Essays in Labor and Education Economics

by

Viola Corradini

M.Sc., Università Commerciale Luigi Bocconi (2017)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2024

Author	Viola Corradini Department of Economics May 14, 2024
Certified by	Joshua Angrist Ford Professor of Economics Thesis Supervisor
Certified by	Parag Pathak Class of 1922 Professor of Economics, Thesis Supervisor Thesis Supervisor
Accepted by	Isaiah Andrews Professor of Economics Chairman, Departmental Committee on Graduate Studies

Essays in Labor and Education Economics

by

Viola Corradini

Submitted to the Department of Economics
on May 14, 2024, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis consists of three chapters on labor economics and the economics of education. The first two chapters study the reasons behind racial disparities in school choices and propose two solutions to alleviate them: providing information about school quality and promoting attendance in racially integrated schools during earlier grades.

Differences in school choice by race contribute to unequal access to effective schools and exacerbate school segregation. Conditional on test score and district of residence, Black and Hispanic families consistently opt for schools with fewer white and Asian students, lower average achievement, and lower value-added. The first chapter asks how information about school quality affects this gap. Specifically, I examine the effects of New York City's introduction of a letter-grade system rating the quality of its high schools. The ratings shifted Black and Hispanic students' choices more than those of white and Asian students, narrowing racial gaps both in enrollment at high-quality schools and in academic achievement. Using a structural model of school choice and surveys of families, I find that race differences in the response to quality information stem in part from different beliefs and preferences. The model estimates suggest that Black and Hispanic students have less accurate perceptions of school quality, making them more receptive to the grade-based scoring system. In addition, white and Asian students are less influenced by information on school quality because they have strong preferences for other school attributes. Simulations suggest that better quality information narrows racial gaps in choice and achievement. Additionally, simulations indicate that the design of information is important in determining who benefits most from its provision. A system that releases coarse quality ratings for high-quality or oversubscribed schools increases test scores among lower achieving students more than perfect information by reducing the competition for high-quality schools from higher achieving students.

The second chapter, joint with Clemence Idoux, combines unique survey data and administrative data from New York City to identify the determinants of racial disparities in school choice and shows that attending a more diverse middle school can mitigate racial choice gaps. A post-application survey with guardians of high school applicants reveals that information gaps and homophily in school preferences explain cross-race differences in choice. In turn, instrumental variable estimates show that middle school students exposed to more diverse peers apply to and enroll in high schools that are also more diverse. These effects are consistent across racial groups, particularly benefiting Black and Hispanic students who enroll in higher value-added high schools. Notably, changes in application patterns due to exposure to diverse middle school peers appear driven by changes in the set of known school options and an increased preference for peer diversity.

The final chapter, joint with Lorenzo Lagos and Garima Sharma, investigates why workplaces are not better designed for women. In particular, we show that changing the priorities of those who set workplace policies can create female-friendly jobs. Starting in 2015, Brazil's largest trade union federation made women central to its bargaining agenda. Using a difference-in-differences design that exploits variation in affiliation to the federation, we find that "bargaining for women" increases female-centric amenities in collective bargaining agreements, which are then reflected in practice. These changes lead women to queue for jobs at treated establishments and separate from them less—both revealed preference measures of firm value. We find no evidence that these gains come at the expense of employment, wages, or firm profits. Our results suggest that changing institutional priorities can narrow the gender compensation gap.

JEL Classification: I20, I21, J52

Thesis Supervisor: Joshua Angrist
Title: Ford Professor of Economics

Thesis Supervisor: Parag Pathak
Title: Class of 1922 Professor of Economics, Thesis Supervisor

Acknowledgments

This thesis would not have been possible without an incredible team of supporters.

First, I would like to thank my advisors Josh Angrist, Parag Pathak, Nikhil Agarwal and Simon Jäger for their guidance and encouragement over the past years. I am grateful for the opportunity to learn from and work with Josh, who introduced me to the field of Education Economics and to its tools of trade. His approach to research, teaching and writing will continue to shape my work for many years to come. I want to express my gratitude to Parag for guiding me to ask important questions and for helping me see the bigger picture behind my research work, as well as for his careful review of the drafts of my job market paper. His insights helped me make progress when I felt stuck. I am thankful to Nikhil for his energy, support, and patience in advising me as I started using tools outside my comfort zone. Lastly, but certainly not least, I want to thank Simon for his kindness, availability, and unwavering encouragement throughout this journey.

I am also thankful for the support of the wider MIT community, including the invaluable assistance of Blueprint Labs staff, notably Eryn Heying and Jim Shen, and the MIT Department of Economics staff. I extend my gratitude to Daron Acemoglu and David Autor for consistently attending my presentations and offering insightful comments, as well as to the numerous fellow PhD students who attended my talks and provided valuable feedback.

A heartfelt thanks goes to my coauthors, Clemence Idoux, Garima Sharma and Lorenzo Lagos. Our meetings and discussions have been central to my learning and made me a better researcher. Thanks to Tishara, Deivy, Dean and Ray for being the best office mates and fostering a sense of community inside and outside our office. Thanks to Marc, Sylvia, Isa and Laura for your friendship and for the countless lunches and coffee chats that made coming to E52 much more fun.

I am lucky to count on the support of numerous friends outside of MIT who have been with me through the highs and lows of this six-year journey. I am grateful for Beatrice's and Silvia's cheerful and warm presence and advice. I am very grateful for Giulia's friendship, steadfast support, and the deep connection we share. Your passion for both research and life inspires me to grow as a person and a researcher. I would like to thank Nicola, for being a significant source of strength, for his patience and encouragement.

My Italian community in Cambridge served as my family during my life as a PhD student. Thank you to Roberto, Francesca, Enrico, Veronica, Matteo, Valerio, Caterina, Clara, Adriano, Valentina, Claudia and Chiara for sharing this journey with me. A special thanks goes to my roommates Antonio and Luca for being an important anchor and for creating a sense of home throughout these past six years.

