estimate $\hat{\beta}_j$ to predict the quantile of i within her priority group among

[74]Lottery-based programs are those with admission methods: Unscreened, Limited Unscreened, Zoned Priority, Zoned Guarantee, and For Continuing 8th Graders programs.

[75]Screen-based programs are those with admission the following admission methods: Audition, Screened, Screened: Language, and Screened: Language & Academics.

[76]We observe some violations in the data. 0.35% of screen-based programs do not assign any tiebreakers to their applicants, and among those that do, 5.32% assign the same tiebreaking number if any.

those who were contemplated, according to the data, by j. Specifically,

$$E_{\text{obj}}[\mathsf{quantile}_{ij}] = \frac{1}{|\mathcal{C}_{ij}|} \sum_{i' \in \mathcal{C}_{ij}} 1(\hat{\beta}_j X_{i'} \leq \hat{\beta}_j X_i)$$

where $\mathcal{C}_{ij} = \{i' : \mathsf{priorityGroup}_{i'j} = \mathsf{priorityGroup}_{ij}, i' \text{ is contemplated by } j \text{ according to the data}\}$.

Educational Option programs, according to the NYC high school directory, "admit students who have high, middle, and low reading levels. Half of the students in each reading level group will be selected based on their rankings from the school using multiple criteria. The other half will be selected randomly from the remaining applicants." Following Che and Tercieux (2019), we create six "virtual subprograms" for each Educational Option program, namely HR, HS, MR, MS, LR, and LS, where H, M, and L indicate high, middle, and low reading levels respectively, while R and S indicate random and select.

We let subprograms HR and HS share the same cutoff level $\mathsf{cutoff}_j(H)$, which is computed as above but conditional on the reading level being high; i.e., $\mathsf{cutoff}_j(H)$ is the summation of the threshold priority group (which is the priority group whose subsequent priority groups have no accepted students with high reading level) and the proportion of accepted students among those who are contemplated by j and have high reading level, within the threshold priority group.

We let $E_{\text{obj}}[\mathsf{quantile}_{ij}(HR)] = 0.5$ as above (since HR is a random subprogram). We calculate $E_{\text{obj}}[\mathsf{quantile}_{ij}(HS)]$ in a similar manner to screen-based programs. This is less straightforward, however, because we do not observe which students are "contemplated" by HS (even though we do observe the students who are contemplated by j as a whole). For this, we run deferred-acceptance algorithm to simulate which students are contemplated at the subprogram level. This requires students to rank the virtual subprograms, for which we again follow Che and Tercieux (2019); a student who applies to an Educational Option program j is assumed to rank the subprograms according to the order HR, HS, MR, MS, LR, and LS. The simulation matches 73.4% of the students who were matched to an Educational Option program (in the data) to the same program. For subprogram matching, we use these correct matches only. Other subprograms, MR, MS, LR, and LS, are treated analogously.

In the end, as we need i's belief on the Educational Option program rather than on its subprograms, we use the maximum[77] of the objective differences of the subprograms to approximate the belief on the program, i.e.,

$$q_{ijk} = \mathbb{P}_i\Big(\max_{s \in \mathcal{S}} [\mathsf{cutoff}_j - E_{\text{obj}}[\mathsf{score}_{ij}(s)]] + \epsilon^b_{ijk} > 0 \Big)$$

where $\mathcal{S} = \{HR, HS, \ldots, LS\}$ is the set of subprograms of j.

In the equation determining c_{ij}, a proxy for objective chance of admission enters the equation: the difference in objective expected scores and cutoffs. They correspond to $\mathsf{cutoff}_j - E_{\text{obj}}[\mathsf{score}_{ij}]$.

[77] We take maximum, instead of, say, average, to reflect that a student is accepted by an Educational Option program if the student is accepted by any of its subprograms.

E.3 Distances

To calculate distance measures, we rely on the centroid of the student's census block and the precise locations of the schools. For this computation, we employ the Haversine formula, a method used in navigation providing great-circle distances between two points on a sphere from their longitudes and latitudes. Distances are expressed in miles.

