# Distributional Impacts of Centralized School Choice*

Jaewon Lee† Suk Joon Son‡

August 17, 2024

First Public Version: November 2020

**Abstract**

Informational frictions in centralized school choice can influence distributional outcomes and welfare. We model school applications, allowing for limited consideration sets and mistaken beliefs about admission chances. Quasi-experimental variation and various aspects of rank-ordered lists are utilized for identification. We then assess the effects of school choice in New York City on racial segregation, equity in student welfare, and matching stability. We find that while school choice enhances welfare across races, limited consideration compromises these gains, particularly for Black and Hispanic students. A counterfactual policy with personalized school recommendations could 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

*An earlier version of this paper was circulated in 2020 as Son's job market paper. 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, Shotaro Beppu, Marianne Bernatzky Koehli, Barbara Biasi, Ali Bray, Jin-wook Chang, Yeon-koo Che, Andrew Chesher, Susan Dynarski, Anaïs Fabre, Chao Fu, Kenneth Gillingham, Charles Hodgson, John Eric Humphries, Mitsuru Igami, Gerald Jaynes, Daiji Kawaguchi, Kei Kawai, Alvin Klevorick, Changhyun Kwak, Koohyun Kwon, Soonwoo Kwon, Mariana Laverde, Kwok Hao Lee, Costas Meghir, Luis Carvalho Monteiro, Christopher Neilson, Cormac O'Dea, Parag Pathak, 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 JSPS KAKENHI 22K13412 and JST ERATO JPMJER2301. It is also financially supported by the Departmental Research Grant of the University of Tokyo. The views expressed herein do not purport to represent the views of Compass Lexecon.

†Compass Lexecon, jaewon.lee@compasslexecon.com

‡The University of Tokyo, sjson@e.u-tokyo.ac.jp

## 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 target desirable properties such as stability or strategy-proofness.[1] However, their real-world impacts remain contested. The theoretical results depend on the assumption that applicants make choices with full knowledge of all available options, which is questionable in many centralized school choice environments with numerous differentiated options. Furthermore, informational frictions may vary across demographic groups, potentially having significant distributional impacts. However, while there is growing evidence documenting the presence of these frictions in school choice settings,[2] building and identifying a model of school applications that incorporates both limited awareness of available options and two-sided matching remains less understood. This is despite its importance for distinguishing the impacts of student preferences from information frictions on racial segregation and equity, accurately measuring the welfare net of confounding effects from optimization frictions, assessing the satisfaction of theoretical targets in practice, and evaluating counterfactual policies, among other considerations.

The centralized high school choice process in New York City (NYC) is one environment where informational frictions can be significant, with more than 700 high school programs in NYC. Guided by the empirical evidence of frictions we find in NYC, we construct a model that allows for limited awareness of school options as well as incorrect beliefs about assignment chances.[3] We demonstrate that the model can be estimated using data commonly available in school choice settings, namely rank-ordered choice data, along with an instrument that shifts awareness but not preferences. We support this claim by providing sufficient conditions for nonparametric identification. In our NYC school choice environment, we utilize the alphabetical ordering of schools in the NYC high school directory as an instrument.

Our results show that Asian and White students' consideration sets are more aligned with their preferred schools compared to those of Black and Hispanic students. Students' preferences, net of the confounding effects of limited information, act to integrate schools. In contrast, informational frictions substantially suppress student welfare, particularly for Black and Hispanic students. These findings underscore the importance of separating preference

[1]See, e.g., Gale and Shapley (1962) and Abdulkadiroğlu and Sönmez (2003).

[2]See, e.g., Corcoran, Jennings, Cohodes, and Sattin-Bajaj (2018), and Arteaga, Kapor, Neilson, and Zimmerman (2022).

[3]Allowing for subject beliefs about assignment chances and nontruthful behavior may be crucial given the complexity of NYC's Deferred Acceptance (DA) mechanism (Calsamiglia, Haeringer, and Klijn, 2010; Hassidim, Marciano, Romm, and Shorrer, 2017).

and information frictions, offering guidance for designing policy interventions. Building on these insights, we propose personalized school recommendations based on the estimated preferences and consideration probabilities. Counterfactual simulations predict that such information interventions will recover 20–36% of welfare losses.

Our paper begins by presenting descriptive evidence suggesting frictions in application behavior 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. These 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[4] and have incorrect beliefs about equilibrium assignment chances. An applicant may fail to consider a school because she is unaware of it 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. For example, while a lack of consideration may affect *which* schools are listed, it cannot affect *where* a listed school will be ranked. Regarding 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.[5] Another assumption that assists identification is that some students have

[4]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.

[5]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

a set of schools (e.g., less selective schools close to home) that they will surely consider. We formalize our intuitive identification strategy by establishing sufficient conditions for non-parametric 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.

Estimates reveal racial differences in both preference and consideration patterns, with the differences in consideration being more pronounced. Moreover, compared to Asian and White students, Black and Hispanic students' consideration sets are less aligned with their preferences. For instance, they are significantly more likely to consider less selective and lower-performing schools, even though their preferences towards them are comparable to those of Asian and White students. 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 modestly enhances racial integration, reducing the isolation index of Black students by about 7.7 percentage points. Student preferences work to integrate races, while limited consideration has mixed impacts across races. Furthermore, school choice also significantly boosts 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. 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.[6] These interventions utilize the estimated preferences and consideration sets to recommend 30 programs to each student. Some of these interventions show significant promise. The most effective one, which recommends programs with the highest predicted utility but low predicted consideration chances, is estimated to address between 20–36% of the welfare losses.

We also measure matching stability by quantifying the prevalence of justified envy, a

considered the effects of the positioning of items in online settings; see, e.g., Feng, Bhargava, and Pennock (2007), Koulayev (2014), Ursu (2018).

[6] Allende, Gallego, and Neilson (2019) also use the estimated model to study alternative designs of their information intervention.

situation where a student and a school program prefer each other over their current matches. Our estimates indicate that 73% of students experience justified envy toward at least one school. On average, however, students view only around three school programs with justified envy, a relatively small number compared to around 750 available programs.

**Relation to the Literature** We contribute to the literature that estimates a model of school applications with limited awareness. Ajayi and Sidibe (2022) study high school choice in Ghana based on a sequential search model, documenting welfare loss from information frictions and estimating their model with surveys on beliefs about admission chances. A recent working paper by Agte, Allende, Kapor, Neilson, and Ochoa (2024) examines primary school assignments in Chile. Using a set of surveys and interventions, they focus on unveiling the search process, assessing the roles of limited awareness, beliefs about school characteristics and admission chances, and search costs, assuming a sequential search model. We complement these studies by using and demonstrating a different approach that does not rely on surveys or interventions but rather only on observational data, coupled with a shifter of information excluded from preferences. We also support our estimation strategy with nonparametric identification results. Also, our model of consideration sets is an alternative-specific consideration model (Manski, 1977; Swait and Ben-Akiva, 1987), not a sequential search model. We do not interpret our consideration model as counterfactual-invariant, but rather focus on the impacts of status quo informational frictions on distributional aspects and the satisfaction of theoretical targets.

Our identification 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,[7] students can list and rank-order multiple schools, which provides additional identifying variation. Our model also distinguishes between consideration and nondegenerate beliefs given consideration. In addition to one instrument, our nonparametric identification results also depend on the presence of a special regressor.[8] Our identification strategy builds upon Agarwal and Somaini (2018), who provide sufficient conditions for nonparametric identification of preferences assuming full consideration and holding fixed a mode of beliefs in a centralized school choice setting.[9] Our paper also relates to the broader literature on the estimation and identification

[7]As a supplementary nonparametric identification result, we discuss the case where both types of instruments are present (Proposition C.4), unlike our empirical setting.

[8]The results depend on a special regressor for utility and one for consideration. These regressors may coincide.

[9]The approaches used in nonparametric identification results are further related to, for example, Thomp-

of discrete choice models with limited consideration,[10] with one important difference being that we also allow and estimate potentially incorrect beliefs about admission chances, which is important in two-sided matching environments such as school choice.

Other studies have documented the importance of limited information about schools. Using surveys and informational interventions, Arteaga, Kapor, Neilson, and Zimmerman (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, Jennings, Cohodes, and Sattin-Bajaj (2018) provide evidence that information intervention affects application behavior in the NYC high school 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. Allende, Gallego, and Neilson (2019) estimate a school choice model featuring imperfect information about the school attributes, highlighting equilibrium effects. Campos (2024) finds that information spillovers between parents are important. Informational frictions are also important in other environments of school or college applications (e.g., Hastings and Weinstein, 2008; Hoxby and Turner, 2013; Ajayi, Friedman, and Lucas, 2017; Dynarski, Libassi, Michelmore, and Owen, 2021). Our contribution is to demonstrate racial disparities in information in NYC school choice setting, to incorporate limited awareness into a model of school applications while allowing for incorrect beliefs, and to identify the model using typical data from centralized school choice settings with presence of an instrument.

We also contribute to the literature by disentangling 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, Sóvágó, and Van Der Klaauw, 2021; Akbarpour, Kapor, Neilson, Van Dijk, and Zimmerman, 2022; Hahm and Park, 2022; Sartain and Barrow, 2022; Idoux, 2023; Park and Hahm, 2023). Calsamiglia, Martínez-Mora, and Miralles (2021) theoretically examine the impact of matching algorithms on segregation.

son (1989), Bresnahan and Reiss (1991), Lewbel (2000), Berry, Gandhi, and Haile (2013), and Berry and Haile (2024).

[10] See, e.g., Goeree (2008), Conlon and Mortimer (2013), Gaynor, Propper, and Seiler (2016), Hortaçsu, Madanizadeh, and Puller (2017), Abaluck and Adams-Prassl (2021), Barseghyan, Coughlin, Molinari, and Teitelbaum (2021a), Barseghyan, Molinari, and Thirkettle (2021b), and Kawaguchi, Uetake, and Watanabe (2021).

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, Huang, and Taber, 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, Neilson, and Zimmerman (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, Agarwal and Somaini (2018), Luflade (2018), and Calsamiglia, Fu, and Güell (2020) estimate preferences and potentially incorrect beliefs with observed choice data without surveys.[11] Some studies propose strategies for estimating preferences while allowing for mistaken beliefs under nontruthful mechanism (He, 2017; Hwang, 2017) and while allowing for nontruthful behavior under (approximately) truthful mechanisms (Artemov, Che, and He, 2017; Fack, Grenet, and He, 2019; Che, Hahm, and He, 2020; Larroucau and Rios, 2020; Idoux, 2023).[12] 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 dropping an unlikely school (Fack, Grenet, and He, 2019; Fabre, Larroucau, Martınez, Neilson, and Rios, 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.[13] Other studies have empirically investigated student welfare or matching stability (Narita, 2016; Abdulkadiroğlu, Agarwal, and Pathak, 2017; He, 2017; Hwang, 2017; Agarwal and Somaini, 2018; Che and Tercieux, 2019; Abdulkadiroğlu, Che, Pathak, Roth, and Tercieux, 2020; Kapor, Neilson, and Zimmerman, 2020; Calsamiglia, Fu, and Güell, 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.

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

[12] Abdulkadiroğlu, Agarwal, and Pathak (2017) and Che and Tercieux (2019) assume weak versions of the truthtelling assumption.

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

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% |
| Mean neighborhood income ($) | 58553.0 | 49469.1 | 47624.1 | 73686.9 | 54119.7 |
| Mean distance to schools (miles) | 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** | | | | | |
| 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% |
| Mean 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. See Appendix D for details about the construction of the table.
[a] 1.6% of students are multi-racial or Native American.

# 2 Overview of New York City's High School Choice

## 2.1 The Context

The NYC public high school choice system annually matches about 80,000 students to cover 700 public high school programs. The system uses the following centralized procedure:[14] (1) applicants submit rankings for up to 12 school programs; (2) school programs rank applicants based on admissions priority groups, screening policies, and/or lotteries; (3) the Student-Proposing DA algorithm is used to assign students to schools using the rankings. 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 (Section 2.2).

Characteristics of our student sample are summarized in Table 1.[15] The district has many minority students and low-income students. Of the students in the sample, 40.5%

[14] 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 B for details about NYC's matching algorithm.

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

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 Supplemental Material A.4 for our definition of *interest area*. *Uses lottery only* are the programs that use admission lotteries and neither screening nor admission priority groups.

are Hispanic, 26.9% are Black, 16.1% are Asian, and 15.0% are White.[16] 71.3% of the students are eligible for free or reduced-price lunch. The table also (partially) demonstrates the housing racial segregation in NYC.

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.

Within a school, multiple programs may be present, each featuring 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.[17] 37.2% of the programs

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

[17] 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." We observe that 4.34% of students experience deviations from this stated lexicographic rule.

exclusively use screening to break ties within a priority group. Screening can be based on various criteria, including grades, test scores, attendance, punctuality, and interviews. Other programs use lotteries to break ties, sometimes jointly with screening.[18]

## 2.2 Deferred Acceptance Algorithm: Theory and Practice

The DA algorithm has been gaining popularity, 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.[19] However, these properties do not directly address distributional outcomes such as racial integration or the equity of assignments.