The most important acknowledgment goes to my family. I am grateful for my sister Gemma, her constant and loving presence in my life. I am incredibly thankful to my parents Vilma and Massimo, who taught me the importance of thoroughness in my work, showing up for others, and how to live a meaningful life. Thank you for your constant and selfless support throughout my educational journey, even when it meant living abroad for many years. This thesis is dedicated to you.

Contents

List of Figures

List of Tables

Chapter 1

Information and Access in School Choice Systems: Evidence from New York City

1.1 Introduction[1]

School choice systems are an increasingly popular alternative to neighborhood-based assignment of students to schools (Neilson, 2019). Proponents argue that such systems can reduce achievement gaps by offering everyone access to high-quality education, increase allocative efficiency and induce schools to improve to keep and attract students (Friedman, 1955; Hoxby, 2000, 2003). Critics counter that market-based education reforms fall short of their goals because the conditions for competition and fair choice are not met in practice (Ravitch, 2010). Families make choices based on the demographics of the student body, rewarding schools that draw from wealthier, more educated communities rather than pressuring them to improve quality (Ladd, 2002; Rothstein, 2006; Barseghyan et al., 2019; Cullen et al., 2006). In addition, critics argue that choice opportunities are more likely to be exploited by students from more affluent and motivated families, further exacerbating racial and socio-economic inequality (Ladd, 2002). Their claims are corroborated by evidence that disadvantaged students are more likely to apply to and attend lower-quality schools, even though higher-quality choices are available

[1]I am thankful to my advisors Josh Angrist, Parag Pathak, Simon Jäger and Nikhil Agarwal for their guidance and advice throughout the project, to the Office of Enrollment Research and Policy and the Office of School Performance and Accountability of the New York City Department of Education for graciously sharing data, and to Daron Acemoglu, David Autor, Clemence Idoux, Frank Schilbach, Benjamin Vatter and participants to the MIT Labor lunch and the MIT Applied Micro lunch for helpful comments. Thanks to Eryn Heying and Jim Shen for dependable administrative support. I am grateful to Jerry Hausman, the George and Obie Shultz Fund, and the Guido Cazzavillan Fellowship for financial support. This paper reports on research conducted under data-use agreements between MIT, the project's principal investigator, and the New York City Department of Education. It reflects the views of the author alone.

(Laverde, 2020; Hastings et al., 2009; Corradini and Idoux, 2023; Hoxby and Avery, 2012; Carlana et al., 2022).

Would providing families better information unleash the potential of school choice to reduce inequality and boost achievement? If so, what are the most effective ways to present this information to families? This paper studies these questions in the context of high school choice in New York. In this setting, I document that Black and Hispanic families apply to lower-quality schools, as measured by causal estimates of school value-added, even after controlling for residential location and differences in attainable options. This gap may be explained by differences in knowledge about school quality or in preferences for school attributes. What would be the impact on achievement inequality if the district made school quality information more accessible? The answer would depend on who made the most use of it: affluent families, who may have stronger preferences for school performance, or disadvantaged ones, potentially less well informed. When high-quality schools are in short supply, inequality could be exacerbated depending on how disadvantaged families respond to increased information compared with advantaged ones. Information could favor those responding more strongly while displacing others.

In 2007, New York City introduced a system that rated high schools by grades A to F, based on factors such as student progress, standardized test scores, and attendance and graduation rates, while controlling for demographic differences. The grading system, which was part of a broader school accountability reform effort, was then removed in 2014. This setting presents several advantages to study my research questions. In the presence of imperfect information about school quality, inferring preferences from realized choices is challenging, as these depend on families's perceptions of schools. Therefore, assuming perfect information may lead to the erroneous conclusion that families do not reward quality and that market-inspired interventions in education are bound to fail. The introduction, changes, and removal of grades provide a natural experiment that can be used to address the key research challenge of separating household beliefs from preferences for school quality. Accounting for heterogeneity in these two components of school demand helps explain who benefits from information interventions in equilibrium and why. Second, while researchers often need to explicitly model school admissions to simulate counterfactual equilibrium outcomes, I am able to draw on detailed data on school capacity and the rules of the centralized admission mechanism to directly simulate assignment of students to schools.

Exploiting within-school changes in letter grades, I find that student choices respond to information about school quality. High grades boost demand for seats, while low grades reduce demand. This shows that families indeed value school effectiveness, apart from other school attributes, such as peer quality, but hold uncertain beliefs about it. Schools with low achievement levels, a commonly perceived indicator of quality, saw demand increase more when they received high letter grades compared to schools with high achievement levels. Correspondingly, schools with high achievement levels saw larger demand declines when they received low scores.

Black and Hispanic applicants respond more strongly to the school grades than do Asian and white applicants. Minority students are 7 percentage points more likely to apply to a school that always received an A after the introduction of letter grades compared to white and Asian students, off a base of 48 percentage points. Similarly, minority students are 9 percentage points (off a base of 34 percentage points) less likely than white and Asian students to apply to a school receiving consistently low grades after their introduction. While letter grades do not substantially affect white and Asian student choices on average, they still do within the subset of schools enrolling higher shares of white and high-performing students. These findings suggest that white students hold strong preferences for school attributes other than quality that attenuate their responses to information.

The larger shifts in demand among minority applicants reduced the cross-race gap in chosen value-added by about 0.03 test score standard deviations (σ), or 4.5 percentiles. After the grading system was discontinued in 2014, student choices partially reversed back to their pre-letter-grade patterns. While the grading system increased Black and Hispanic students' applications to high grade schools, these students did not always gain in admissions. As a result, information reduced the racial gap in applications to high grade schools more than in admission. In some cases, high grade schools screened out students on the basis of test scores, disproportionately favoring white students. In other cases, the increased demand from Black and Hispanic students led to greater competition for the schools they were selecting.