E.4 Program Admission Methods and Interest Area

The following admission methods, as defined by NYC DOE, use lottery to break the ties: "Unscreened", "Limited Unscreened", "Zoned Priority", "Zoned Guarantee", and "For Continuing 8th Graders". The following admission methods use screening policies to break the ties: "Audition", "Screened", "Screened: Language", and "Screened: Language & Academics". "Educational Option" programs use both screening and lotteries to break the ties. Such programs were counted towards the calculation of proportion of programs that *uses screening* in Table 2.

In terms of programs' interest areas, Table 2 and in our estimation of the model of application behavior (Table 6) defines some programs to be of *Arts* programs or *STEM* programs. Arts programs are the programs that have one of "Performing Arts", "Visual Art & Design", and "Performing Arts/Visual Art & Design" as their interest area as defined by NYC DOE. Similarly, STEM programs are those that have "Computer Science & Technology", "Engineering", and "Science & Math" as their interest area.

E.5 Construction of Figures and Tables

Table 1 presents the variable *State Reading Category*, which indicates students' reading performance on the New York State English Language Arts (ELA) test scores, whenever available. The "Low" category indicates students who scored in the bottom 16%, the "Middle" category indicates students who scored in the middle 68%, and the "High" category represents students who scored in the top 16%. For non-public school students, the reading category is calculated based on another standardized assessment, which is then used for the purposes of categorization for admissions criteria for Educational Option programs.

Tables 3, 4, and 5 control for the following variables. Student-specific variables are ethnicity, sex, subsidized lunch, math score, disability status, and borough. Program- or school-specific variables include page rank (for Table 5), coed, school borough, graduation rate, the percentage of students who enroll in college or career programs, attendance rate, admissions method, interest area, the percentage of students who feel safe on the premises, and (log of) the number of enrolled students. Match-specific variables are the priority group (for Table 3 and 5), the interaction between sex and coed, whether the student's borough matches the school's borough, whether the student's feeder school is close (less than 0.5 miles) to the high school, whether the high school is the feeder school, the distance (between the student and the school), its square, and the student's own ethnicity interacted with the percentages of each ethnicity group in the school.

In Table 9, Panel A discusses two counterfactual matchings without school choice: Random matching and Neighborhood matching. *Random* matching randomly allocates the students to the programs respecting the capacity constraints of the programs. *Neighborhood* matching approximately minimizes the total distance traveled by the students to the programs subject to the capacity constraints.[78]

Other matchings (Panel B) in Table 9 reflect different versions of school choice. Matchings in Panel B.a. represent the status-quo school choice. *Actual* matching is the actual school choice matching in 2017 from the main round of DA. *Estimated* matching is the result from a simulated DA using the estimated model of student behavior, coupled with the approximated admission policies by the programs.[79]

The matchings in Panel B.b shut off different factors' influences one by one. *Change to Truthful among Considered* is the same as the Estimated matching except that students truthfully report their considered programs in the order of their preferences until they run out of the programs that are preferred to the outside option or reach the twelve-program threshold. Note that the students may still drop programs they are unaware of or they feel out of reach. *Change to Full Consideration* matching then further turns off limited consideration by assuming that students consider every eligible program. *Change to Random Screening* matching turns off the schools' screening policies by forcing the programs endowed with screening ability to randomly screen students. *Change to No Admissions Priorities* matching then removes the admissions priority groups. Note that this matching purely reflects student preferences without the influences of limited consideration, non-truthtelling behavior, nor the schools' admissions priorities and screening policies. In this regard, an alternative name for the matching is *Student-Preferences-Only Choice*.