Even the two desirable outcomes promised by the theoretical results, namely, truthful reporting and stability, may fail. Survey- and experiment-based evidence shows that a fraction of applicants do not truthfully report even in DA mechanisms (Chen and Sönmez, 2006; Calsamiglia, Haeringer, and Klijn, 2010; Hassidim, Marciano, Romm, and Shorrer, 2017; Rees-Jones, 2018; Hassidim, Romm, and Shorrer, 2021). Complementing these results, Ashlagi and Gonczarowski (2018) theoretically show that DA is generally not obviously strategy-proof in the sense of Li (2017); applicants with limited rationality may not understand its strategy-proofness. Stability may also fail; nontruthful reporting (Gale and Shapley, 1962; Artemov, Che, and He, 2017; Fack, Grenet, and He, 2019) or limited information about schools may undermine stability.

Furthermore, theoretically ideal versions of DA that promise strategy-proofness and stability are only occasionally implemented in practice (Abdulkadiroğlu, Pathak, and Roth, 2009; Haeringer and Klijn, 2009). In particular, the matching procedure in NYC creates incentives for the applicants to deviate from truthfully reporting their preferences, because its implementation deviates from the canonical DA 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 for whom such length constraint binds must then consider their admission chances to the schools.[20] Second, in NYC, there is

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

[19] 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).

[20] Reflecting this, the 2017 NYC High School Directory states that "[i]f you are applying to 'reach' pro-

an aftermarket that follows the main round.[21] If a student believes that she can be matched to a school in this aftermarket, she may choose not to list this school in the main round.

In addition, it is plausible that students have limited information about NYC's 763 school programs. Corcoran, Jennings, Cohodes, and Sattin-Bajaj (2018) find that providing information about schools altered the students' choices in NYC. Moreover, lower-income families may have differentially less information about high-performing schools (Sattin-Bajaj, 2016).

# 3 Evidence of Frictions and Disparities

Motivated by the aforementioned discussions, we explore empirical evidence of deviations from the theoretical aspects of the DA algorithm. We also investigate racial disparities within the context of the school choice program.

## 3.1 Data

Our main 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—we refer to a student's rank-ordered list of choices as the student's *report*—as well as school assignments, admissions priority groups, schools' screening rankings over students, and demographic information. The demographic information includes students' race, home address, subsidized lunch 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 characteristics for other students.[22] We also use some public school-level data, including those from NYC's High School Directory and School Quality Reports.

## 3.2 Evidence of Frictions

Our descriptive analysis here focuses on two types of deviations from the theory in school program application—that students are not fully aware of all the options available to them, and that students are less likely to apply to schools with lower admission chances. These are at odds with aspects of the canonical DA algorithm, namely full awareness and truthful reporting. The empirical findings will guide us in the next section when we model the choice behavior of students.

grams, be sure to include 'target' or 'likely-match' programs on your application."

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

[22] 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: Regressions 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:* *p<0.1; **p<0.05; ***p<0.01. The table reports the results from probit regressions. The *page rank / 100* variable indicates the within-borough rank of the program's school in terms of the order in which it is listed in the school directory, divided by 100. The coefficients in the Page rank / 100 row are the estimated Probit coefficients, with standard errors in parentheses. *ADE* (average difference effect) is the average increment in the predicted application 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. Standard errors are clustered at the program level. 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. A student is considered to have *applied* to a program if that program appears anywhere in their rank-ordered list. *All eligible* indicates the sample of all student-program pairs for which the student is eligible for applying to 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. Appendix D describes the controls used.

We first examine the students' awareness of schools by inspecting whether the page at which a school appears in the NYC High School Directory, which is about 600 pages long, affects the application rates to the school's programs.[23] In the directory, schools are first grouped into the five boroughs of NYC, and within each borough, they are ordered alphabetically by their names. Provided that 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 aware of all the schools.

Table 3 reports estimates from a probit model predicting whether a student applies to a program (i.e., lists it anywhere in her report), focusing on the effect of *Page Rank*, which denotes the within-borough rank of a program in terms of the order in which it is listed in the NYC's High School Directory. The top panel is based on 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 (equating to 125 pages on average) is associated with a 24.16% decrease in application rates, even after controlling for a rich set of observables, as suggested by the average difference effect (ADE) relative to the mean application rate; we define the ADE as the average increment in the (predicted) application 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 stronger negative associations for Black and Hispanic students. This could be due to Asian and White students having access to better information sources or preferable schools nearby, leading to less reliance on the directory. Our main analysis suggests that Asian and White students' consideration sets are more aligned with their preferences (Table 6). Their neighborhood schools also tend to be more selective and higher-performing (Figures 1 and A.1).

We now assess the assumption that page rank is uncorrelated with unobserved preferences. Table A.1 regresses the page rank on observable school characteristics. The $F$-statistic has a $p$-value of 0.163, indicating that page rank is largely uncorrelated with preferences as captured by observable school characteristics. Table 3 also supports the assumption. The *Near* samples 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

[23] According to Sattin-Bajaj, Jennings, Corcoran, Baker-Smith, and Hailey (2018), guidance counselors reported that the printed directory is the main source of information for the applicants.

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 group | −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. *Near and likely* sample 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. *Priority group* refers to the predicted admissions priority group; see Supplemental Material A.1. See Appendix D for the controls used and other details.

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 the *Near* samples, given that students are likely aware of these nearby schools. Our findings align closely with the latter scenario: the ADE divided by the mean application rates is either smaller or reverses sign, though not statistically significant, in the *Near* samples (Columns 7–12). 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.

Next, we turn to another phenomenon: students base their reports on admission chances. 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. 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 priority groups on whether the student lists a program; lower-priority students (i.e., those with higher numerical values of the variable *priority group*) tend not to list the program. This effect holds true irrespective of whether we account for the potential correlation of priorities with unobserved program qual-

Table 5: Regressions of Submitted Rank on Priority Group

| | *Dependent variable: rank in submitted report* | | | | | |
|---|---|---|---|---|---|---|
| No. of listed programs | 2 | 4 | 6 | 8 | 10 | 12 |
| Priority group | −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 |

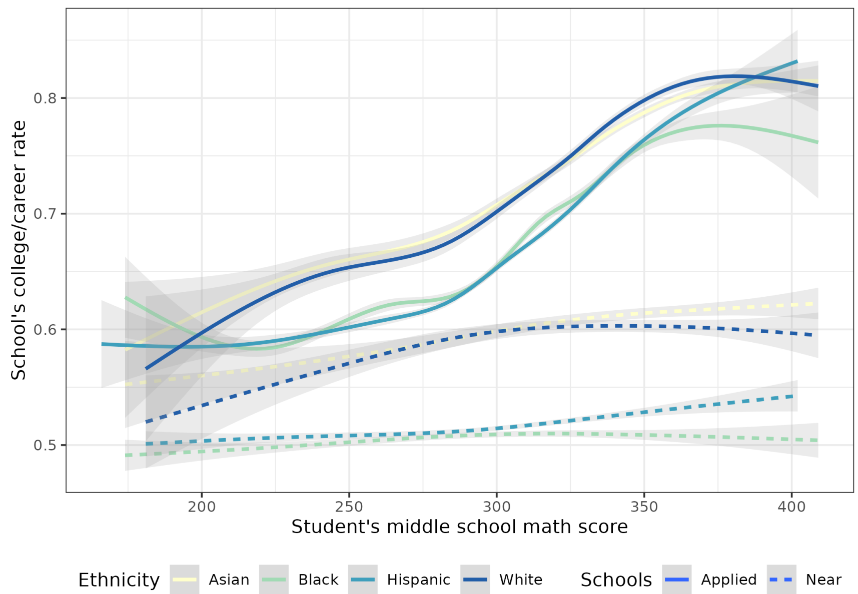
*Notes:* *p<0.1; **p<0.05; ***p<0.01. An observation refers is a student-program pair such that the student has included the program in their list, with the list's length corresponding to the number indicated in the column heading. *Priority group* refers to the predicted admissions priority group; see Supplemental Material A.1 for details. Refer to Appendix D for the controls used. Standard errors are clustered at the program level.

ity via program dummies.[24] 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, the effect persists albeit at a weaker strength relative to the mean application rates, using our preferred estimates with program dummies.

Given the 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 effects of priorities on rankings are somewhat mixed and are milder than 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

[24] We believe that the correlation between predicted priority groups and unobserved preferences of students are likely minimal even within schools. The predicted priority groups can only be based on geographic boundaries and whether the student attends the same middle school as the high school in question, and our controls include rich controls for distance to schools in addition to the indicator for whether the middle school is the same as the high school.

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.

has the opposite (and insignificant) sign. This result may be influenced in part by selection issues; the sample only includes listed programs. If a student lists a lower-priority program, which Table 4 shows is uncommon, it likely indicates a strong preference, leading her to rank it higher on her list.

Overall, the analyses suggest that students may not be aware of all the programs and admission chances may influence their application behavior. As for the latter, the students seem to factor in the chances of admission when choosing which schools to list, although the influences of the chances on their ranking behavior remains inconclusive. The main model, as we develop in Section 4, takes these descriptive findings into account, by allowing for limited consideration and incorrect beliefs.

### 3.3 Disparities in Residential Location and Choice Behavior

Now we document empirical findings regarding racial disparities both in terms of the students' proximity to and their application to better-performing schools, as illustrated in Figure 1. 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 schools.[25] In terms of the schools that applicants apply to, the racial disparities appear reduced. High-performing applicants are more likely to apply 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, as suggested by our empirical evidence. 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.

A school program is *considered* by an applicant if (1) he is aware of that 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.[26] The *consideration set* of applicant $i$, denoted by $\mathcal{C}_i$, is the set of school programs considered by applicant $i$. Consideration of school program $j$ by applicant $i$ is determined by a latent variable $c_{ij} \in (-\infty, \infty]$.[27] A school program is considered if and only if $c_{ij} > 0$. We assume that students do not consider ineligible schools.

Each applicant $i$ solves

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

[25]Figure A.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.

[26]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 two-sided matching settings, 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 (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.

[27]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.

where $j \in \{1, \ldots, J\} \equiv \mathcal{J}$ denotes a school program available through the application procedure, and $j = 0$ is the outside option.[28] Given the consideration set $\mathcal{C}_i \subseteq \mathcal{J}$, a report $r \in \mathcal{R}(\mathcal{C}_i)$ is either an empty list $\emptyset$ or an ordered list of school programs in $\mathcal{C}_i$ with length at most 12.[29] Formally, $\mathcal{R}(\mathcal{C}_i) \equiv \{\emptyset\} \cup \bigcup_{k=1}^{12} \left\{(j_1, \ldots, j_k) \in \mathcal{C}_i^k | j_m \neq j_n \text{ for } m \neq n\right\}$. Although $r$ is an ordered list, we occasionally abuse notation to 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 $r$. $p_{ij}^r \in [0, 1]$ 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 $j$, with normalization $v_{i0} = 0$. The solution to the maximization problem in Equation 4.1 is denoted by $r_i$. Multiple solutions can occur with probability zero under our assumptions and are ignored.

We model nondegenerate[30] beliefs about assignment probabilities similarly to Kapor, Neilson, and Zimmerman (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: $\mathsf{score}_{ij}$ and $\mathsf{cutoff}_{ij}$. Being a function of admissions priority groups, screening rankings, and lotteries, $\mathsf{score}_{ij}$ represents program $j$'s evaluation of applicant $i$, with a lower score denoting higher preference. One important aspect of DA is that $\mathsf{score}_{ij}$ 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 $\mathsf{type}_i$ are admitted by program $j$. In NYC, $\mathsf{type}_i$ indicates whether the student has disabilities. Separate capacities are set for each type.[31]

Under the cutoff-score representation, each student is matched to his first school program in the list for which $\mathsf{score}_{ij}$ is below $\mathsf{cutoff}_{ij}$. We model beliefs about the assignment probabilities based on this process. 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 rankings of programs in students' reports cannot affect the scores in DA, students may not recognize this property of DA. On the other hand,

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

[29]The empty list represents non-participation in the main (first) round of the application process.

[30]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.

[31]For programs using the *Educational Option* admission method, the type also depends on the applicant's reading category, determined by the middle school English Language Arts (ELA) score.

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 write the nondegenerate subjective belief, for program $j$ listed in report $r$, as

$$p^r_{ij} = \mathbb{P}_i \left( \widetilde{\mathsf{diff}}_{ij'}(k^r_{j'}) < 0 \; \forall j' \text{ s.t. } k^r_{j'} < k^r_j, \quad \widetilde{\mathsf{diff}}_{ij}(k^r_j) > 0 \right) \tag{4.2}$$

where $k^r_j$ denotes the rank of $j$ in report $r$. As mentioned above, report $r$ consists only of considered school programs, and a program is considered only if the student feels the program is reachable. Therefore, in effect, we assume $p^r_{ij} > 0$ for all $i$, $r \in \mathcal{R}(\mathcal{C}_i)$, and $j \in r$.

Finally, 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

Before presenting our empirical specification of the model, 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. Appendix C develops sufficient conditions for nonparametric identification, formalizing and extending the ideas presented here.

First, there is variation in the data that is affected only by preferences and consideration, and not by nondegenerate beliefs: (1) the number of programs in an applicant's list and (2) whether a program is listed in 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 C.4. Intuitively, if a student is not constrained by the length constraint and he considers a program (thus aware of the program and perceive it as reachable), he has no reason to drop it from his report, as long as he prefers it to the outside option. Conversely, if he does not prefer it to the outside option or does not consider it, he 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 are some school programs that are "surely" considered by an applicant or if (2) there are shifters of consideration that are excluded from utilities (Goeree, 2008). We define the *surely considered set* of applicant $i$, denoted by $\mathcal{S}_i$, as the set of programs (assumed to be) surely considered by applicant $i$. Formally, $\mathcal{S}_i \subseteq \mathcal{C}_i$ with probability 1. The following observation, a corollary of Observation 1, aids in separating preferences and consideration using 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 C.1 and C.2 in Appendix C formalize the intuition by providing sufficient conditions for nonparametrically identifying the distributions of preferences and consideration sets. 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.