To better understand what drives or constrains the beneficial effects of information interventions, I specify and estimate a model of demand for schools using data on rank-ordered preference lists. Departing from the standard typically adopted in the school choice literature, the model allows imperfectly informed students to hold prior beliefs about school quality and to update them when receiving quality signals using Bayes rule. Adapting the argument used in Vatter (2022), I show how variation in school quality ratings within schools and their availability over time separately identifies student preferences and beliefs over quality. I let preferences and beliefs vary across students with different demographic characteristics and let beliefs about school quality depend on school average achievement levels. These modeling assumptions are informed by differences in student responses to letter grades across schools of different achievement levels.

Estimates suggest that racial differences in beliefs and, even more so, preferences explain the larger response to information among minority students. Applicants of all racial backgrounds believe that high-performing schools are high quality. This belief is most firmly held by white and high-achieving students. As a result, they are less likely to change their choices in response to letter grades than students who are low-achieving, Black or Hispanic. Survey data that I collected among a more recent cohort of high school applicants validates the structural belief estimates. Beliefs elicited in my survey are also positively correlated with achievement levels and value-added, particularly among white respondents. I interpret the findings as suggesting that differences in perceptions stem from signals families receive from their social networks. Black and Hispanic families receive more mixed signals from their social networks before observing the quality ratings. High-achieving, white students receive more precise signals from their social networks that equate schools with high-achievement with high quality.

Racial differences in preferences for school attributes are even more important than beliefs to explain differing responses to the ratings. All students similarly trade-off preferences for attending higher quality schools with distaste for commuting. White and Asian students, however, prefer the few public schools that are majority white and Asian, which reduces their responsiveness to information about other schools. Overall, this subset of schools is also of higher quality, which explains why white and Asian students choose better schools even without letter grade information. Minority students, by contrast, prioritize quality over other attributes more than non-minority students.

The model provides an opportunity to test whether information design can increase student achievement and close opportunity gaps. Providing perfect information about school value-added would cause students to rank schools with 0.07σ higher value-added on average. The larger response among Black and Hispanic students would close cross-race choice gaps conditional on baseline test scores. Because there is slack in the capacity of high-quality schools, changes in choices would translate into average test score gains of 0.01σ in equilibrium, with marginally larger gains among minority and high-achieving students. This number correspond to 24% of the maximum possible achievement gains that would be realized if school seats were filled in order of quality. The ability to accurately measure school value-added is crucial, as simply providing information about school average achievement levels yields less than half of these test score gains. Crucially, what allows Black and Hispanic students to benefit from information in equilibrium, even in the presence of test-based admission standards, is that white and Asian students hold relatively stronger preferences for schools enrolling more advantaged students. The distribution of achievement gains would be different if students exclusively valued school quality and commuting time. In this scenario, information would improve the quality of school offers for white and Asian students substantially but hurt minority students.

Achievement gains for Black and Hispanic students under full information are comparable with those obtained through more controversial school admission reforms often targeted at reducing racial inequalities in New York City (Cohen, 2021). Information yields 80% of the minority gains realized by removing admission priorities based on where students live or their baseline test scores. School match simulations also show that combining information and leveling the playing field in admission rules are not substitute policies but their redistributive effects are cumulative. Information amplifies the displacement effects of removing screens in admissions on high-achieving white and Asian students, rather than causing these students to reallocate to better schools.

In the absence of changes in admissions rules, however, information displaces many lower achieving students. Targeted information can help this group of students obtain access to high-quality education when seat capacity is fixed. However, if fairness concerns prevent targeted outreach, information can be distributed to everyone and designed in a way to favor one group over another. Notably, coarser information, such as partitioning value-added into school grades, can lead to better educational outcomes for low-achieving students compared to offering more detailed information. Intuitively, information about the quality of schools that are considered non-desirable for other reasons would not induce large choice responses. Therefore providing more detailed information about schools valued relatively more by low achievers, while coarsening information about schools valued more by high achievers, would redistribute quality. This policy limits shifts in choices of high achieving students and therefore the competition for high quality seats, and allows low achievers to match to better schools. A system that provides more precise quality ratings for schools at the bottom of the value-added distribution and coarser for those at the top, or a system that provides more precise ratings for undersubscribed schools than for oversubscribed schools, would benefit lower achieving students more than a system which provides perfect information about the value-added of every school.

This paper contributes to the literature studying household preferences for schools (Abdulkadiroğlu et al., 2020; Beuermann and Jackson, 2018; Hastings et al., 2009; Allende, 2020; Abdulkadiroğlu et al., 2017a). These studies explore whether the strength of preferences for school quality is sufficient to enable school choice and accountability to produce efficiency gains (Hoxby, 2000, 2003; Ladd, 2002; Rothstein, 2006; Cullen et al., 2006; Barseghyan et al., 2019; Campos, 2023a; Walters, 2018; Abdulkadiroğlu et al., 2018). Abdulkadiroğlu et al. (2020) find that preferences for schools are uncorrelated with school value added after controlling for peer quality, casting doubts on the strength of competitive incentives introduced by school choice. A recent group of papers, however, considers the role of information frictions and inaccurate beliefs in explaining this choice pattern, often using a combination of surveys and experiments (Allende et al., 2019; Bergman et al., 2020; Hastings et al., 2015; Kapor et al., 2020; Ainsworth et al., 2023; Campos, 2023b; Corradini and Idoux, 2023). This paper contributes to this debate by showing that students care about school value-added separately from peer quality. It also quantifies the importance of value-added and peer quality relative to other attributes by specifying a model where families have imperfect information about school quality. The model does not rely on the direct elicitation of beliefs, which may be unfeasible when the set of schools is large.[2] I show that the two methods nevertheless provide qualitatively similar results comparing the model-based estimates against prior means elicited using survey data.