E.6 Missing School Characteristics

The dataset had some instances of missing values for the following school characteristics: graduation rate, college/career rate, and percentage of students feeling safe. To perform our counterfactual analysis, we needed to predict students' utilities, consideration probabilities, and beliefs for every program. Therefore, we took the approach of imputing the missing values. We used the predicted values from the ordinary least squares regressions of each of these variables on the following characteristics: attendance rate, average grade 8 math proficiency, percentage of students eligible for Human Resources Administration, enrollment size, and the percentage of White, Black, and Asian students. These regressions utilized only the non-missing observations.

[78] Refer to Appendix G.2 for approximation details.

[79] Refer to Appendix E for details on the estimation of the screening policies and the approximation of the priority groups. See Appendix G.3 for details on our implementation of Deferred Acceptance algorithm.

F Estimation: Details

F.1 Likelihood of Inclusion

Here we derive the formula of likelihoods of school inclusions and discuss why the true parameters maximize the likelihoods. The likelihoods that we consider are not standard in the sense that (1) they select students with $s_{ij} := \mathbb{1}(|r_i \backslash \{j\}| < 11) = 1$ and (2) one of the likelihoods is weighted. We show that the true parameters maximize the likelihoods despite being non-standard.

We first derive the formula of log-likelihood of inclusion of school j in the report of applicant i. The log-likelihood reflects the identifying information in Observations 1 and 2. It selects individuals with $s_{ij} = 1$ (rather than those with $|r_i| < 12$) to resolve selection issues explained in Section 7.1; Lemma F.1 shows that, given $(\epsilon^v_{ij}, \epsilon^c_{ij})_{j\in\mathcal{J}}$ is $i.i.d$ across j, $|r_i \backslash \{j\}| < 11$ is independent of $(\epsilon^v_{ij}, \epsilon^c_{ij})$ conditional on observables (x_j, z_{ij}), where x_j is the union of all variables in x^v_j and x^c_j (as defined in Section 6), and similarly for z_{ij}. Let $\iota_{ij} := \mathbb{1}(j \in r_i)$ denote the random variable indicating whether school j was included in the report r_i. Let $w_{ij} := \mathbb{1}(v_{ij} > 0)\mathbb{1}(c_{ij} > 0)$ and note that $w_{ij} = \iota_{ij}$ whenever $s_{ij} = 1$ following Observation 1. Let $f_{w|z,s}(\cdot|z', s'; \theta)$ denote the density of w_{ij} given $z_{ij} = z'$, $s_{ij} = s'$, and θ. Similarly define $f_{\iota|z,s}(\cdot|z', s'; \theta)$ and $f_{w|z}(\cdot|z'; \theta)$. We treat $(x_j)_j$ as nonrandom in this subsection. Then,

$$
\begin{aligned}
&\log \Pi_i \Pi_{j:s_{ij}=1} f_{\iota|z,s}(\iota_{ij}|z_{ij}, 1; \theta) \qquad \text{(F.1)}\\
&= \log \Pi_i \Pi_{j:s_{ij}=1} f_{w|z,s}(w_{ij}|z_{ij}, 1; \theta)\\
&= \log \Pi_i \Pi_{j:s_{ij}=1} f_{w|z}(w_{ij}|z_{ij}; \theta)\\
&= \log \Pi_i \Pi_{j:s_{ij}=1, j\notin\mathcal{S}_i} \big(1 - \mathbb{P}(v_{ij} > 0|z_{ij}; \theta^v)\mathbb{P}(c_{ij} > 0|z_{ij}; \theta)\big)^{1-w_{ij}} \cdots\\
&\qquad\qquad \big(\mathbb{P}(v_{ij} > 0|z_{ij}; \theta^v)\mathbb{P}(c_{ij} > 0|z_{ij}; \theta^c)\big)^{w_{ij}} \cdots\\
&\qquad \Pi_{j:j\in\mathcal{S}_i} \big(1 - \mathbb{P}(v_{ij} > 0|z_{ij}; \theta^v)\big)^{1-w_{ij}} \mathbb{P}(v_{ij} > 0|z_{ij}; \theta^v)^{w_{ij}}\\
&= \sum_i \Bigg[\sum_{j:s_{ij}=1, j\notin\mathcal{S}_i} \Big[\big(1 - w_{ij}\big) \log \big(1 - \overline{\Phi}(-\psi^v_{ij})\overline{\Phi}(-\psi^c_{ij})\big) + w_{ij} \log \big(\overline{\Phi}(-\psi^v_{ij})\overline{\Phi}(-\psi^c_{ij})\big) \Big]\\
&\qquad + \sum_{j:s_{ij}=1, j\in\mathcal{S}_i} \Big[(1 - w_{ij}) \log \big(\Phi(-\psi^v_{ij})\big) + w_{ij} \log \big(\overline{\Phi}(-\psi^v_{ij})\big) \Big] \Bigg]
\end{aligned}
$$