In Section 7.1, we discuss how the potential selection issues—Observations 1 and 2 only utilize the students who do not exhaust all slots in the report—are resolved by an independence assumption. Propositions C.1 and C.4 show the conditions under which the selection issues regarding the exhaustion of the slots do not arise even without the independence assumption.[32]

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 the applicants for whom the list length constraint binds. These aspects of data variation are affected by beliefs in addition to preferences and consideration.

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

*(i) Suppose that an 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*

[32]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 C.4 further assumes the presence of a utility shifter that is excluded from consideration. On the other hand, case (ii) of Proposition C.5, 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.

*a function of* $\left(v_{ij}, c_{ij}, (p^r_{ij})_{r\in\mathcal{R}(\mathcal{J})}\right)_{j\in\mathcal{J}}$*. 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* $\left(v_{ij}, c_{ij}, (p^r_{ij})_{r\in\mathcal{R}(\mathcal{J})}\right)_{j\in\mathcal{J}}$*. In particular, the function is not constant in* $(p^r_{ij})_{j\in\mathcal{J},r\in\mathcal{R}(\mathcal{J})}$.[33]

In a restricted setting, Proposition C.3 outlines the conditions for nonparametric identification of beliefs. As nonparametric identifiability in a general setting is ambiguous, our 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. These distributions then determine the distribution of the counterfactual truthful reports constructed only using the 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 A.5 implements the comparison.

## 6 Empirical Specification

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

$$v_{ij} = x^v_j \beta^v_{\mathsf{eth}_i} + z^v_{ij}\alpha^v_{\mathsf{eth}_i} + \epsilon^v_{ij},$$

where $x^v_j$ denotes the vector of observed program characteristics and $z^v_{ij}$ 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 $(x^v_j, z^v_{ij})$.[34] The scale of $v_{ij}$ is normalized by setting the standard deviation of $\epsilon^v_{ij}$ equal to 1. The location is normalized by $v_{i0} = 0$. Thus, $v_{ij}$ is interpreted as the utility of $j$ relative to the outside option 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.[35] The vector $x_j$ includes, for example, college/career rate, average middle school math achievement, ethnic composition, and program interest area

[33]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 $(p^r_{ij})_{j\in\mathcal{J},r\in\mathcal{R}(\mathcal{J})}$, see the cases in Proposition C.3 and the corresponding proof.

[34]The assumption may be mild in the sense that we do not need to regard the coefficients on $(x^v_j, z^v_{ij})$ as causal in the counterfactual analyses.

[35]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.

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

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

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 $\epsilon^c_{ij} \sim_{i.i.d} N(0,1)$, 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 through observables.[36] The scale is normalized by setting the variance of $\epsilon^c_{ij}$ 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.[37] 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 600-page long NYC school directory (ranked within its borough). Because applicants may overlook the schools listed later, the page rank may shift consideration. However, as the schools follow alphabetical ordering within their boroughs, we argue that page rank is excluded from preferences.[38] We also allow the indicator of whether a program is within one mile from an applicant's middle school to affect consideration. 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 equa-

[36]Proposition C.4 and Agarwal and Somaini (2022) suggest joint distribution of $(\epsilon^v_{ij}, \epsilon^c_{ij})_j$ can be non-parametrically 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.

[37]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.

[38]We discussed how Tables 3 and A.1 are consistent with this assumption.

tion. The proxy is calculated as the difference between her objective expected scores and cutoffs; see Supplemental Material A.2 for details. Whether a student resides in the same borough as a school program could influence the student's awareness, partly 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 group. Consistent with the usage of the term in Table 4, a program is *likely* for a 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 the 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. Thus, we require further that the student be within the program's 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, Neilson, and Zimmerman, 2020).

Starting from Equation 4.2, we further assume

$$p_{ij}^r = \mathbb{P}_i\left(\widetilde{\mathsf{diff}}_{ij'}(k_{j'}^r) < 0 \ \forall j' : k_{j'}^r < k_j^r\right) \mathbb{P}_i\left(\widetilde{\mathsf{diff}}_{ij}(k_j^r) > 0\right) = \prod_{l=1}^{k-1}(1 - q_{ij_{r_l}l})q_{ijk}$$

where $q_{ijk}$ denotes $\mathbb{P}_i(\widetilde{\mathsf{diff}}_{ij}(k) > 0)$, and $j_{r_l}$ denotes the school program listed at the $l$th spot in $r$. 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, Fu, and Güell (2020) (Appendix F). This makes the computation feasible even with a vast choice set of rank-ordered reports.

Now, to parametrize $q_{ijk}$, we model $\widetilde{\mathsf{diff}}_{ij}(k)$ 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 Supplemental Material A.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 mistakenly 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 $\text{Logistic}(\delta^{\mathsf{diff}}_{ijk}, \sigma^{\mathsf{eth}_i}_{\nu})$, implying $q_{ijk} \equiv \mathbb{P}_i(\widetilde{\mathsf{diff}}_{ij}(k) > 0) = \left(1 + \exp(-\delta^{\mathsf{diff}}_{ijk}/\sigma^{\mathsf{eth}_i}_{\nu})\right)^{-1}$.

For each ethnicity, the two parameters that govern belief are $(\beta^{\mathsf{eth}_i}_{\mathsf{rank}}, \sigma^{\mathsf{eth}_i}_{\nu})$. 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 too low that he decided to exclude it from his twelve slots to list another program; the latter case only arises when the length constraint binds. If $\beta^{\mathsf{eth}_i}_{\mathsf{rank}} < 0$, on the other hand, the submitted rankings may not be truthfully ordered in terms of utilities even among the listed programs.[39] The level of $\sigma^{\mathsf{eth}_i}_{\nu}$, which governs the level of doubt the student has about his prediction, can also affect the degree of truthtelling. If $\sigma^{\mathsf{eth}_i}_{\nu} = \infty$, 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

[39] We assume students know that lower rankings cannot improve scores, ruling out $\beta^{\mathsf{eth}_i}_{\mathsf{rank}} > 0$.

option, until they run out of such programs or exhaust all the 12 slots. In this case, the ranked programs are always preferred to any considered but unranked program.

# 7 Estimated Preferences, Consideration, and Beliefs

## 7.1 Estimation

Estimation proceeds in two stages. We first estimate preference and consideration parameters using the partial likelihood of inclusion of programs in submitted reports. 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 C.1 and C.2), depends only on preference and consideration parameters, excluding belief parameters.[40] This likelihood uses a sample of student-program pairs that meet a specific criterion ($|r_i \backslash \{j\}| < 11$), which implies the criterion in Observations 1 and 2 ($|r_i| < 12$), to address selection issues. Assuming $(\epsilon^v_{ij}, \epsilon^c_{ij})$ is $i.i.d$ across $j$,[41] we establish the distribution of the unobservables $(\epsilon^v_{ij}, \epsilon^c_{ij})_{j\in\mathcal{J}}$ is independent of our selection criterion (Lemma E.1). Appendix E.1 delineates the partial likelihood and shows that the true parameters maximize it. We randomly sample 4,000 students per ethnicity to facilitate estimation.[42]

In the second stage, belief parameters $(\beta^{\mathsf{eth}}_{\mathsf{rank}}, \sigma^{\mathsf{eth}}_{\nu})_{\mathsf{eth}}$ are estimated using the Generalized Method of Moments, taking as given the first-stage estimates for preference and consideration.[43] 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 C.3. The exact moment conditions are provided in

[40] A part of this partial likelihood depends solely on preference parameters, following Observation 2. See Appendix E for details.

[41] This assumption excludes a random coefficient model, but this may not be overly restrictive (Pathak and Shi, 2020).

[42] 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. 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.

[43] The moment conditions also contain information about preference and consideration parameters. Hence, joint estimation of all the parameters using the scores of the first-stage likelihood stacked with the moments here would be more efficient. For computational tractability, we proceed in two stages.

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% |

Appendix E.2.

### 7.2 Estimates

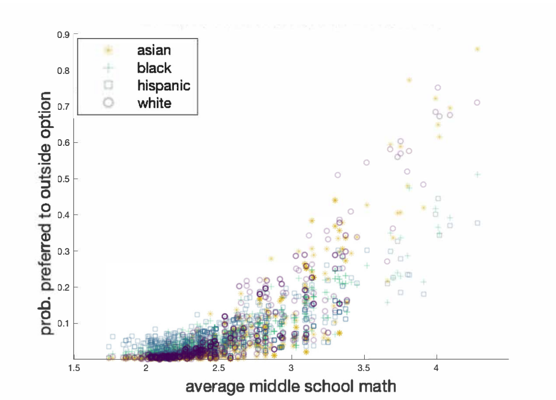
**Preference and Consideration** In Table 6, we provide a summary of the key features of the estimated parameters (raw parameter estimates are shown in Table A.2). Students across all ethnicities are estimated to consider approximately 10.6% of programs on average. White students consider the smallest proportion of schools, potentially because 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.[44]

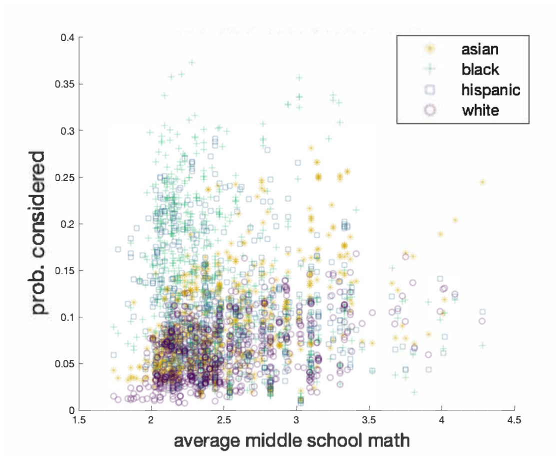
Our findings reveal that Asian and White students have stronger preferences for more selective programs, represented by the average middle school math proficiency of incoming students, compared to Black and Hispanic students. Although this trend might simply be

[44] Specifically, we use extrapolated values of $x_j^c \beta^c_{\mathsf{eth}_i} + z_{ij}^c \alpha^c_{\mathsf{eth}_i} + \epsilon^c_{ij}$ (which equals $c_{ij}$ for the not surely considered programs) even for the surely considered programs. The 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 $\hat{v}_{ij} = \tilde{v}_{ij} - \epsilon^v_{ij}$.

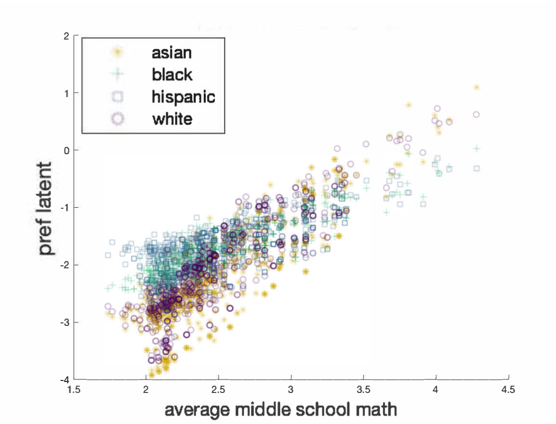
Figure 2: Probability and Latent Values for Preference and Consideration



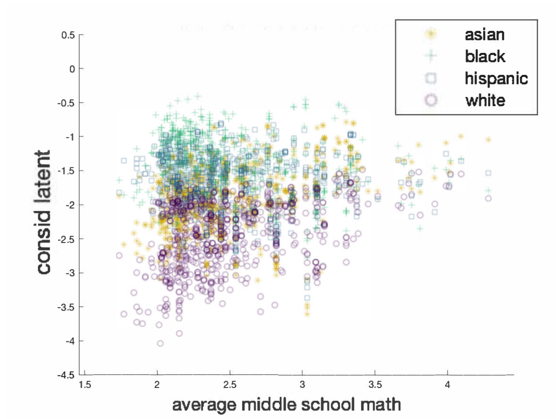

(a) Probability of Being Preferred to the Outside Option



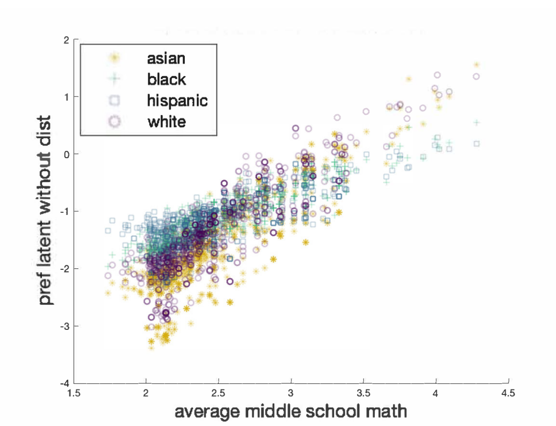

(b) Probability of Being Considered



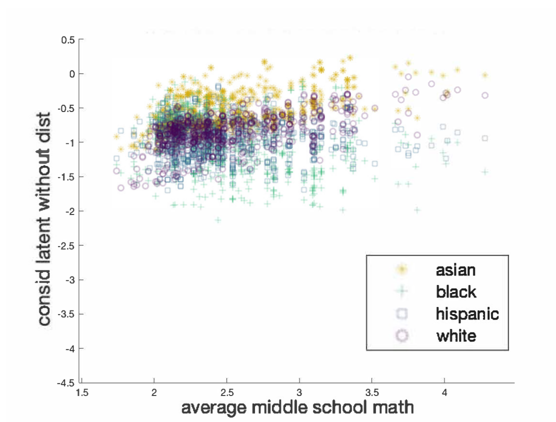

(c) Mean Latent Values for Preference




(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. 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 (see footnote 44 for details). Figures 2e and 2f show the latent values when the distance to school is set to zero.

mirroring the geographical distribution of less selective schools, which tend to be located further from Asian and White students' homes (Figure A.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 when school selectivity is replaced with school performance gauged through college/career rates (Figure A.2) or when we estimate preferences using only the surely considered programs (Figure A.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. This pattern emerges partly because Asian and White students typically live farther from less selective programs and because the distance to schools is an important determinant of consideration, especially for these groups (Table A.2). 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 A.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 A.2). 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 7 indicates that students tend to truthtell in both scenarios. The fractions represent how many of the simulated subjectively optimal reports from our estimated model exactly match the simulated truthful-among-considered reports, where a report is said to be *truthful among considered* if the considered programs are ranked truthfully according to the utilities until no more program is preferred to the outside option or all 12 slots are filled.[45] Notably, even when the length constraint binds, the subjectively optimal reports approximate the truthful-among-considered reports. Section 8.3 further explores the implication for the truthfulness of reports among all eligible (not just considered) programs. Figure A.5 performs a diagnostic com-

[45]Such a report may still skip some programs deemed unreachable or unknown to the student.