Second, the paper relates to studies evaluating the effects of information interventions in education. Following Hastings and Weinstein (2008), several papers study the effects of providing information about schools on choices and education outcomes. Effects range from zero to large and positive (Mizala and Urquiola, 2013; Cohodes et al., 2022; Corcoran et al., 2022; Allende et al., 2019; Andrabi et al., 2017). The majority of studies focuses on changes in the demand side of the market. Notable exceptions are Rockoff and Turner (2010), who show that school accountability reforms may directly affect school incentives before changes in choices are realized and Allende et al. (2019) and Andrabi et al. (2017) who study changes in both the demand for and the supply of quality. Several have relied on experimental evidence in which a random subset of students receive information and do not directly observe equilibrium effects. Moreover, most of these experiments provide information about school outcome levels rather than value-added, so it is unclear what families learn about school quality and what they are responding to. The setting I study has several distinct advantages. First, letter grades were framed as measures of school effectiveness, which lets me learn about household preferences and beliefs for quality. Second, the information was provided by an actual policy rather than researchers, which may be more informative about the potential effects of a large-scale intervention conducted by a school district. This distinction seems important in light of the fact that most families primarily rely on institutional sources to gather information about schools (Corradini and Idoux, 2023). Third, thanks to detailed data on applications, school admission rules, capacity and offers, I can directly observe congestion and displacement effects in the equilib-

[2]Papers eliciting beliefs usually focus on a small subset of applicants and schools (Kapor et al., 2020; Campos, 2023b)

rium when everyone is informed, which is important when school seats are scarce.

Finally, the paper relates to studies empirically examining the distributional and efficiency effects of assignment reforms, including affirmative action (Barahona et al., 2023a; Idoux, 2021; Black et al., 2023; Tincani et al., 2021; Bleemer, 2021; Kapor, 2020; Ellison and Pathak, 2021) and changes in admission rules (Dur et al., 2018; Park and Hahm, 2023). The distributive role of quality ratings design connects this study also to the literature on the design of information disclosure policies (Vatter, 2022; Kamenica, 2019). This paper contributes to both literatures by studying how information and its design benefits and hurts individuals and by considering how to optimally coarsen ratings to implement distributional objectives.

1.2 Background

1.2.1 Institutional Context

The NYC High School Match Every year New York City public schools enroll roughly 80,000 ninth graders at more than 400 high schools. Rising ninth graders apply for school seats by submitting an application to the centralized assignment system, ranking up to 12 academic programs.[3] Seats are allocated using the student-proposing deferred acceptance (DA) algorithm (Abdulkadiroğlu et al., 2005, 2009). Student priorities at a program depend on different factors, which vary depending on the program admission method type. There are three types of programs. Unscreened programs give priority to students based on residential zones and in some cases to those who attend an information session. Screened programs use these factors and also rank applicants based on prior grades, standardized test scores, attendance and/or program-specific requirements, such as essays, or auditions. Educational option programs use screened criteria for some of their seats and unscreened criteria for the rest. Random numbers are used to break ties among applicants with equal priority.

Sources of Information About Schools Navigating the high school admission process is a daunting task for many families. Parents lament that gathering information about more than 700 programs with different admission methods, is difficult, costly, and time consuming (Corradini and Idoux, 2023; Son, 2020). To aid families in their decision-making, the NYC Department of Education (DOE) assembles every year a high school directory and maintains a website with detailed measures about school performance. Before 2019 and throughout the period I study,

[3] Schools may run more than one academic program, but most schools (70%) offer only one. For the purposes of this paper, programs and schools should be treated as synonyms.

the directory was provided in paper copy to every 8th grader in the city.[4] This printed booklet was the main tool used by families to choose schools, as confirmed by conversations with staff at the DOE and by interviews conducted among middle school counselors by Sattin-Bajaj et al. (2018). Today, the DOE application portal (MySchools.nyc) hosts a virtual version of what used to be the printed high school directory booklet before 2018. Survey evidence confirms it is still the source of information most widely used by high school applicants across all demographic groups, while reliance on other information sources varies across race (Corradini and Idoux, 2023).[5]

The directory provided an overview of the high school admission process, key dates, and an information page for each high school, which always included the school address, total enrollment, offered programs and their admission methods, courses and extracurricular activities, and a brief statement of its mission. The school pages also provided measures of school performance and student achievement, such as graduation rates.

Accountability Reforms and Changes in Information Over the course of the years, the NYC Department of Education (DOE) changed its way of measuring and reporting school quality metrics on the school directory and online. Table A.1.1 summarizes these changes during the study period.[6] The most noticeable addition to the information provided on the high school directory was the inclusion, from 2010 to 2015, of letter grades that graded schools from A to F. The letter grades were introduced as part of a broader set of education reforms adopted by the Bloomberg administration after taking mayoral control of the city schools in 2002. Bloomberg's approach to reforming schools was designed around market-based principles of increasing quality through teacher incentives, school competition and accountability (Ravitch, 2010). Beginning in the fall of 2007, the DOE started issuing yearly school progress report cards that evaluated school performance and provided a summative assessment in the form of letter grades.[7] They were provided in addition to statistics about the school average achievement levels, such as graduation rates, and were meant to measure the school's contribution to student academic progress.[8] On the high school directories they were described with words similar to what one

[4] Since Fall 2019, families receive thinner admission guides with general information that point to an online portal for more detailed, school-by-school data (Amin, 2019).

[5] Black and Hispanic applicants use fewer sources of information than white and Asian households, are 19 p.p. less likely to rely on their family and friend networks for information about schools, and are 9 p.p. less likely to attend individual high school information sessions.

[6] Years indicate fall of 9th grade enrollment.