where $\bar{\Phi}(\cdot) := 1 - \Phi(\cdot)$, $\psi^v_{ij} := v_{ij} - \epsilon^v_{ij}$, $\psi^c_{ij} := c_{ij} - \epsilon^c_{ij}$, θ^v denotes the preference parameters, and θ^c denotes the consideration parameters. For notational convenience, the dependence of ψ^v_{ij} on θ^v and the dependence of ψ^c_{ij} on θ^c are made implicit. The second equality comes from $s_{ij} = 1$ being independent of $(\epsilon^v_{ij}, \epsilon^c_{ij})$ and therefore also of (v_{ij}, c_{ij}) conditional on observables. Note that the second summation in the last expression, which consists only of the programs that are surely considered, exclusively reflects the variation in Observation 2 and therefore is a function solely of the preference parameters. It is also possible to estimate preference parameters using this part of the partial likelihood. The summary of the results are given in Figure B.3.

We now show that the population version of the log-likelihood is maximized by the true parameters θ_0. Define

$$Q(\theta) := E_{\theta_0} \sum_{j:s_{ij}=1} \log f_{\iota|z,s}(\iota_{ij}|z_{ij}, 1; \theta) \equiv E_{\theta_0} \sum_{j:s_{ij}=1} \log f_{w|z}(w_{ij}|z_{ij}; \theta)$$

where both w_{ij} and z_{ij} are random variables. This is the population version of the log-likelihood (Equation F.1) in the sense that $\operatorname{plim}_{n\to\infty} n^{-1} \log \Pi_i \Pi_{j:s_{ij}=1} f_{\iota|z,s}(\iota_{ij}|z_{ij}, 1; \theta) = Q(\theta)$ where n denotes the number of students in the sample and with the understanding that in the left-hand side (ι_{ij}, z_{ij}) are realized values.

Now we show $Q(\theta_0) \geq Q(\theta)$ for all θ. Note that

$$\begin{aligned}
Q(\theta) - Q(\theta_0) &= \sum_j E_{\theta_0}\Big[s_{ij} \log \frac{f_{w|z}(w_{ij}|z_{ij}; \theta)}{f_{w|z}(w_{ij}|z_{ij}; \theta_0)}\Big] \\
&= \sum_j E_{\theta_0}\Big[s_{ij} E_{\theta_0}\Big[\log \frac{f_{w|z}(w_{ij}|z_{ij}; \theta)}{f_{w|z}(w_{ij}|z_{ij}; \theta_0)}\Big| s_{ij}, z_{ij}\Big]\Big] \\
&\leq \sum_j E_{\theta_0}\Big[s_{ij} \log E_{\theta_0}\Big[\frac{f_{w|z}(w_{ij}|z_{ij}; \theta)}{f_{w|z}(w_{ij}|z_{ij}; \theta_0)}\Big| s_{ij}, z_{ij}\Big]\Big] \\
&= \sum_j E_{\theta_0}\Big[s_{ij} \log E_{\theta_0}\Big[\frac{f_{w|z}(w_{ij}|z_{ij}; \theta)}{f_{w|z}(w_{ij}|z_{ij}; \theta_0)}\Big| z_{ij}\Big]\Big] = 0
\end{aligned}$$