Table 7: 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:* The bottom panel tabulates subjectively optimal reports that list the same set of programs as truthful-among-considered reports, ignoring the ordering.

parison of the slopes of mean utilities against rank across different reports as discussed in Section 5.[46]

Since we do not accommodate individual heterogeneity in truthtelling attitude within race,[47] 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 Policy

## 8.1 Impacts of School Choice: A Decomposition Analysis

**Effects on Racial Integration** We find that NYC's school choice slightly promotes racial integration relative to neighborhood matching. Our analyses also reveal that student preferences, net of the confounding effects from limited information and potential nontruthful behavior, contribute to integration. 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. The indices under school choice matchings—Actual and Estimated—are similar to or lower than those under Neighborhood matching. We then sequentially shut off each channel as described

[46] The downward trend in mean utility along the subjectively optimal reports well approximates that of the observed reports from the data. The subjectively optimal reports, in turn, are almost indistinguishable from the truthful-among-considered reports, reflecting that the belief parameters are in line with truthtelling among considered programs.

[47] The error term $\nu_{ij}$ 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 $q_{ijk}$.

Table 8: 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 (Supplemental Material B.1). *Actual* matching refers to the actual school choice matching in 2017 from the main round of DA. See Appendix D for other details about the implementation of the matchings in the table.

in Table 8. Changing the estimated beliefs to Truthful among Considered does not lead to significant changes, which is natural given our belief estimates are 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.

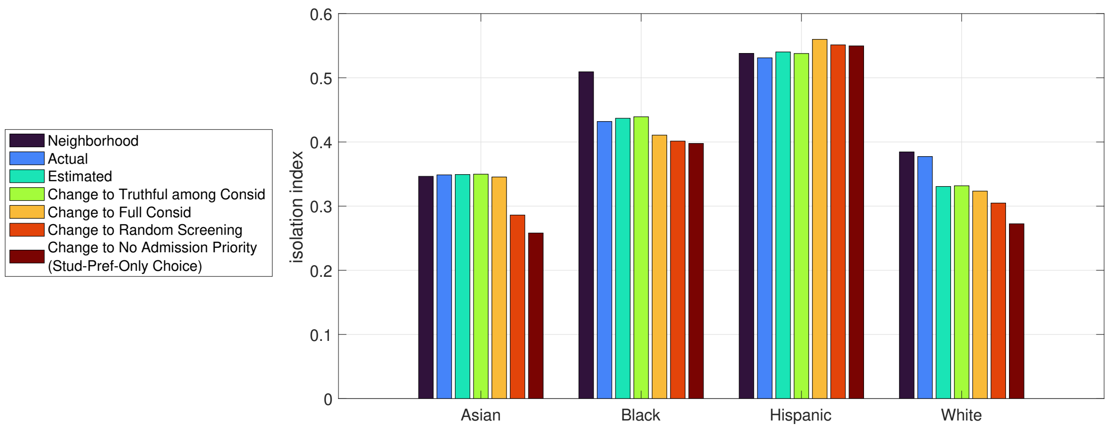
While we observed both similarities and differences in preferences across races (e.g., in Figure 2), overall, student preferences work to integrate races. This is evident when we compare Student-Preference-Only School Choice allocation—the DA with fully-informed students making truthful reports and programs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

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

Each bar in Figure 4 depicts the fraction of students who are matched to one of their top five[48] preferred programs based on their utilities. These are the top five programs they

[48] Figure A.7 presents the equivalent figure for top ten preferred programs.

Figure 3: Isolation Indices by Matching—Decomposition




*Notes:* Each bar represents isolation index of an ethnic group in a matching. See Table 8 for the definitions of the matchings.

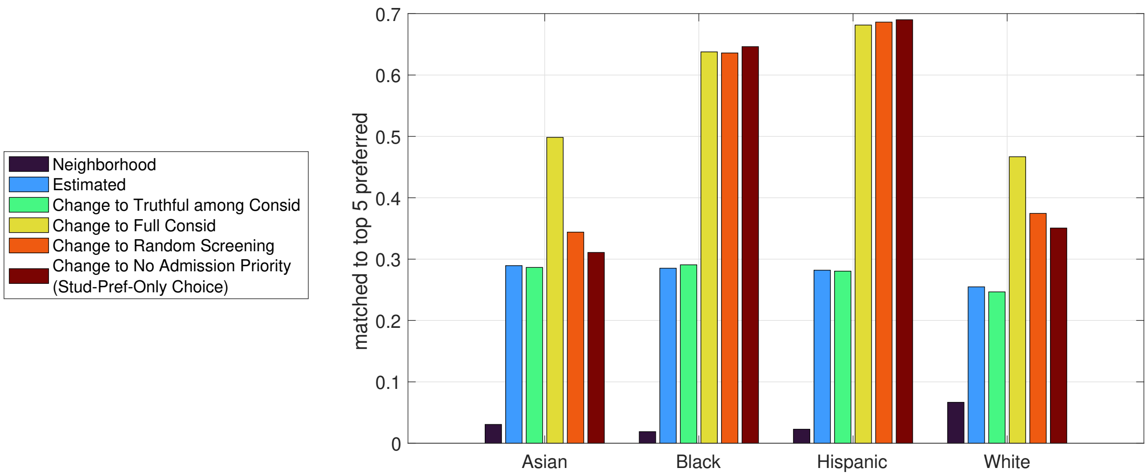
would have preferred the most if they considered all eligible programs. First, focusing on Neighborhood matching, we see that only a small fraction of the students are placed in their top five preferred programs. White students are the most likely to be matched to one of their preferred programs in this 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.

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 are conducive to matching 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 screening programs. We also see that admissions priorities act to place Asian and White students in their preferred programs. This may reflect that most 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 Personalized School Recommendations: A Counterfactual Analysis

Although the preceding section showed that substantial welfare gains can arise when students consider *all* eligible programs, this is unlikely in practice. As such, we assess various

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 based on utilities, accounting for all eligible programs, whether considered or not. See Table 8 and the discussions for the definitions of the matchings.

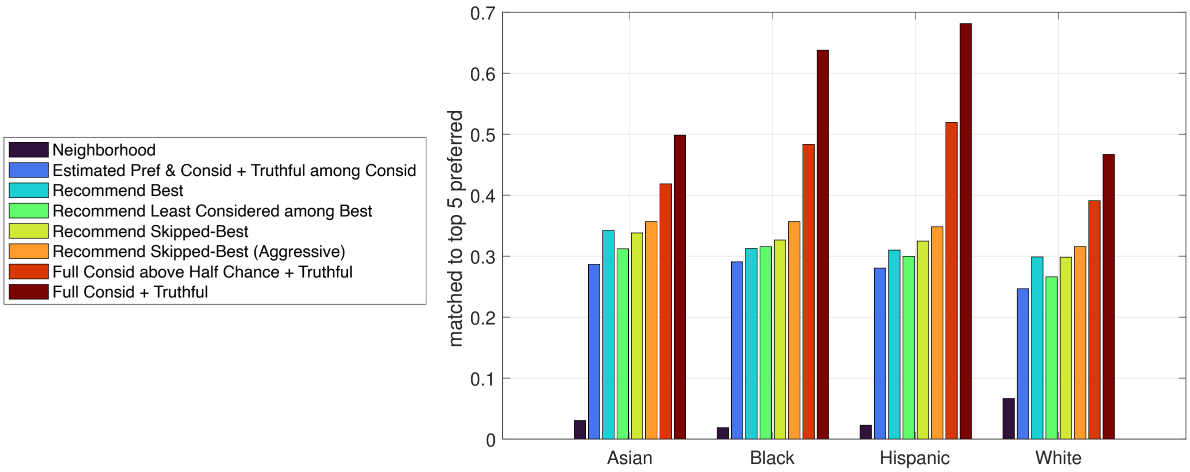
feasible interventions involving personalized school recommendations. A judicious use of both the preference and consideration estimates turns out to be useful for designing effective interventions.

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 with objective admission chances exceeding 50%,[49] 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 the 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.
- *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, unlike the three other interventions, it also recommends programs with objective admission

[49] 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.

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. Refer to the text for definitions of the matchings.

chance below 50%.

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 our estimates that indicate students are essentially truthfully reporting among the considered programs, we impose such truthful reporting in this subsection to facilitate the computation.[50]

Figure 5 summarizes the results.[51] 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[52] and Full Consideration + Truthful matching.[53] This is an encouraging result, recognizing that we recommended only 30 out of approximately 750 programs.

Notably, both preference and consideration estimates are useful for designing interventions. The highest-performing intervention, Aggressive Skipped Best, employs both pref-

[50]After intervention, simulating subjectively optimal reports takes longer due to enlarged consideration sets.

[51]Figure A.8 shows the impacts on isolation indices and the proportions matched to the top ten preferred programs. The results on matchings to the top ten preferred programs are qualitatively similar to the results here. Aggressive Skipped Best intervention slightly increases isolation indices.

[52]This is the Change to Truthful Among Considered matching in Table 8.

[53]This is the Change to Full Consideration matching in Table 8. In comparison, Full Consideration above Half Chance + Truthful matching represents the counterfactual scenario where the students are considering all eligible programs with objective admission chances exceeding 50%.

Table 9: 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 $\geq 5$ 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 8) was used.

erence 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.

### 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.[54] We say that a student *views* an eligible program with *justified envy* if the student and the program 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 are determined by the estimated utilities, which are defined regardless of whether the programs were considered or not.[55]

Table 9 shows that 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.

[54]This is true if each program has a responsive preference, 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 $\mathsf{score}_{ij}$. 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 19 defines *individual irrationality* and *blocking pairs*.

[55]Programs' preferences are determined by the expected scores $E_{\text{obj}}[\mathsf{score}_{ij}]$, which is a function of the admissions priority groups and, for screening programs, the expected screening ranking for each applicant. Lotteries have the same distribution across applicants and does not affect the expected scores. See Supplemental Material A.2 for the definition of expected scores.

**Truthful Reporting** As we have seen in Table 7, students typically rank their *considered* programs truthfully based on their utilities. Closer inspection allows us to discuss truthful reporting among all (eligible)—not just considered—programs.

There are two potential types of deviations 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 programs from the length-constraint-binding lists due to a subjectively low (albeit nondegenerate) probability of admission. The middle panel of Table 7 shows that students seldom play the first type of deviation: reversals. Note that the reversals can occur only among considered programs by design. Hence, the middle panel's figures are also 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 definitively distinguish between the two reasons for not considering a program, our findings indicate that subjective assessments of reachability is also important.[56] In Section 7.2, we have also 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 the researcher can specify, for each student, a set of school programs surely considered by

[56]For instance, Table 4 showed that admissions priority is a key determinant of inclusion of a program in the submitted report and Table A.2 showed that objective admission probability positively affects consideration chances.

the student—e.g., highly likely high school programs near their home or middle school. We have also developed nonparametric identification results that further clarify the sources of identification.

Compared to neighborhood-based school program allocation, school choice slightly improves racial integration and markedly boosts the number of students matched to their preferred schools across all races. We find that admissions priorities and screening policies tend to segregate races. We also show 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 an essentially truthful manner.

Some key aspects highlighted in our paper align with the NYC DOE's recent policies after our analysis. Some NYC DOE schools adopted "Diversity in Admissions" policies, which prioritize admissions for students of lower socioeconomic status and English Language Learner students. 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.

## References


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

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

Atila Abdulkadiroğlu, Parag A. Pathak, and Alvin E. Roth. Strategy-Proofness Versus Efficiency in Matching With Indifferences: Redesigning the NYC High School Match. *American Economic Review*, 99(5):1954–1978, December 2009. ISSN 0002-8282. doi: 10.1257/aer.99.5.1954. URL https://pubs.aeaweb.org/doi/10.1257/aer.99.5.1954.