[7] An example of what they looked like is provided in figure A.1.1

[8] Table A.1.1 shows that graduation rates were always printed on the school directory, except between 2008-2010, while college rates were introduced on the booklet in 2013 and average performance on Regents exams was only reported in 2006 and 2007. These statistics were based on students enrolled at the schools 2 years before

would use to explain what school value-added is.[9]

In practice, the scoring rule used to assign grades was not based on causal estimates of school quality, but it attempted to control for underlying differences in the student body and was positively correlated with causal test scores value-added. Rules varied slightly by education level. For high schools, they were based on three continuous measures of school performance: school environment, based on student attendance and answers to a school environment survey (around 14% of the total score), student performance, based on graduation rates and average student performance on Regents exams (around 30% of the total score) and student progress, based on test score growth for all students and for students in the lowest third of all students citywide (around 50% of the total score). The formula was tweaked in 2010 to add measures of college readiness as a fourth component of the total score (around 10% of the total score), reducing the importance of the other components.

A school's score for each element was determined not only by the school relative performance city-wide, but also relative to a group of 40 "peer schools" with similar student demographics. Performance relative to peer schools was given double the weight of citywide relative performance in an attempt to separate school quality from student selection (Rockoff and Turner, 2010). A school's overall score was calculated using the weighted sum of the scores within each element plus any additional credit received.[10]

Letter grades were assigned on the basis of the school score percentile and had consequences for school closures, financing and school principals.[11] Rockoff and Turner (2010) studied the effects of introducing the grades on the incentives of elementary and middle schools to raise students test score in the first year the policy was introduced, finding that only receiving an F induced schools to raise test scores.

Applicant cohorts of 2008 and 2009 could see the letter grades on the DOE website online,

when applicants are applying. When measuring the effects of letter grades on choices in section 1.3.1, I separately control for all these changes.

[9]The description of the progress report on the school directories reads: *"The Progress Report measures each school's contribution to student academic progress, no matter where each child begins his or her journey to proficiency."*

[10]Schools could also receive additional points for improving student achievement from year to year among particularly vulnerable student subgroups (English Language Learner, special education students, and Black, Hispanic or LatinX students with performance in the lowest third of all students citywide). Appendix Table A.1.2 describes the education outcomes used to compute the score in each component and reports the component weight in each year, before and after the assignment of the extra points.

[11]Schools receiving low grades could face leadership changes or closure, and students enrolled in F schools were eligible to transfer out through a special application process. Schools receiving an A grade received additional funding for the following school year of roughly $33 per student, and were eligible together with B schools for payments of $1500 to $3000 per student per year for any student accepted as a transfer from a failing school. Principals in the top 20% of scores were eligible to receive bonuses of $7000 to $25,000 (Rockoff and Turner, 2010).

searching school by school, while the 2010-2014 cohorts could read the letter grades directly on the school directory. In 2014, the newly elected mayor, Bill de Blasio, removed letter grades from school quality reports (that were renamed school quality snapshots) and the following year his administration introduced a new approach to school quality measurement, vowed to be more holistic and less focused on test scores.[12] These new quality snapshots, however, never made it to the printed school directory but could be consulted online, on a school by school basis.

1.2.2 Data

Sources of Data I combine three main sources of data. The first is publicly available data from the school directories and online school quality reports issues by the NYC DOE between 2006 and 2016. The second is administrative data provided by the DOE covering all students enrolled in New York City public high schools between the 2006-2007 and the 2016-2017 school years. These data include student demographics and residence, school enrollment, student educational outcomes, including test scores on New York State standardized tests in middle school and high school (Regents exams), SAT and high school graduation, along with preferences submitted to the centralized high school assignment mechanism. An additional file from the National Student Clearinghouse (NSC) reports college enrollment and is internally linked to the DOE administrative data. I supplement the administrative data with public transport commuting time between schools and students' addresses measured at 7:30AM that I obtain using publicly available APIs. The third source is data from a survey of 3500 parents of 9th grade applicants collected between February and March of 2023 and analyzed more extensively in a companion paper (Corradini and Idoux, 2023). Here, I use the survey to elicit parents' beliefs about high school effectiveness in preparing students for their end-of high school (Regents) exams.[13]

Measures of School and Peer Quality I use student achievement data to construct two key attributes of schools: school quality, measuring the causal contribution of schools to student achievement, and peer quality, measuring the average ability of students enrolling in a school. I define peer quality to be the average 7th grade standardized NY state Math test scores of students enrolled at a school in a year. I measure school quality using a standard school value added model (VAM) of high school standardized test scores, namely Regents and SAT Math test scores. Regents exams are New York state standardized exams in core high school subjects that

[12] These changes apply to the cohort applying to enroll in the fall of the following year (in 2015 and 2016 respectively).

[13] Appendix C.2 provides more detail on the data and the survey.

are required in order to graduate high school. My main value-added measure is given by OLS estimates of α_j in the following regression:

$$Y_i = \alpha_0 + \sum_{j=1}^{J} \alpha_j D_{ij} + X_i' \Gamma_{t(i)} + \epsilon_i \tag{1.1}$$

where Y_i measures student i's standardized Regents or SAT math score, D_{ij} is a dummy indicating 9th grade enrollment in school j and X_i is a vector of baseline controls including race and ethnicity, subsidized-lunch and English Language Learner (ell) status, and lagged test scores (7th grade Math and English standardized state test scores). I allow the effects of X_i to vary by cohort, as denoted by $t(i)$. This model assumes that school quality is fixed over time and across student demographics and relies on a standard conditional independence assumption (CIA) that states that potential outcomes are independent of school fixed effects after controlling for the vector of student covariates X_i.

In robustness checks in appendix A.2, I relax these assumptions in two ways. First, I relax the CIA by estimating risk-controlled (RC) VAM, as introduced by Angrist et al. (2021). RC VAM supplements the vector of controls with applicant characteristics integral to school matching, such as where they apply and the priority status that a school assigns them.[14] I use random variation in school offers embedded in the centralized school match to test how well conventional and RC VA estimates predict student outcomes (Angrist et al., 2016, 2021, 2022b). These tests show that conventional and RC measures are equally unbiased and well predictive of student Regents test scores, while adding risk-related controls substantially improves the predictive validity of SAT VA measures. For this reason, I will use Regents scores as the primary outcome to measure school quality unless otherwise noted.[15] More details about the test statistics and the test implementation are provided in Appendix A.2.