where the inequality holds by Jensen's inequality, the penultimate inequality holds from $(c_{ij}, v_{ij}) \perp\!\!\!\perp s_{ij}|z_{ij}$ and therefore $w_{ij} := \mathbb{1}(c_{ij} > 0)\mathbb{1}(v_{ij} > 0) \perp\!\!\!\perp s_{ij}|z_{ij}$, and the last equality holds from

$$E_{\theta_0}\Big[\frac{f_{w|z}(w_{ij}|z_{ij}; \theta)}{f_{w|z}(w_{ij}|z_{ij}; \theta_0)}\Big| z_{ij}\Big] = \frac{f_{w|z}(0|z_{ij}; \theta)}{f_{w|z}(0|z_{ij}; \theta_0)} f_{w|z}(0|z_{ij}; \theta_0) + \frac{f_{w|z}(1|z_{ij}; \theta)}{f_{w|z}(1|z_{ij}; \theta_0)} f_{w|z}(1|z_{ij}; \theta_0) = 1.$$

As the sample size of (i, j) pairs such that i surely considers j is small relatively those that do not have the sure-consideration relationship, we weight higher the sure-consideration pairs. The weighted log-likelihood is

$$\begin{aligned}
\omega_{\text{NSC}} \sum_i \sum_{j:s_{ij}=1, j\notin \mathcal{S}_i} &\Big[(1 - \mathbb{1}(j \in r_i)) \log\big(1 - \overline{\Phi}(-\psi^v_{ij})\overline{\Phi}(-\psi^c_{ij})\big) + \mathbb{1}(j \in r_i) \log\big(\overline{\Phi}(-\psi^v_{ij})\overline{\Phi}(-\psi^c_{ij})\big)\Big] \\
&+ \omega_{\text{SC}} \sum_i \sum_{j:s_{ij}=1, j\in \mathcal{S}_i} \Big[(1 - \mathbb{1}(j \in r_i)) \log\big(\Phi(-\psi^v_{ij})\big) + \mathbb{1}(j \in r_i) \log\big(\overline{\Phi}(-\psi^v_{ij})\big)\Big]
\end{aligned}$$

for some weights ω_{NSC} and ω_{SC} such that $\omega_{NSC} \sum_{j\notin \mathcal{S}_i} s_{ij} + w_{SC} \sum_{j\in \mathcal{S}_i} s_{ij} = \sum_{j\in \mathcal{J}} s_{ij}$. That is, the schools that are not surely considered are weighted by ω_{NSC} and those that are surely considered are weighted by ω_{SC}.

The true parameters maximize the population version of the weighted likelihood. To see this, it suffices to show that the true preference parameters maximize the second term above (as $\omega_{\text{SC}} > \omega_{\text{NSC}}$ and the weighted likelihood can be expressed as the sum of unweighted likelihood multiplied by ω_{NSC} and the second term weighted by $\omega_{\text{SC}} - \omega_{\text{NSC}}$). But the sure-consideration event $j \in \mathcal{S}_i$ is determined by the observables and is independent of (c_{ij}, v_{ij}) conditional on z_{ij}. Then, the Jensen's inequality above holds with the s_{ij} replaced as $\tilde{s}_{ij} := s_{ij}\mathbb{1}(j \in \mathcal{S}_i)$.

F.2 Simulated Ordering Moments

In this section, we denote the union of variables in $(x_j^v, x_j^c, z_{ij}^v, z_{ij}^c)$, as defined in the empirical specification (Section 6), as simply z_{ij} for notational convenience. For any $f : \mathcal{R} \to \mathbb{R}^m$,

$$0 = \mathbb{E}\big[f(r_i) - \mathbb{E}[f(r_i)|z_i]\big|z_i\big] = \mathbb{E}\big[f(r_i) - \mathbb{E}[f(r(z_i, e_i; \theta_0))|z_i]\big|z_i\big]$$