Atila Abdulkadiroğlu, Nikhil Agarwal, and Parag A. Pathak. The Welfare Effects of Coordinated Assignment: Evidence from the New York City High School Match. *American Economic Review*, 107(12):3635–3689, December 2017. ISSN 0002-8282. doi: 10.1257/aer.20151425. URL https://pubs.aeaweb.org/doi/10.1257/aer.20151425.

Atila Abdulkadiroğlu, Yeon-Koo Che, Parag A Pathak, Alvin E Roth, and Olivier Tercieux.

Efficiency, justified envy, and incentives in priority-based matching. *American Economic Review: Insights*, 2(4):425–442, 2020.

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

Nikhil Agarwal and Paulo J Somaini. Demand analysis under latent choice constraints. *NBER Working Paper*, number 29993, 2022.

Patrick Agte, Claudia Allende, Adam Kapor, Christopher Neilson, and Fernando Ochoa. Search and biased beliefs in education markets. *Working Paper*, 2024.

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

Kehinde Ajayi and Modibo Sidibe. School choice under imperfect information. *Working Paper*, 2022.

Kehinde F. Ajayi, Willa H. Friedman, and Adrienne M. Lucas. The Importance of Information Targeting for School Choice. *American Economic Review*, 107(5):638–643, May 2017. ISSN 0002-8282. doi: 10.1257/aer.p20171131. URL https://pubs.aeaweb.org/doi/10.1257/aer.p20171131.

Mohammad Akbarpour, Adam Kapor, Christopher Neilson, Winnie Van Dijk, and Seth Zimmerman. Centralized school choice with unequal outside options. *Journal of Public Economics*, 210:104644, 2022.

Claudia Allende, Francisco Gallego, and Christopher Neilson. Approximating the equilibrium effects of informed school choice. *Working Paper*, 2019.

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

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

Georgy Artemov, Yeon-Koo Che, and Yinghua He. Strategic 'Mistakes': Implications for Market Design Research. *Working Paper*, 2017.

Itai Ashlagi and Yannai A. Gonczarowski. Stable Matching Mechanisms Are Not Obviously Strategy-Proof. *Journal of Economic Theory*, 177:405–425, September 2018. ISSN 00220531. doi: 10.1016/j.jet.2018.07.001. URL https://linkinghub.elsevier.com/retrieve/pii/S0022053118303454.

Christopher Avery and Parag A. Pathak. The distributional consequences of public school choice. *American Economic Review*, 111(1):129–52, January 2021. doi: 10.1257/aer.20151147. URL https://www.aeaweb.org/articles?id=10.1257/aer.20151147.

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

Levon Barseghyan, Maura Coughlin, Francesca Molinari, and Joshua C Teitelbaum. Heterogeneous choice sets and preferences. *Econometrica*, 89(5):2015–2048, 2021a.

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

Steven Berry, Amit Gandhi, and Philip Haile. Connected Substitutes and Invertibility of Demand. *Econometrica*, 81(5):2087–2111, 2013.

Steven T Berry and Philip A Haile. Nonparametric identification of differentiated products demand using micro data. *Econometrica*, 92(4):1135–1162, 2024.

Robert Bifulco and Helen F Ladd. 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, 2007.

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

Caterina Calsamiglia, Guillaume Haeringer, and Flip Klijn. Constrained School Choice: An Experimental Study. *The American Economic Review*, 100(4):1860–1874, 2010. URL http://www.jstor.org/stable/27871278.

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

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

Christopher Campos. Social interactions, information, and preferences for schools: Experimental evidence from los angeles. *Working Paper*, 2024.

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

Yeon-Koo Che, Dong Woo Hahm, and Yinghua He. Leveraging uncertainties to infer preferences: Robust analysis of school choice. *Working Paper*, 2020.

Yan Chen and Tayfun Sönmez. School Choice: An Experimental Study. *Journal of Economic Theory*, 127(1):202–231, March 2006. ISSN 00220531. doi: 10.1016/j.jet.2004.10.006. URL https://linkinghub.elsevier.com/retrieve/pii/S0022053104002418.

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

Sean P Corcoran, Jennifer L Jennings, Sarah R Cohodes, and Carolyn Sattin-Bajaj. Leveling the Playing Field for High School Choice: Results From a Field Experiment of Informational Interventions. *NBER Working Paper*, number 24471, 2018.

Susan Dynarski, CJ Libassi, Katherine Michelmore, and Stephanie Owen. 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, 2021.

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

Anaïs Fabre, Tomás Larroucau, Manuel Martınez, Christopher Neilson, and Ignacio Rios. Application mistakes and information frictions in college admissions. *Working Paper*, 2021.

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

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

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

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

Michelle Sovinsky Goeree. Limited Information and Advertising in the U.S. Personal Computer Industry. *Econometrica*, 76(5):1017–1074, 2008. URL http://www.jstor.org/stable/40056493.

Guillaume Haeringer and Flip Klijn. Constrained School Choice. *Journal of Economic Theory*, 144(5):1921–1947, September 2009. ISSN 00220531. doi: 10.1016/j.jet.2009.05.002. URL https://linkinghub.elsevier.com/retrieve/pii/S002205310900057X.

Dong Woo Hahm and Minseon Park. 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, 2022.

Avinatan Hassidim, Déborah Marciano, Assaf Romm, and Ran I. Shorrer. The Mechanism Is Truthful, Why Aren't You? *American Economic Review*, 107(5):220–224, May 2017. ISSN 0002-8282. doi: 10.1257/aer.p20171027. URL https://pubs.aeaweb.org/doi/10.1257/aer.p20171027.

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

Justine S Hastings and Jeffrey M Weinstein. Information, School Choice, and Academic

Achievement: Evidence From Two Experiments. *The Quarterly Journal of Economics*, 123(4):1373–1414, 2008.

Yinghua He. Gaming the Boston School Choice Mechanism in Beijing. *Working Paper*, 2017.

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

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

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

Chang-Tai Hsieh and Miguel Urquiola. 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, 2006.

Sam Il Myoung Hwang. How Does Heterogeneity in Beliefs Affect Students in the Boston Mechanism. *Working paper*, 2017.

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

Adam J. Kapor, Christopher A. Neilson, and Seth D. Zimmerman. Heterogeneous Beliefs and School Choice Mechanisms. *American Economic Review*, 110(5):1274–1315, may 2020. doi: 10.1257/aer.20170129. URL https://doi.org/10.1257%2Faer.20170129.

Kohei Kawaguchi, Kosuke Uetake, and Yasutora Watanabe. Designing context-based marketing: Product recommendations under time pressure. *Management Science*, 67(9):5642–5659, 2021.

Dany Kessel and Elisabet Olme. School Choice Priority Structures and School Segregation. *Working paper*, 2018.

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

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

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

Arthur Lewbel. Semiparametric Qualitative Response Model Estimation With Unknown Heteroscedasticity or Instrumental Variables. *Journal of Econometrics*, 97(1):145–177, July 2000. ISSN 03044076. doi: 10.1016/S0304-4076(00)00015-4. URL https://linkinghub.elsevier.com/retrieve/pii/S0304407600000154.

Shengwu Li. Obviously Strategy-Proof Mechanisms. *American Economic Review*, 107(11): 3257–3287, November 2017. ISSN 0002-8282. doi: 10.1257/aer.20160425. URL https://pubs.aeaweb.org/doi/10.1257/aer.20160425.

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

Charles F Manski. The Structure of Random Utility Models. *Theory and Decision*, 8(3): 229, 1977.

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

Daniel McFadden. A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration. *Econometrica*, 57(5):995, September 1989. ISSN 00129682. doi: 10.2307/1913621. URL https://www.jstor.org/stable/1913621?origin=crossref.

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

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

Hessel Oosterbeek, Sándor Sóvágó, and Bas Van Der Klaauw. Preference heterogeneity and school segregation. *Journal of Public Economics*, 197:104400, May 2021. ISSN 00472727. doi: 10.1016/j.jpubeco.2021.104400.

Minseon Park and Dong Woo Hahm. Location choice, commuting, and school choice. *Working Paper*, 2023.

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

Alex Rees-Jones. Suboptimal Behavior in Strategy-Proof Mechanisms: Evidence From the Residency Match. *Games and Economic Behavior*, 108:317–330, March 2018. ISSN 08998256. doi: 10.1016/j.geb.2017.04.011. URL https://linkinghub.elsevier.com/retrieve/pii/S0899825617300751.

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

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

Lauren Sartain and Lisa Barrow. The Pathway to Enrolling in a High-Performance High School: Understanding Barriers to Access. *Education Finance and Policy*, 17(3):379–407, July 2022. ISSN 1557-3060. doi: 10.1162/edfp_a_00349.

Carolyn Sattin-Bajaj. *Unaccompanied Minors: Immigrant Youth, School Choice, and the*

*Pursuit of Equity*. Harvard Education Press, September 2016. ISBN 978-1-61250-710-1. Publication Title: Harvard Education Press.

Carolyn Sattin-Bajaj, Jennifer L. Jennings, Sean P. Corcoran, Elizabeth Christine Baker-Smith, and Chantal Hailey. Surviving at the Street Level: How Counselors' Implementation of School Choice Policy Shapes Students' High School Destinations. *Sociology of Education*, 91(1):46–71, January 2018. ISSN 0038-0407, 1939-8573. doi: 10.1177/0038040717751443. URL http://journals.sagepub.com/doi/10.1177/0038040717751443.

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

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

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

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

# Online Appendix for
# Distributional Impacts of Centralized School Choice

Jaewon Lee     Suk Joon Son

August 17, 2024

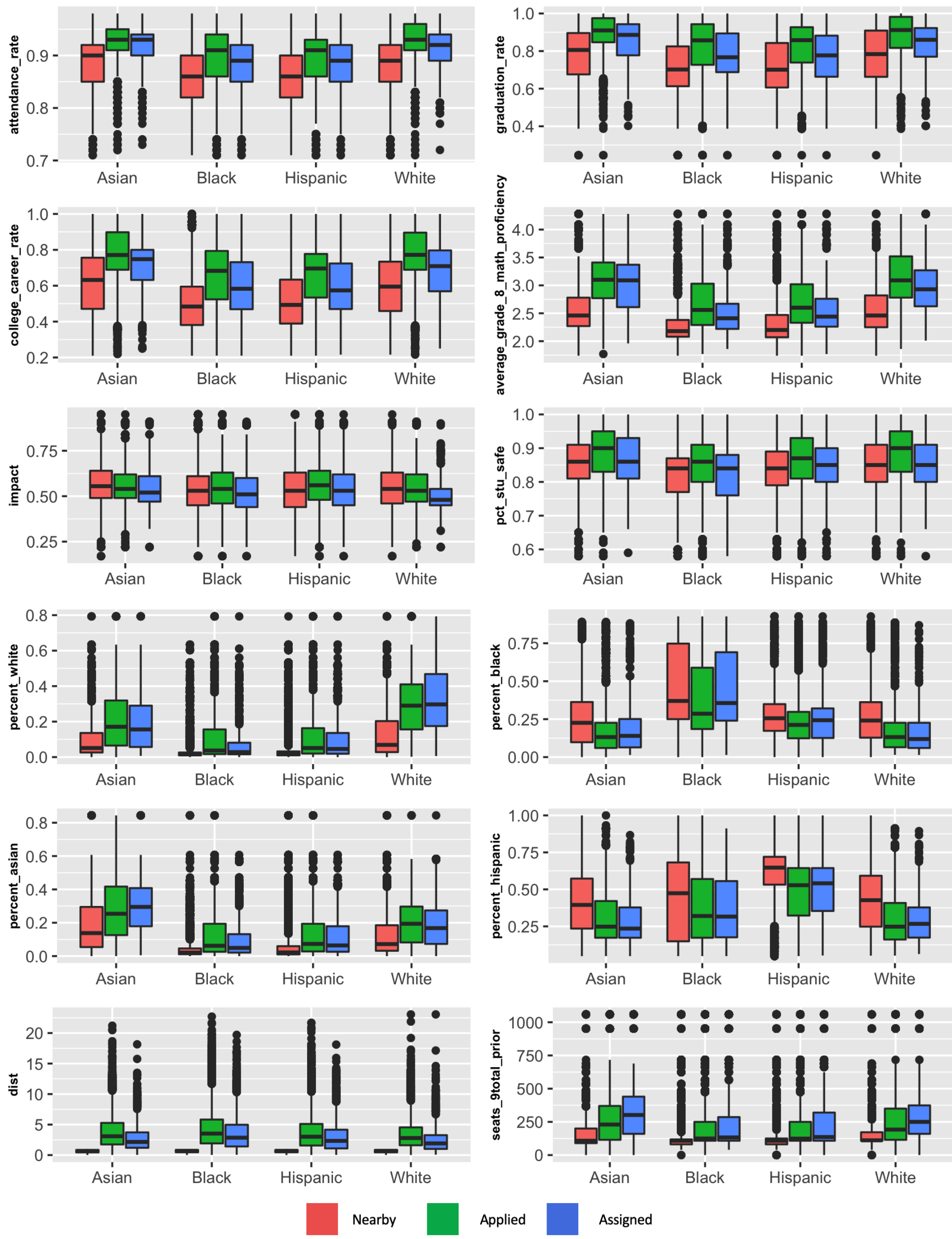
## A Additional Tables and Figures

Table A.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 A.1: Schools Nearby, Applied to, and Matched, by Ethnicity



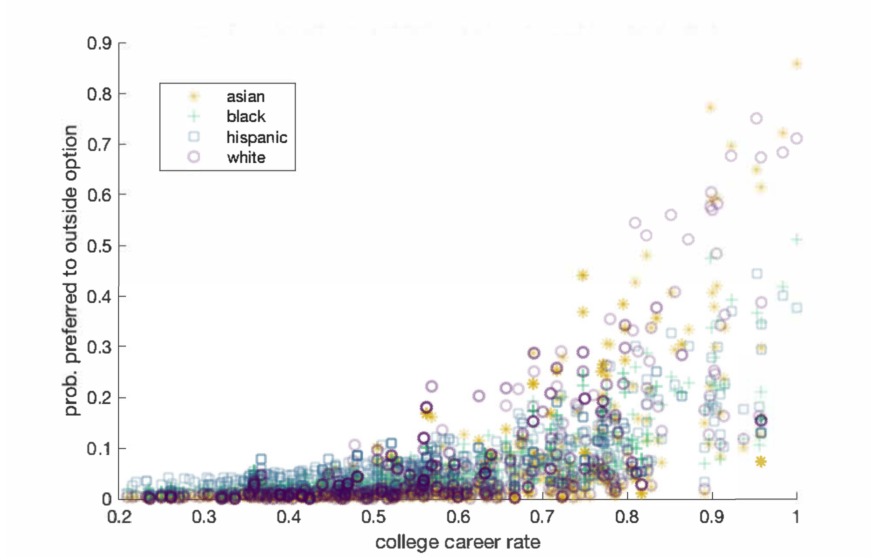

*Notes:* Nearby schools are the schools within one mile from student's home. The applied and assigned schools are from the main round of applications. *Pct_stu_safe* denotes the proportion of students who have reported that they feel safe in the school. *College_career_rate* indicates the proportion 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.