In a second robustness check, I relax the constant-effect model and allow school effectiveness to vary by student race. Value-added measures for different races are strongly correlated within schools. Lottery-based tests of bias in appendix table A.2.21 confirm that measures of VA that do not vary by race ("Pooled VA") have a good predictive validity for student Regents

[14]RC VAM estimates are not available for a subset of schools in my sample and use data from a shorter time span because they rely on the possibility of replicating the high school match. I have the necessary information to do this starting from the 2012 cohort of applicants. Some schools in my sample were phased out before then. For these reasons I rely on conventional OLS VAM estimates of school quality and I provide evidence that conventional and risk-controlled VAM measures in this setting are largely equivalent.

[15]There are additional reasons to prefer measures of school effectiveness based on Regents rather than SAT. First, not everyone takes the SAT while the great majority of students take Regents exams. Second, school accountability measures in NYC have always been based on Regents test scores rather than on SAT, which also resulted in a stronger correlation between school letter grades and Regents VA as compared to SAT value added, as described in greater detail in the next section.

scores of both races. As noted above, OLS SAT VA is more biased, but this is true regardless of whether VA is estimated by race or on the pooled sample, suggesting that bias is unrelated with heterogeneity in treatment effects by race.

Measures of Demand for Schools To measure demand for schools, I use the rank-ordered lists submitted by 9th grade applicants. For the reduced form analysis, I aggregate individual level choices to school shares at the level of student application cohort by demographic cell, defined by the combination of student race, baseline test score tercile and residential borough.[16] School shares, denoted by s_{jtc}, measure the share of applicants in cohort t and belonging to demographic cell c that rank school j among their first or first three choices, depending on the case.

Analysis Samples and Descriptive Statistics I build two main datasets. The first includes student-level data on high school applicants from 2006 to 2016 and the second is a yearly panel of high schools.[17] The student analysis sample includes high school applicants applying for enrollment in 9th grade in NYC public schools with baseline (middle school) demographic, test scores and address information. I exclude special education students because they participate in a fully separate school match with a different set of programs. I use this set of students to study patterns of high school choice, student achievement outcomes, and to construct measures of school value added and peer quality.

Table 1.1 describes the students in my sample, their choices and achievement outcomes. The district serves a racially mixed and disadvantaged urban population, with over 77% of students eligible for free or subsidized lunch. Throughout the analysis, I compare Black and Hispanic students (labeled as “Minority") to white and Asian students (labeled as “Non-Minority"). Panel B shows that school choice attributes are very similar within this binary race definition and significantly different across the two groups, supporting the decision to divide students along this line. On average, white and Asian students choose schools enrolling higher achieving peers, a higher share of non-minority students and with 5 p.p. higher graduation rates. They also choose higher quality schools, as measured by value-added: their choices rank 14 and 19 percentiles higher in the distribution of Regents and SAT VA. Panel C restricts the sample to

[16]While I could also use choice indicators to study demand, the size of the data would explode considering that I should consider the cartesian product of 60,000 applicants and 350 schools in each of the 11 years of data. Moreover, the variation in quality signals is at the school-year level and student characteristics could only be useful for increasing precision and for studying heterogeneous responses.

[17]School attributes in a year refer to what would have been observed at the time of application. For instance, peer quality is measured using students enrolled at the school in the year before applicants enroll in 9th grade.

students enrolling in a NYC public high school with non-missing education outcomes.[18] High school education achievement also varies greatly by race. White and Asian students' SAT Math (English) scores are 1σ (0.8σ) higher than those of minority students. Non-minority students are also 17 p.p. more likely to graduate in time, and 23 p.p. more likely to enroll in college.

Table 1.2 provides descriptive statistics about the school panel, focusing on the years when letter grades were issued. The first two columns pool all school-years together, while columns (3) to (6) split observations by letter grade.[19] Schools receiving higher letter grades enrolled larger shares of white students, fewer students eligible for subsidized lunch, and have higher peer quality compared to schools receiving lower grades. On average, they are also of higher quality. Schools receiving an A have a 0.25σ higher Regents VA than schools receiving a C, D or an F. Scatter plots of school value-added against progress report quality score in appendix Figure A.1.2 show that the two measures are indeed positively correlated. The quality score was more positively correlated with Regents VA than with SAT VA, since it was primarily based on Regents performance and it did not use SAT scores. The last three columns of table 1.2 compare the characteristics of schools by grade, if grades had been assigned based on Regents Math VA alone.[20] Classifying schools correctly using causal estimates of quality would have resulted in twice as large differences in Regents VA between grade A and low grade schools.

1.2.3 Documenting the Race Quality Gap

On average, Black and Hispanic high school applicants choose lower quality schools. Here I document that this gap cannot be entirely explained by differences in the set of schools attainable due to residential segregation or differences in baseline achievement and that the gap narrowed after the introduction of school letter grades.

Panel (a) of figure 1-1 illustrates racial differences in student top choices and in the best schools within their attainable options. The graph plots the relation between applicants' baseline achievement and the average quality percentile ranking of their top 3 choices or the average of the best three schools in a student's "feasible" choice set, by student race. I construct a student's feasible choice set to include the schools reachable within 38 minutes by public transport - the mean student commute - in which the student had a non-zero probability of admission.[21]

[18]The samples used to study outcomes exclude students enrolled at the nine specialized high schools because they admit students via a separate process.

[19]Because the same school might receive different grades in different years, an observation is a school-year.

[20]I re-assign letter grades to schools based on the school Regents Math value added ranking, keeping the distribution (the count) of letter grades within years constant.

[21]These are all schools at which the student has below marginal priority, or schools where she has marginal priority that use lotteries to admit students or use academic screens that weren't binding for the student in that

Across the distribution of baseline achievement, Black and Hispanic students choose schools with 8 percentiles lower quality than white and Asian students. Differences in the maximum attainable quality by race and baseline achievement are negligible, therefore minority and lower achieving students are leaving more value-added on the table.