where e_i denotes the vector of unobservables $(\epsilon_i^v, \epsilon_i^c, \eta_i)$, θ denotes the parameter vector, θ_0 denotes the true parameter vector, and $r(z_i, e_i; \theta)$ denotes the subjectively optimal report under (z_i, e_i, θ) which is uniquely defined with probability 1. Section G.1 describes the procedure for simulating $r(z_i, e_i; \theta)$. It follows that

$$\mathbb{E}\Bigg[\Big(f(r_i) - \mathbb{E}\big[f\big(r(z_i, e_i; \theta_0)\big)\big|z_i\big]\Big)h(z_i)\Bigg] = \mathbb{E}\Bigg[\mathbb{E}\Big[f(r_i) - \mathbb{E}\big[f\big(r(z_i, e_i; \theta_0)\big)\big|z_i\big]\Big|z_i\Big]h(z_i)\Bigg] = 0$$

where $h(z_i)$ may be a $m-$dimensional vector.

The sample equivalent of this condition is

$$\frac{1}{I}\sum_i \Big(f(r_i) - \mathbb{E}^{\text{sim}}\big[f\big(r(z_i, e_i; \theta_0)\big)\big|z_i\big]\Big)h(z_i) = 0 \tag{F.2}$$

where in the brute-force version of simulation $\mathbb{E}^{\text{sim}}\big[f\big(r(z_i, e_i; \theta_0)\big)\big|z_i\big] = \frac{1}{S}\sum_s f\big(r(z_i, e_i^s; \theta_0)\big)$ where the distribution of e_i^s is completely governed by θ_0 and not by z_i due to independence.

The simulated ordering moments gives information about how individuals order the schools:

$$\mathbb{E}\Big[\frac{1}{J}\sum_j \Big(\mathbb{1}(j \in r_i^k) - \mathbb{P}\big(j \in r^k(z_i, e_i; \theta)\big)h_j(z_i)\Big] = 0 \quad \forall k = 1, \dots, 12$$

where r_i^k is represents the report r_i truncated up to the kth slot, $r^k(\cdot)$ is the equivalent for the simulated report, and the set inclusion notation is used towards r_i^k and $r^k(\cdot)$ with a slight abuse. The condition uses $f(r_i) = \frac{1}{J}\big(\mathbb{1}(j \in r_i^k)\big)_{j \in \mathcal{J}}$ in the notation of Equation F.2. The moment condition is implemented by

$$\frac{1}{IJ}\sum_i \sum_j \Big(\mathbb{1}(j \in r_i^k) - \mathbb{E}^{\text{sim}}\big[\mathbb{1}\big(j \in r^k(z_i, e_i; \theta)\big)\big|z_i\big]\Big)h_j(z_i)$$

with $S = 1$. Using one simulation draw per observation is justified as the simulator $\mathbb{E}^{\text{sim}}\big[\mathbb{1}\big(j \in r^k(z_i, e_i; \theta)\big)\big|z_i\big]$ is unbiased for $\mathbb{E}[\mathbb{1}(j \in r_i^k)]$ and therefore rely on the law of large numbers with respect to the observations to control for simulation error (McFadden, 1989). And we use

$h(z_i) = \left(1, z_{ij}, (z_{ij} - \bar{z}_i)^2, \mathsf{cutoff}_{ij} - E_{\text{obj}}[\mathsf{score}_{ij}]\right)_{j \in \mathcal{J}}$ where we remind the readers that we are using a shorthand expression: z_{ij} includes all variables in $(x_j^v, x_j^c, z_{ij}^v, z_{ij}^c)$.

Potentially because of non-smoothness of the criterion function with respect to the parameters due to simulations, in the second stage of estimation—where we use these ordering moments to recover the belief parameters—the traditional gradient-based algorithms or Knelder-Mead algorithms did not work well to find the minimizer. We instead relied on grid search on the two-dimensional grid (per each ethnicity) to find the minimizer.