Table A.2: Estimated Parameters by Race

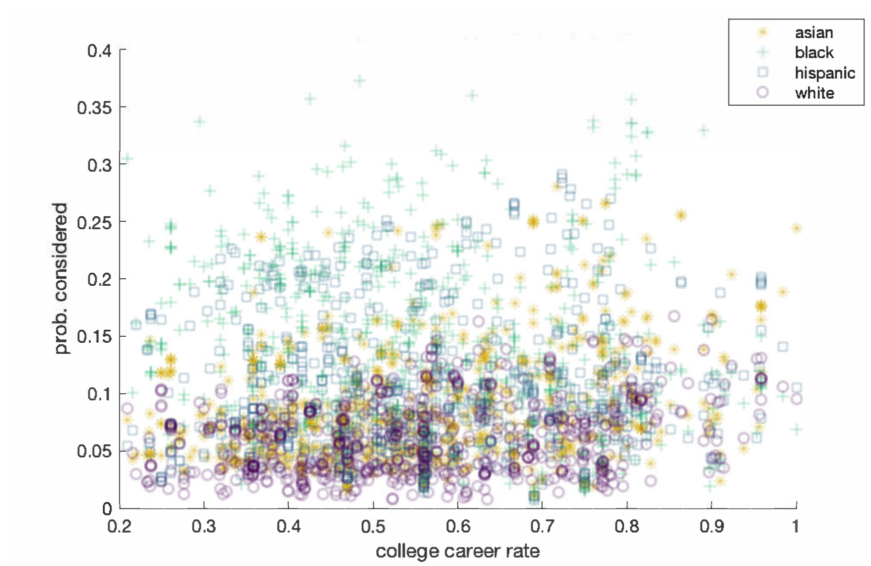
| Parameter | Asian | | Black | | Hispanic | | White | |
|---|---|---|---|---|---|---|---|---|
| *Preference* | | | | | | | | |
| Subsidized lunch | 0.13 | (0.01) | 0.17 | (0.01) | −0.01 | (0.01) | 0.20 | (0.01) |
| Lives in Brooklyn | −0.51 | (0.02) | −0.29 | (0.02) | 0.00 | (0.02) | −0.52 | (0.02) |
| 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 $\epsilon_{ij}^{v}$ | 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 $\epsilon_{ij}^{c}$ | 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.

Figure A.2: Preference and Consideration and School Performance



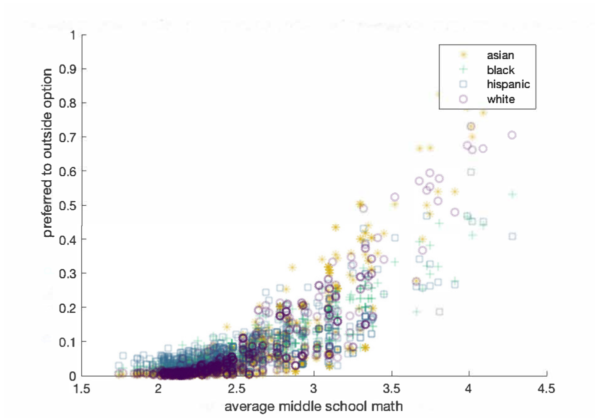

(a) Probability of Being Preferred to the Outside Option



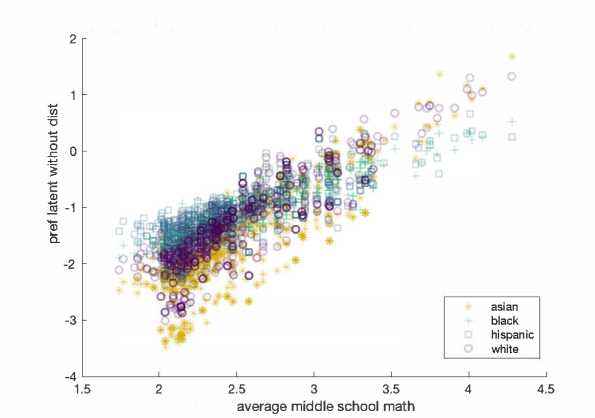

(b) Probability of Being Considered

*Notes*: For each ethnicity, each point in the scatter plot denotes a program. Figures (a) and (b) respectively depicts the within-race average probability of a school being preferred to the outside option and being considered. *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.

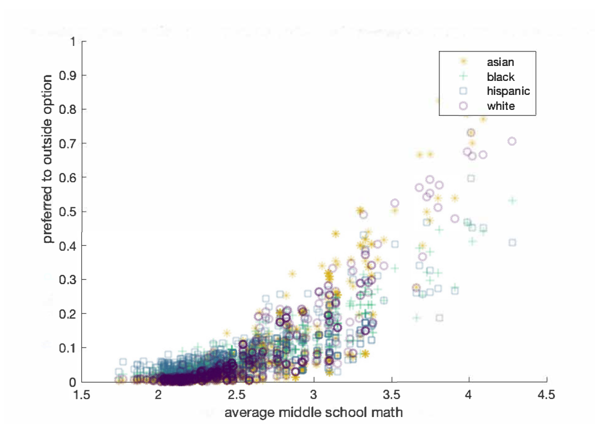
Figure A.3: Summary of Preference Estimates Using Only Surely Considered Programs




(a) Probability of Being Preferred to the Outside Option



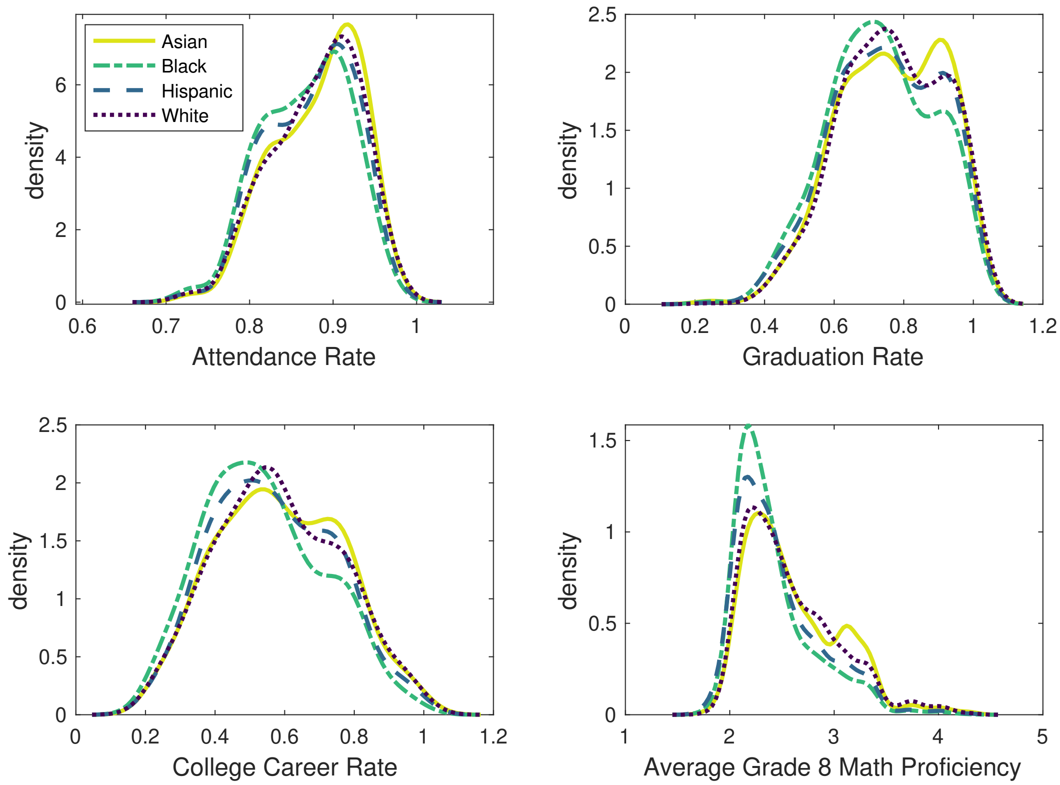

(b) Mean Latent Values for Preference




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

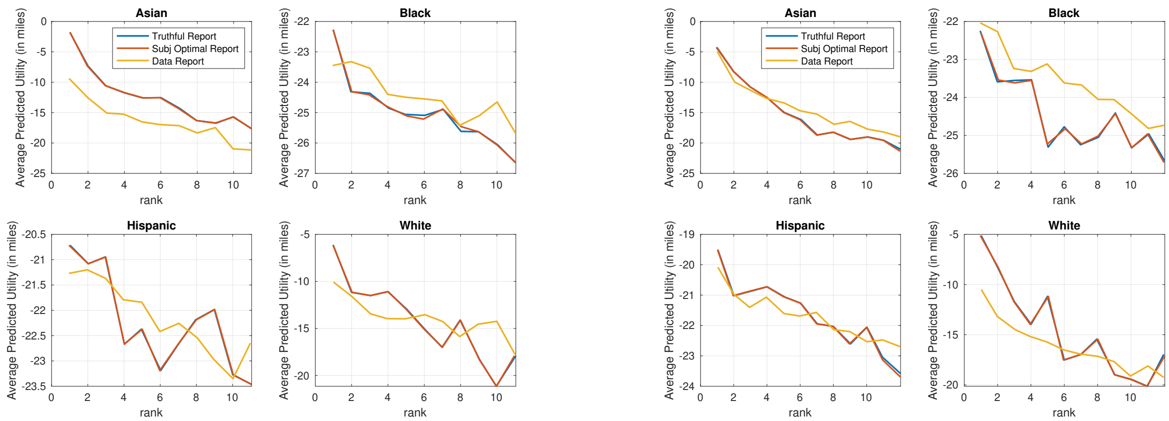
*Notes*: For each ethnicity, each point in the scatter plot denotes a program. Figure (a) depicts the within-race average probability of a school being preferred to the outside option. Figure (b) presents the within-race average predicted latent values of $\tilde{c}_{ij}$, where $\tilde{c}_{ij}$ adjusts $c_{ij}$ for the fact that sure consideration implies $c_{ij} = \infty$ by construction (see footnote 44). Figure (c) shows the latent values when the distance to school is set to zero. The estimates are obtained using only the surely considered programs, i.e., by maximizing the second sum in the partial log-likelihood expression E.1.

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

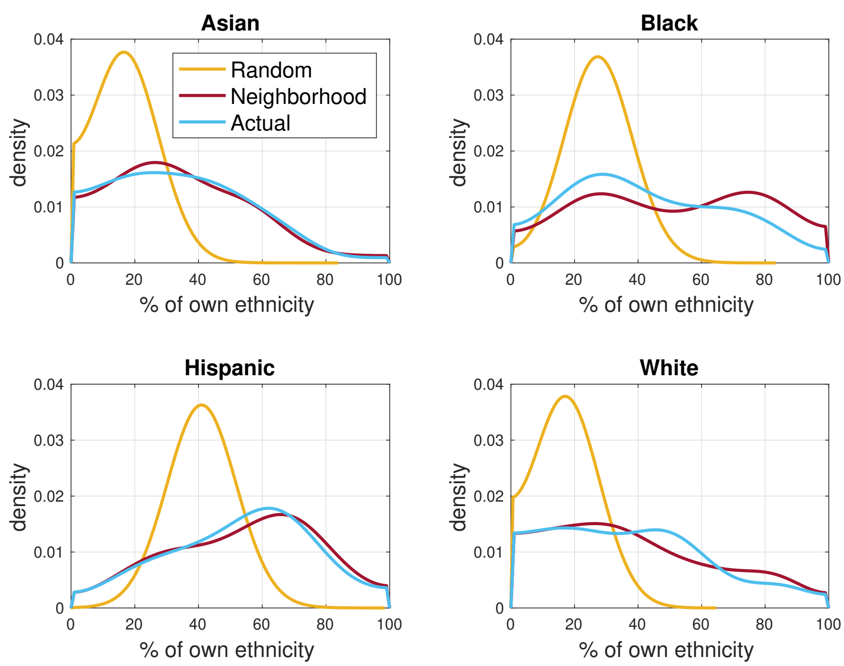



(a) Among 11-program reports

(b) Among 12-program (full) reports

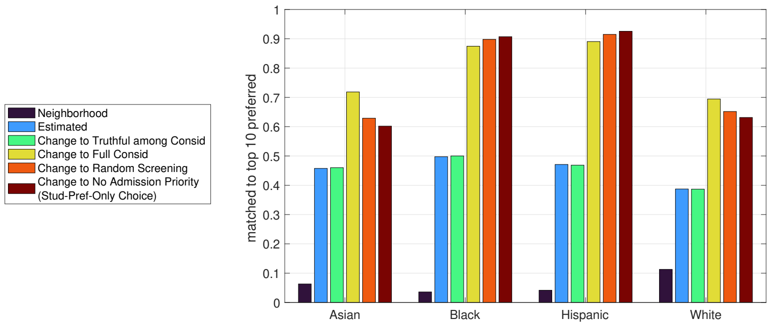
*Notes*: The figures overlay the plots of mean utilities against the rank at which the program is listed. The mean utilities reflect the within-race average of predicted utilities (i.e., net of $\epsilon_{ij}^{v}$) normalized by the coefficient on the race-specific coefficient on distance.