To better understand if differences in observable student characteristics can explain this gap, I estimate the following equation:

$$Q_i = \alpha + \beta M_i + X_i' \gamma + \epsilon_i \tag{1.2}$$

. Q_i is the mean value added of applicant i's top three school choices (or in the school of enrollment, for comparison), M_i indicates Black and Hispanic applicants, and X_i is a vector of controls. Table 1.3 shows estimates of the coefficient β. The first column reports raw differences: Black and Hispanic students, on average, choose schools that have 14 (18) percentiles lower Regents (SAT) VA. These differences translate into enrollment gaps and contribute to achievement disparities: if minority students attended the same schools as their white peers, they would have 0.1σ higher test scores. A first candidate explanation for this gap is that Black and Hispanic students live in neighborhoods with lower quality schools and traveling to better schools would be too costly. Differences in residential locations (zip codes), however, only account for a third of the differences in value-added of top choices.[22] Another explanation is that Black and Hispanic students may not meet the test score criteria for high-quality schools, rendering their applications to such schools futile. Even after controlling also for baseline achievement, however, 35% of the choice gap remains unexplained. The gap unexplained by disparities in available schooling options is likely to be larger, because residential zip code and baseline test score may be associated with differences in information and school preferences. In columns (4) and (8) I directly control for the mean quality and the quality of the best three schools in students' feasible choice sets as sufficient statistics for differences in geographic proximity to schools and availability of school options due to academic screening. Differences in attainable schooling options only explain between 25% and 30% of the gap.

School quality measures that do not account for variations in school effectiveness across

year, meaning their test scores are higher than the minimum score among admitted students with marginal priority. This definition does not take into account differences in non-zero probability of admission across race, which may matter if students ranked schools strategically. However, differences in application behavior due to strategic concerns should be minimized in this context for two reasons. The first is that most students do not fill all 12 positions in their list, and truthful ranking is a dominant strategy in this situation. The second is that I am considering only students' first three choices. Even for applicants constrained by the 12 choices cap, it would be rational to rank high-quality options with a non-zero probability among the first three choices rather than safe bets.

[22] Zip codes in NYC correspond to relatively small geographies. There are 204 different zip codes values in my sample.

student race may be potentially missing whether students are choosing schools that are a better match for their demographic group. The discussion in section 3.3 suggests that these concerns should be limited. Table A.2.22 confirms that choice differences are remarkably similar, and if anything larger, when using a measure of value-added that varies by race.

Panel (b) of figure 1-1 instead plots the change in the quality of applicants' top three choices relative to the mean of the 2007 cohort, separately by race.[23] These trends show that the previous regressions were masking substantial changes in choices over time. While everyone's choices improve relative to baseline, changes are larger for Black and Hispanic students. The raw school choice racial gap in 2007 is 19 percentiles of Regents VA, but it shrinks up to 13 percentiles during the study period. The increase in chosen quality is more marked during 2010-2014, the years when letter grades were printed on the high school directory. The trend reverses in 2015, when letter grades where removed.[24] These patterns suggest that letter grades might have played an impact in directing choice towards higher quality schools, especially among Black and Hispanic applicants. Changes in SAT VA of top choices, shown in figure A.1.3, follow a similar but less pronounced pattern.

What, then, could drive the cross-race choice gap? One explanation is differences in school preferences: white and Asian students might value school quality, or other school attributes correlated with it, more than Black and Hispanic students. In particular, preferences for schools enrolling similar students could be behind the choice gap, since in my sample higher quality schools enroll more white and higher achieving students.[25] An alternative hypothesis is differences in information about school quality. The trends documented in figure 1-1 suggest that lack of information about quality might indeed be an issue in this setting. As a first step to distinguish between these two explanations, I survey families who had just applied to NYC high schools asking them to situate real schools within the quality distribution of their residential borough.[26] Answers could vary from 1, corresponding to the worst 25% of schools, to 4, for the best 25%.[27]

[23]The percentile position is measured using the school relative ranking within the high schools participating in the high school match in that year to keep the measure comparable across years.

[24]In 2017 the DOE introduced an online search engine - the *School Finder* - on the high school admissions website that simplified the information search process. The 2017 and 2018 cohorts also took part in a large RCT that provided information about high school graduation rates conducted by Cohodes et al. (2022). To keep the information environment comparable across years, I truncate my study period to 2016.

[25]The rank-rank correlation coefficient between school quality and share of non-minority students is 0.38 and the one between school and peer quality is 0.51.

[26]The exact text of the question read: "How well does *school name - (school code)* prepare students for their Regents exams compared to other schools in your borough?". The distribution of school VA within each borough is essentially a replica of the distribution of VA within the city and most students rank schools in their borough among their first three choices.

[27]I randomize schools across respondents, sampling among relatively well known schools situated close to the

In table 1.4, I study the relationship between elicited beliefs and school quality and how this varies across respondents' race. Beliefs are positively correlated with value-added, and more so for white and Asian students, though the race difference is not statistically significant. The correlation between beliefs and school achievement levels is even stronger, and is also significantly higher among white and Asian respondents. When controlling for both achievement levels and VA, beliefs are positively correlated only with the first, which is consistent with the findings of Abdulkadiroğlu et al. (2020) who document similar patterns in measures of revealed preferences for schools. These correlations suggest that families rely on easily observed school attributes, such as average student achievement, to form opinions about a school's quality as it might be difficult for them to separate value-added from the composition of a school's student body (Rothstein, 2006; Ainsworth et al., 2023). By design, the schools that respondents have to assess are not statistically different across respondent race after controlling for district of residence. The different responses by race are thus only the result of differences in perceptions about identical schools.[28] White and Asian students appear to interpret achievement levels as a stronger signal of school effectiveness, which results in a slightly stronger correlation between their beliefs and actual quality.