F.3 Lemmas

Lemma F.1. *The event $|r_i \setminus \{j\}| < 11$ is independent of $(\epsilon_{ij}^c, \epsilon_{ij}^v)$ conditional on observables.*

Proof. Fix the observables (x, z). We shall show that the event $|r_i \setminus \{j\}| < 11$ is the same as the event $\sum_{j' \neq j} \mathbb{1}\{c_{ij'} > 0, v_{ij'} > 0\} < 11$. Being determined by only $(\epsilon_{ij'}^c, \epsilon_{ij'}^v)_{j' \neq j}$, the latter is independent of $(\epsilon_{ij}^c, \epsilon_{ij}^v)$ as desired.

Note that

$$
\begin{aligned}
|r_i \setminus \{j\}| < 11 \quad & \text{iff} \quad |r_i \setminus \{j\}| < 11 \quad \text{and} \quad |r_i| < 12 \\
& \text{iff} \quad \sum_{j' \neq j} \mathbb{1}\{c_{ij'} > 0, v_{ij'} > 0\} < 11 \quad \text{and} \quad |r_i| < 12 \\
& \text{iff} \quad \sum_{j' \neq j} \mathbb{1}\{c_{ij'} > 0, v_{ij'} > 0\} < 11.
\end{aligned}
$$

The second equivalence holds due to the first statement of Observation 1 (since $|r_i| < 12$). For the last equivalence, "only if" holds trivially. The "if" holds due to the second statement of Observation 1. ☐

G Simulations

G.1 Simulating Subjectively Optimal Reports

Here we describe the procedure for calculating the subjectively optimal reports:

$$
r(z_i, e_i, \theta) = \arg \max_{r \in \mathcal{R}(\mathcal{C}_i)} \sum_{j=0}^{J} p_{ij}^r v_{ij} \tag{G.1}
$$

where the distribution of $(\mathcal{C}_i, v_{ij}, p_{ij}^r)_{ij}$ depends on θ. We ignore ties in optimal reports as they occur with probability zero. Note that $\arg\max_{r \in \mathcal{R}(\mathcal{C}_i)} \sum_{j=0}^{J} p_{ij}^r v_{ij} = \arg\max_{r \in \mathcal{R}(\mathcal{J}_i^+)} \sum_{j=0}^{J} p_{ij}^r v_{ij}$ where $\mathcal{J}_i^+ = \{j \in \mathcal{C}_i | v_{ij} > 0\}$ is the set of schools that are considered by i and are preferred to the outside option. The equality holds since students will never wish to list any school outside $\mathcal{J}_i^+$.

The optimization problem is difficult to solve since the size of a choice set, even after being reduced to $\mathcal{R}(\mathcal{J}_i^+)$, can be large. For instance, with $|\mathcal{J}_i^+| = 20$, the choice set $\mathcal{R}(\mathcal{J}_i^+)$ is all possible ordered lists using the schools in $\mathcal{J}_i^+$ which has as many as $20!/(20-12)! \simeq 6.03 * 10^{13}$ elements. As in Calsamiglia et al. (2020), to make this problem solvable through backward induction, we represent this problem as what resembles a finite-horizon dynamic programming problem, where a "period" is a slot in the list and a state is the set of schools already listed.

Let j_k represent the school listed in the kth spot. Note $p_{ij}^r = \Pi_{l=1}^{k-1}(1 - q_{ij_{r_l}l})q_{ijk}$. Let $K = \min\{12, |\mathcal{J}_i^+|\}$, which represents the last slot (or period) that the student optimally fills in. Each student solves the following problem:

$$\arg \max_{r \in \mathcal{R}(\mathcal{J}_i^+)} \sum_{j=0}^{J} p_{ij}^r v_{ij}$$
$$= \max_{\{j_1, \cdots, j_K\} \subset \mathcal{J}_i^+} q_{ij_1 1} v_{ij_1} + (1 - q_{ij_1 1})\Big(q_{ij_2 2} v_{ij_2} + \cdots + (1 - q_{ij_2 2}) \cdots (1 - q_{ij_{11} 11}) q_{ij_K} v_{ij_K}\Big).$$