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



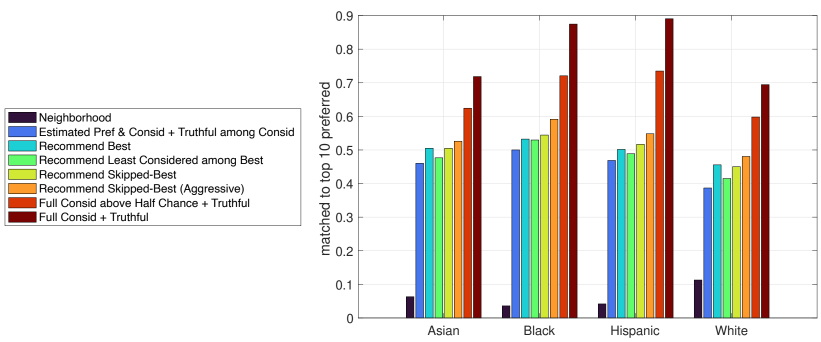

*Notes:* For each ethnicity and matching, the plot represents the kernel-smoothed density of the proportion of students with the same ethnicity in the students' assigned programs. The kernel density estimation uses Gaussian kernel with bandwidth 10, and is boundary corrected. See Table 8 for the definitions of the matchings.

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

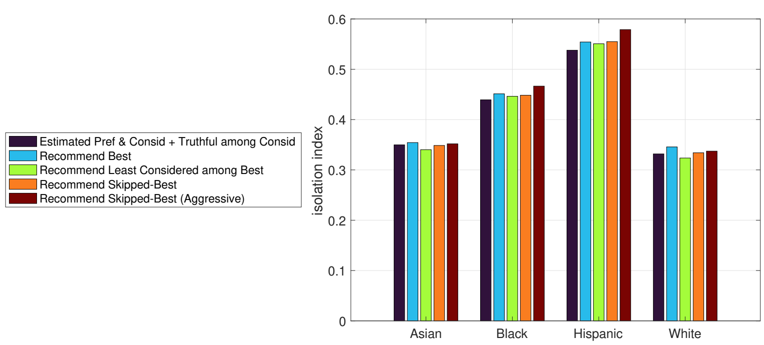



*Notes:* Each bar represents the fraction of the students matched to their top ten preferred programs. The sample includes both the programs that are considered and those that are not. See Table 8 and the discussions for the definitions of the matchings.

Figure A.8: Impacts of Information Interventions—Top Ten Preferred Programs 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.

# B 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.

- *Step* $k \geq 2$*:* 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 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.

# C Identification: Details

## C.1 Nonparametric Identification

In this section, we provide sufficient conditions for the nonparametric identification of the model. The main results are provided here, and Appendix C.2 provides additional results under stronger and weaker sets of assumptions. Proofs are in Appendix C.4.

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 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.[1]

**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[2] by any exogenous $(i,j)-$level observables, including distance to school, and the interactions between school characteristics and the student-level observables.[3]

We first establish the nonparametric identifiability of preference. Proposition C.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 C.1** (Identification of preferences)**.** *Suppose that we observe:*

*(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 $z_i \backslash (z_i^v, a_i)$.*

[1]To complement our main result, Proposition C.5 only assumes the presence of surely considered sets.

[2]We conjecture that results in Berry and Haile (2024) may be used to formally show how different variables can form an index that mimics the role of a special regressor.

[3]Note that most results—except case (ii) of Proposition C.4, which uses identification-at-infinity argument—can be extended to allow for limited support of the special regressor at the cost of identifying the distribution of the utilities or the latent variables for consideration on limited support.

*Then, the joint distribution of utilities conditional on observables,* $\mathbb{P}\big(v_i \le v \big| z_i\big)$, *is identified for almost all* $(v, z_i) \in \operatorname{supp}(v_i, z_i)$.[4]

All proofs are in Appendix C.4. 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 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 C.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 C.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.[5] 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 C.2** (Identification of consideration). *Suppose that* $\mathbb{P}\big(v_i \le v | z_i = z\big)$ *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,* $\mathbb{P}\big(c^*_i \le c^* \big| z_i\big)$, *is identified for almost all* $(c^*, z_i) \in \operatorname{supp}(c^*_i, z_i)$.[6]

*(ii) if* $c_i$ *is not independent of* $v_i$ *conditional on* $z_i$, $\mathbb{P}\big((c^*_{ij})_{j\in\mathcal{A}} \le c^* \big| (v_{ij})_{j\in\mathcal{A}} > 0, z_i\big)$ *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 C.1, it is allowed that $a_i = z^c_i$ or $z^c_i = z^v_i$.[7]

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

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

[5] Proposition C.4 presents a result with length constraints with stronger data requirements.

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

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

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

where $(p^r) = (p_j^r)_{j\in\mathcal{J}}$ 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.

**Proposition C.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.*

## C.2 Supplementary Results on Nonparametric Identification

**Proposition C.4** (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* $(v, z) \in \operatorname{supp}(v_i, z_i)$*.*[8]

**Proposition C.5** (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* $L = J$*,* $\mathbb{P}\big((v_{ij})_{j\in\mathcal{S}(z_i)} \leq v | z_i\big)$ *is identified for all* $(v, z_i)$ *in its support.*

*(ii) if* $L < J$*, then for all* $x$*,* $z = (z^v, z^-)$*, and* $\mathcal{A} \subseteq \mathcal{S}(z)$ *with* $|\mathcal{A}| \leq L$*,* $\mathbb{P}\big((v_{ij})_{j\in\mathcal{A}} \leq x | z\big)$ *is set-identified by an interval. Specifically, the interval has the endpoints given by* $\mathbb{P}(|r_i| < L, j \notin r_i \ \forall j \in \mathcal{A} | z_i^v = z^v - x, z^-)$ *and* $\mathbb{P}(|r_i| < L, j \notin r_i \ \forall j \in \mathcal{A} | z_i^v = z^v - x, z^-) + \mathbb{P}(|r_i| = L, r_i \cap A = \emptyset | z_i^v = z^v - x, z^-)$*, hence of the length* $\mathbb{P}(|r_i| = L, r_i \cap A = \emptyset | z_i^v = z^v - x, z^-)$*.*

[8] 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).

### C.3 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 C.1.** *Consider a list $r$ that contains an unacceptable program before an acceptable program, and the lowest-ranked program is an acceptable program. Then, $r$ gives a weakly lower expected utility than an alternative list that switches the lowest-ranked unacceptable program with the program that gives the maximum utility among the programs that follow this lowest-ranked unacceptable program.*[9]

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

### C.4 Proofs

*Proof of Lemma C.1.* By assumption, the list $r$ has an unacceptable program before an acceptable program. Let $j_-$ denote the lowest-ranked unacceptable program in the list. Then, there are some programs that follow $j_-$ and these programs are all acceptable. Let the utility-maximum of these programs be indicated by $j_{\max}$. Then, the report $r$ reads: $r = (\underbrace{\cdots}_{A}, j_-, \underbrace{\cdots}_{B}, j_{\max}, \underbrace{\cdots}_{C})$ where $A$, $B$, and $C$ denote (potentially empty) ordered sublist of the programs in each respective position. Consider an alternative list $r'$ that switches $j_{\max}$ with $j_-$, as in the statement: $r' = (\underbrace{\cdots}_{A}, j_{\max}, \underbrace{\cdots}_{B}, j_-, \underbrace{\cdots}_{C})$.

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

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

where $r(j)$ and $r'(j)$ denote the ranks of program $j$ in $r$ and $r'$, respectively. This is because otherwise, he gets into the same program regardless of reporting $r$ or $r'$ and obtains the same utility. It is impossible that he gets rejected in one report but not in the other report—his scores for any $j \in A$ 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 the subjective assessment of scores is restricted to depend only on certain aspects of the report—i.e., the rank.

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

Also, it must be that he gets rejected by $j_-$ under $r$. Otherwise, given that the student is rejected by all programs in $A$, this is the worst that can happen to him under $r$ or $r'$ because $B$ and $C$ can never have an unacceptable program. 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 be strictly preferred to $j_{\max}$. Thus, $\widetilde{\mathsf{cutoff}}_{j_-}(\omega) < \widetilde{\mathsf{score}}_{ij_-}\big(r(j_-);\omega\big)$ and $\widetilde{\mathsf{cutoff}}_{j_{\max}}(\omega) < \widetilde{\mathsf{score}}_{ij_{\max}}\big(r'(j_{\max});\omega\big)$.

Similarly, it must be that he fails to make the cutoffs (in either reports) by all programs in $B$. Otherwise, he gets same utility under the two reports. Note that he makes the cutoff in any of these programs in $B$ by submitting $r$ iff he does so in $r'$; the score for the program 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 what we had before: $\widetilde{\mathsf{cutoff}}_{j_-}(\omega) < \widetilde{\mathsf{score}}_{ij_-}\big(r(j_-);\omega\big) \leq \widetilde{\mathsf{score}}_{ij_-}\big(r'(j_-);\omega\big)$ and $\widetilde{\mathsf{cutoff}}_{j_{\max}}(\omega) < \widetilde{\mathsf{score}}_{ij_{\max}}\big(r'(j_{\max});\omega\big) \leq \widetilde{\mathsf{score}}_{ij_{\max}}\big(r(j_{\max});\omega\big)$.

By the same reasoning, it must be that he fails to make the cutoffs (in either reports) by all programs in $C$. Otherwise, he gets same utility under the two reports. Note that he makes the cutoffs in all of these programs 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 programs 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 C.2.* If $r$ ends with an unacceptable program, removing it weakly increases expected utility. Continuously drop any unacceptable programs from the end, each time weakly increasing expected utility. If unacceptable programs are interspersed, (repeatedly) apply Lemma C.1 to switch the lowest-ranked unacceptable program with the lower-ranked highest-utility acceptable programs, moving them to the end for removal, each time weakly increasing expected utility. This process, repeated until all unacceptable programs are removed, results in a list providing strictly higher expected utility compared to the original $r$; it has at least one instance of removing a considered (and therefore positive-assignment-chance) unacceptable program. □

*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 C.2, such a list with an unacceptable but considered program 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 program has (subjectively) positive admission chance upon listing. Addition of a program 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 $L$ programs if and only if $\{j \in \mathcal{J} | v_{ij} > 0, j \in \mathcal{C}_i\}$ has $L$ programs or more. Suppose first that $|r_i| = L$ but $|\{j \in \mathcal{J} | v_{ij} > 0, j \in \mathcal{C}_i\}| < L$. Because all programs in $r_i$ must be considered by definition, there must be some programs in $r_i$ that is subjectively reachable but is unacceptable. By Lemma C.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 program $j \notin r_i$ 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 C.1.* 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,

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

where generalizability of Observation 1 to $L \neq 12$ is immediate from its proof. As the first line represents what 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 C.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

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

where generalizability of Observation 1 to $L \neq 12$ is immediate from its proof. 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 $c_{ij}^* := \mathbb{1}(c_{ij} > 0)$.

The proof of (ii) follows analogously by noting that $\mathbb{P}(j \in r_i \ \forall j \in \mathcal{A}|z_i^c = z^c) = \mathbb{P}((\tilde{c}_{ij})_{j\in\mathcal{A}} > (z_j^c)_{j\in\mathcal{A}}|(v_{ij})_{j\in\mathcal{A}} > 0)\mathbb{P}((v_{ij})_{j\in\mathcal{A}} > 0|z^c)$ 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 C.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 $p' \equiv \{p_j'^r\}_{j\in\mathcal{J},r\in\mathcal{R}(\mathcal{J})}$ are behaviorally equivalent if and only if for all $v \in \mathcal{R}^J$, $\arg\max_{r\in\mathcal{R}(\mathcal{J})} v\cdot p^r = \arg\max_{r\in\mathcal{R}(\mathcal{J})} v\cdot p'^r$. 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 $\big(C^r(p)\big)_{r\in\mathcal{R}(\mathcal{J})}$ 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 $C^{\emptyset}(p) = \{(v_1, v_2) \in \mathbb{R}^2 | v_1, v_2 \le 0\}$, $C^{(1)}(p) = \{(v_1, v_2) \in \mathbb{R}^2 | v_1 \ge 0, v_2 \le 0\}$, $C^{(2)}(p) = \{(v_1, v_2) \in \mathbb{R}^2 | v_1 \le 0, v_2 \ge 0\}$, $C^{(1,2)}(p) = \{(v_1, v_2) \ge 0 | v_2/v_1 \le \delta,\}$, and $C^{(2,1)}(p) = \{(v_1, v_2) \ge 0 | v_2/v_1 \ge \delta\}$. 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 $\mathbb{P}(\{v_{i2}/v_{i1} \ge \delta\} \cap \{v_i \ge 0\}) = \mathbb{P}(v_i \in C^{(2,1)}(p)) = \mathbb{P}(r_i = (2, 1))$. As $\mathbb{P}(v_i \le v)$ is known, the left-hand side of the equation is calculable as a function of $\delta$. 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 $C^{\emptyset}(p) = \{(v_1, v_2) \in \mathbb{R}^2 | v_1, v_2 \le 0\}$. 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.

Now we can use Corollary 1 of Berry, Gandhi, and Haile (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).[10] To see that the corollary applies, note first that $\mathcal{X}$ is a Cartesian product. Moreover, $\sigma_j(\tilde{p})$ is strictly

[10]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 $p$.

decreasing in $\tilde{p}_k$ for all $j = \{1, \dots, J\}$ 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. $\square$

*Proof of Proposition C.4* . 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 $z^- \in \mathrm{supp}(z_i^-)$. Then, $\mathbb{P}\big(j \in r_i \ \forall j = 1, \cdots, J | z_i^v = z^v, z_i^c = z^c, z_i^- = z^-\big) = \mathbb{P}(\tilde{v}_i - z^v > 0, \tilde{c}_i - z^c > 0 | z^v, z^c, z^-) = \mathbb{P}(\tilde{v}_i > z^v, \tilde{c}_i > z^c | z^-) = \mathbb{P}(-\tilde{v}_i < -z^v, -\tilde{c}_i < -z^c | z^-)$ and since the first expression is observed for any $z^v \in \mathbb{R}^J$, $z^c \in \mathbb{R}^J$, and $z^- \in \mathrm{supp}\left(z_i^-\right)$, 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.

To show the first part of case (ii), note that $\mathbb{P}\big(r_i = \emptyset | z_i^v = z^v, z_i^c = z^c, z_i^- = z^-\big) = \mathbb{P}(v_{ij} \leq 0 \text{ or } c_{ij} \leq 0 \ \forall j \in \mathcal{J} | z^v, z^c, z^-) = \mathbb{P}(\tilde{v}_{ij} \leq z^v \text{ or } \tilde{c}_{ij} < z^c \ \forall j \in \mathcal{J} | z^-)$. Now, send all of the elements in $z^c$ to negative infinity. By the dominated convergence theorem, the last expression converges to $\mathbb{P}(\tilde{v}_{ij} \leq z^v \ \forall j \in \mathcal{J} | z^-)$. 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. $\square$

*Proof of Proposition C.5.* Proof of part $(i)$ follows by noting that $\mathbb{P}(r_i \cap \mathcal{S}(z_i) = \emptyset | z_i = z) = \mathbb{P}((v_{ij})_{j \in \mathcal{S}(z_i)} \leq 0 | z_i = z) = \mathbb{P}(\tilde{v}_{ij} \leq z_{ij}^v \ \forall j \in \mathcal{S}(z_i) | z_i^v = z^v, z_i^- = z^-) = \mathbb{P}((\tilde{v}_{ij})_{j \in \mathcal{S}(z_i)} \leq (z_j^v)_{j \in \mathcal{S}(z_i)} | z_i^- = z^-)$ 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 $z$. We shall use the fact that, for any two events $A$ and $B$, $\mathbb{P}(A \cap B) \leq \mathbb{P}(A) \leq \mathbb{P}(A \cap B) + \mathbb{P}(B^c)$. Consider the events $A = \{v_{ij} < 0 \ \forall j \in \mathcal{A}\}$ and $B = \{|r_i| = L, r_i \cap \mathcal{A} = \emptyset\}^c \equiv \{|r_i| < L\} \cup \{r_i \cap \mathcal{A} \neq \emptyset\}$. We have $A \cap \{r_i \cap \mathcal{A} \neq \emptyset\} = \emptyset$ by Lemma C.2. Therefore, $A \cap B = A \cap \{|r_i| < L\} = \{v_{ij} < 0 \ \forall j \in \mathcal{A}, |r_i| < L\} = \{j \notin r_i \ \forall j \in \mathcal{A}, |r_i| < L\}$, where the last equality is due to Observation 2. Note that both $\mathbb{P}(j \notin r_i \ \forall j \in \mathcal{A}, |r_i| < L)$ and $\mathbb{P}(B)$ are observable. Therefore the interval that bounds $\mathbb{P}(A) \equiv \mathbb{P}(v_{ij} < 0 \ \forall j \in \mathcal{A}) = \mathbb{P}(\tilde{v}_{ij} < z_{ij}^v \ \forall j \in \mathcal{A})$ is identified and given by

$$\Big[\mathbb{P}(j \notin r_i \ \forall j \in \mathcal{A}, |r_i| < L), \quad \mathbb{P}(j \notin r_i \ \forall j \in \mathcal{A}, |r_i| < L) + \mathbb{P}(|r_i| = L, r_i \cap A = \emptyset)\Big],$$

having the length $\mathbb{P}(|r_i| = L, r_i \cap A = \emptyset)$. One can then use the special regressor similarly as before to bound $\mathbb{P}\big((v_{ij})_{j\in\mathcal{A}} \leq x|z\big)$. $\square$

# D Construction of Figures and Tables

Table 1 shows the *State Reading Category*, based on 7th-grade reading scores from the New York State English Language Arts test. The categories are "Low" (bottom 16%), "Middle" (middle 68%), and "High" (top 16%). For non-public school students, categories are determined by another standardized assessment for admissions to Educational Option programs. Neighborhood income, derived from students' ZIP codes, uses median household income from the 2013–2017 U.S. Census Bureau's American Community Survey five-year estimates in 2017 dollars.

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 4, we use predicted priority groups for the reasons explained in Supplemental Material A.1. We find that 97.05% of the prediction error in priority groups (based on a sample of 20,000 students) are from programs that employ (as a device to determine priority groups) "attendance at information sessions." Because of this, we exclude these 238 programs (out of 743) programs.

In Table 8, 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.[11]

Other matchings (Panel B) in Table 8 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

[11] Refer to Supplemental Material B.1 for approximation details.

admission policies by the programs.[12]

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 Estimation: Details

## E.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) the likelihoods are 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 E.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 $\theta$. 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

[12] Refer to Appendix A for details on the estimation of the screening policies and the approximation of the priority groups. See Supplemental Material B.2 for details on our implementation of Deferred Acceptance algorithm.

subsection. Then,

$$\log \Pi_i \Pi_{j:s_{ij}=1} f_{\iota|z,s}(\iota_{ij}|z_{ij}, s_{ij}=1;\theta) = \log \Pi_i \Pi_{j:s_{ij}=1} f_{w|z}(w_{ij}|z_{ij};\theta)$$

$$= \sum_i \Bigg[ \sum_{j:s_{ij}=1, j\notin \mathcal{S}_i} \Big[ \big(1-w_{ij}\big) \log \big(1-\overline{\Phi}(-\psi_{ij}^v)\overline{\Phi}(-\psi_{ij}^c)\big) + w_{ij} \log \big(\overline{\Phi}(-\psi_{ij}^v)\overline{\Phi}(-\psi_{ij}^c)\big)\Big]$$

$$+ \sum_{j:s_{ij}=1, j\in \mathcal{S}_i} \Big[ (1-w_{ij}) \log \big(\Phi(-\psi_{ij}^v)\big) + w_{ij} \log \big(\overline{\Phi}(-\psi_{ij}^v)\big)\Big]\Bigg] \tag{E.1}$$

where $\bar{\Phi}(\cdot) := 1-\Phi(\cdot)$, $\psi_{ij}^v := v_{ij} - \epsilon_{ij}^v$, $\psi_{ij}^c := c_{ij} - \epsilon_{ij}^c$, and $\theta^v$ and $\theta^c$ denotes the preference and consideration parameters, respectively. For convenience, the dependence of $\psi_{ij}^v$ on $\theta^v$ and the dependence of $\psi_{ij}^c$ on $\theta^c$ are made implicit. The second equality comes from $s_{ij}=1$ being independent of $(\epsilon_{ij}^v, \epsilon_{ij}^c)$ and therefore also of $(v_{ij}, c_{ij})$ conditional on observables. The second summation in the final expression, including only surely considered programs, is based solely on the variation in Observation 2 and therefore depends only on preference parameters. This part of the partial likelihood can also be used to estimate preference, whose results are summarized in Figure A.3.

We now show that the population version of the log-likelihood is maximized by the true parameters $\theta_0$. Define

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

This is the population version of the log-likelihood (Equation E.1) in the sense that $Q(\theta) = \text{plim}_{n\to\infty} n^{-1} \log \Pi_i \Pi_{j:s_{ij}=1} f_{\iota|z,s}(\iota_{ij}|z_{ij}, s_{ij}=1;\theta)$ where $n$ denotes the number of students in the sample.

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

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

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 put larger weights on sure-consideration pairs. The weighted log-likelihood is

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

for some weights $\omega_{\text{NSC}}$ and $\omega_{\text{SC}}$ 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 $\omega_{\text{NSC}}$ and those that are surely considered are weighted by $\omega_{\text{SC}}$.

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 $\omega_{\text{NSC}}$ 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 $\tilde{s}_{ij} := s_{ij}\mathbb{1}(j\in\mathcal{S}_i)$.

### E.2 Simulated Ordering Moments

In this subsection, we denote the union of variables in $(x^v_j, x^c_j, z^v_{ij}, z^c_{ij})$, as defined in the empirical specification (Section 6), as simply $z_{ij}$. 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 $(\epsilon^v_i, \epsilon^c_i, \eta_i)$, $\theta$ denotes the parameter vector, $\theta_0$ denotes the true parameter vector, and $r(z_i,e_i;\theta)$ denotes the subjectively optimal report under $(z_i,e_i,\theta)$ which is uniquely defined with probability 1. Section F describes the procedure for simulating $r(z_i,e_i;\theta)$. It follows that

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

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

The sample equivalent of this condition is

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

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

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

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

where $r_i^k$ is represents the report $r_i$ truncated up to the $k$th 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 E.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 $h(z_i) = \big(1, z_{ij}, (z_{ij} - \bar{z}_i)^2, \mathsf{cutoff}_{ij} - E_{\text{obj}}[\mathsf{score}_{ij}]\big)_{j\in\mathcal{J}}$ 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.

### E.3 Lemmas

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

*Proof.* Fix the observables $(x, z)$. Note that it suffices to show that $|r_i \setminus \{j\}| < 11$ 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 $|r_i \setminus \{j\}| < 11$ holds if and only if both $\sum_{j'\neq j} \mathbb{1}\{c_{ij'} > 0, v_{ij'} > 0\} < 11$ and $|r_i| < 12$ holds due to the first statement of Observation 1. This is then equivalent to iff $\sum_{j'\neq j} \mathbb{1}\{c_{ij'} > 0, v_{ij'} > 0\} < 11$ due to the second statement of Observation 1. □

# F Simulating Subjectively Optimal Reports

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

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

where the distribution of $(\mathcal{C}_i, v_{ij}, p_{ij}^r)_{ij}$ depends on $\theta$. 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, Fu, and Güell (2020), to make this problem solvable through backward induction, we represent this problem as what resembles a finite-horizon dynamic programming problem.

Let $j_k$ represent the school listed in the $k$th 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:

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

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.

# Supplemental Material for
# Distributional Impacts of Centralized School Choice

Jaewon Lee    Suk Joon Son

August 17, 2024

## A Data Appendix

### A.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.

### A.2 Scores and Cutoffs

As in the main text, we model student $i$'s belief regarding program[1] $j$ written at the $k$-th slot of his report as $q_{ijk} = \mathbb{P}_i\big(\mathsf{cutoff}_j - E_{\text{obj}}[\mathsf{score}_{ij}] + \epsilon^b_{ijk} > 0\big)$ where $\mathbb{P}_i$ is the probability measure of $\epsilon^b_{ijk}$. The belief consists of two parts, namely the objective difference $\mathsf{cutoff}_j - E_{\text{obj}}[\mathsf{score}_{ij}]$ and the subjective assessment $\epsilon^b_{ijk}$. In this section we explain the empirical specification of the objective difference, starting with $\mathsf{cutoff}_j$.

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 $\mathsf{cutoff}_j$ 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.[2]

Now we turn to $E_{\text{obj}}[\mathsf{score}_{ij}]$. 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 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 $\nu_{ij}$.

We do not observe $\mathsf{priorityGroup}_{ij}$ for all $(i, j)$ pairs, and hence we impute their values. As explained in Supplemental Material A.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,

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

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

screen-based programs, and Educational Option programs. For lottery-based programs,[3] 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,[4] the tiebreaker is the screening priority and we observe how programs ranked a subset of the applicants by their screening policies.[5] 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 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.

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

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

[5]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.

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[6] of the objective differences of the subprograms to approximate the belief on the program, i.e.,

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

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 $\mathsf{cutoff}_j - E_{\text{obj}}[\mathsf{score}_{ij}]$.

### A.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.

[6] 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.

### A.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.

### A.5 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.

## B Simulations of matchings

### B.1 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.

### B.2 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.