We solve the problem backwards from the last school the student puts in the list. Let $\mathcal{J}_k = \{j_1, \cdots, j_k\}$. Let

$$V_K^i(\{j_1, \cdots, j_{K-1}\}) = \max_{j \in \mathcal{J}_i^+ \setminus \mathcal{J}_{K-1}} q_{ijK} v_{ij}$$

and, for $1 \leq k < K$, let

$$V_k^i(\{j_1, \cdots, j_{k-1}\}) = \max_{j \in \mathcal{J}_i^+ \setminus \mathcal{J}_{k-1}} q_{ijk} v_{ij} + (1 - q_{ijk}) V_{k+1}^i(\{j_1, \cdots, j_{k-1}, j\}).$$

Then,

$$V_1^i = \max_{j \in \mathcal{J}_i^+} q_{ij1} v_{ij} + (1 - q_{ij1}) V_2^i(\{j\}) = \max_{r \in \mathcal{R}(\mathcal{J}_i^+)} \sum_{j=0}^{J} p_{ij}^r v_{ij},$$

which shows that the original problem may be solved via the dynamic formulation.

G.2 Simulation of Neighborhood Matching

Finding the optimal matching that minimizes the sum of distance-to-school is an integer linear programming problem, and its computation time increases nonlinearly with the number of students. To reduce the computation time, we adopt an iterative approach to approximate the total distance minimization. In the initial step, we randomly select 10,000 students and match them to programs to minimize the sum of distance traveled, with program capacities adjusted proportionally. We then iterate this procedure, considering the remaining students and program seats, until all students are matched.

G.3 Simulation of Deferred Acceptance Algorithm

Capacities The simulation exercises require program capacities (for each disability type) as one of their inputs, which we take from 2018 High School Directory because it states the capacities for the year 2017. For zoned programs and feeder-only programs, however, the directory does not state the capacities. In these cases, we use the number of students who are in the corresponding school zones and who are from the corresponding feeder schools, respectively. For programs that appear in 2017 High School Directory but not in 2018, we use the capacities as stated in 2017 High School Directory.

In the data, students are often matched beyond the stated capacities; for example, 51.34% of the programs admitted more students (13.86 students on average) in round 1 than their capacity as stated in 2018 High School Directory. Therefore we take the number of actually matched students as the program capacity if it is greater than what we have obtained in the previous paragraph.

We have complete non-missing data only for the students who are attending NYC public schools and therefore have used only such students for estimation. In the simulation of DA, we also use only these students, which are 92.2% of the total students. To account for this, we reduce the capacities of programs proportionately to be 92.2% of their estimated capacities.

Finally, for Educational Option programs, the capacity is divided into six virtual subprograms: $50\% \times 16\%$ of the total capacity goes to each of HR, HS, LR, and LS subprograms, and $50\% \times 68\%$ goes to each of MR and MS subprograms.

Preferences of Students and Rankings by Programs Other inputs of simulation include preferences of students and programs. The preferences of students are formed as described in Section 4, given parameter values and policies such as information interventions. A slight complication involves Educational Option programs; for those, we follow Che and Tercieux (2019) and let a student, whenever she includes an Education Option program in her report, rank its subprograms in the order of HR, HS, MR, MS, LR, and LS.

For simulation, we need each program to rank all the counterfactual applicants for the program, which does not necessarily coincide with the set of its actual applicants in our data. Because of this, we let the programs rank the students according to the objective expected score $E_{\text{obj}}[\mathsf{score}_{ij}]$, defined in Section 6, instead of the actual ranking reported by the program in the data; the former is defined for each pair of student i and program j, whereas the latter is available only when i is an actual applicant for j in the data. In case of lottery-based programs or ties in the objective expected scores, we use individual-specific lottery number to break ties.