Figure 1-2 plots the distribution of responses by respondent race for schools with achievement levels above and below the median. Most parents select the middle response, in line with a Bayesian model of belief formation in which families shade their evaluations towards the city mean when observing imprecise signals about school quality. Through the lenses of this framework, white and Asian respondents seem to receive signals of school quality that are either more strongly correlated with peer achievement, more precise, or both.

1.3 Information and Choice

1.3.1 Applicants Respond to the Introduction, Changes and Removal of School Letter Grades

Grades might be correlated with demand for reasons unrelated to quality. By exploiting within-school changes in grades and demand, I establish that information about quality has a causal effect on demand for schools. I use two empirical strategies.

respondent's address. More information on the survey and the selection of schools for this specific question is provided in appendix C.2.

[28]Appendix table A.4.25 confirms that schools populating respondents' questions are observably identical by showing balance of school value-added and achievement levels across respondent race, also conditional on the other school attribute.

The first compares changes in demand before and after the introduction or the removal of quality ratings across schools consistently receiving the same letter grade. I divide schools into 4 categories: *Type A* schools are those receiving a grade of A in at least 5 out of the 7 years of school quality reports; *Type low* schools receive a low grade, C D or F, in at least 5 out of the 7 years; *Never graded* schools are those that were never graded; all remaining schools are pooled in the residual category of *Type Average* schools.[29] This classification into types is fixed over years, which allows me to compare choices for the same set of schools over time even in the absence of letter grades.

Figure 1-3 plots the raw trends in the average share of students ranking a school in their top three choices ("school share") by school category between 2006 and 2016. The vertical lines indicate the introduction of letter grades online, on the school directory, and their removal. There is a marked substitution away from Type Low schools in favor of Type A schools after the introduction of letter grades, which affects for the first time the 2008 cohort of applicants.[30] The shares diverge substantially especially when letter grades are introduced on the school directory in 2010, while the trend reverses immediately after their removal in 2015.

An event-study model isolates grade effects over time. This can be written:

$$s_{cjt} = \sum_{L} \sum_{\tau=2006}^{2014} \beta_L^{t=\tau} (D_{jL} \times \lambda_{t=\tau}) + X'_{jt}\gamma + \mu_{ct} + \alpha_{cj} + \epsilon_{cjt} \tag{1.3}$$

. Rather than comparing demand to one of the four categories of schools as in a standard differences-in-differences, I estimate a fully interacted model of year dummies $\lambda_{t=\tau}$ with dummies D_{jL} indicating whether school j belongs to letter category $L \in \{$*Type A, Type Average, Type Low, Never graded*$\}$. I normalize $\beta_L^{t=2007}$ to zero for all letter categories, so that the coefficients of interest, β_L^t, captures the average within-school change in shares among applicants' top three choices, for schools of category L in year t relative to their level in 2007. For the effect of removing letter grades from school directories in 2015, I estimate similar regressions using years between 2011 and 2016, normalizing the coefficients $\beta_L^{t=2014}$ to zero.

I control for unobserved preferences for school characteristics that are fixed over time with school-cell fixed effects α_{cj}, and for a time-varying vector of school attributes X_{jt} that families

[29]There are 74 schools in the Type A category, 38 in the Type low, 273 are Average schools and 135 are never graded.

[30]The 38 schools that were always receiving low grades were large low performing schools evenly distributed across the four main boroughs of the city. Their median size in 2007 was 887 students, almost three times the median school size in that year (332), and their total enrollment share in the city was around 19%. 10 of them were closed between 2013 and 2016 , while those that remained open experienced a drop in enrollment of around 63%. Their median size in 2014 was 582, and their total enrollment share in the city only 8%.

could easily observe. This includes measures of student achievement (average student performance on English and Math Regents exams, graduation and college rates) and the share of white and Asian students enrolled at the school. To account for the fact that average Regents performance, graduation and college rates were not always printed on the school directory, I also control for their interaction with an indicator for years when these statistics were included in the printed school directory.[31] Finally, I include fixed effects for combinations of cell-year, μ_{ct}, to account for possible changes in the set of schools available due to opening and closure of schools that might affect demand differently across demographic cells and years.[32] Standard errors are clustered at the school-year level.

Figure 1-4 plots the coefficients β_L^t from estimation of equation (1.3) for years around the introduction or the removal of letter grades and table A.1.3 summarizes them in pooled pre-post coefficients. The plots paint a consistent picture: choices respond positively to the introduction of positive quality signals and negatively to the introduction of negative signals and with a reverse sign when these signals are removed. The demand for schools consistently receiving an A increased by 26% while that for Type Low schools dropped by 66% when considering pooled pre-post changes. Changes after the grades removal are smaller compared to when letters were introduced, especially for Type Low schools. These patterns are consistent with student learning and sticky school reputation, particularly in the form of stigma associated with being marked as a bad school for a long time.[33] The figure also confirms that trends in demand of the different school categories were parallel prior to the introduction of letter grades (and are similarly parallel after their removal).[34]

The second strategy studies the effects on demand of year-to-year changes in grades. This strategy similarly exploits within-school changes in signals and demand, but focuses on the variation coming from schools receiving different grades over the years. I regress school shares among student top choices on letter grade dummies, controlling for school fixed effects and

[31] Table A.1.1 summarizes the information printed on the school directories during the study sample.

[32] During my study period, 10 low performing schools (always receiving low grades) were closed and I observe 135 new small schools being opened. These were mostly small high schools, whose cumulative enrollment share grew up to 12% of the city-wide high school enrollment at the end of 2018.

[33] Another explanation could be the downsizing of schools consistently receiving low grades, which could affect demand through preferences for size.

[34] The parallel trends and the timing of the changes support the hypothesis that they are due to the introduction of school grades, rather than to the bundle of reforms introduced by the Bloomberg administration after it took mayoral control of the city schools in 2002. Moreover, there is no reason to believe that these reforms, including changes in the school district management and structure and teacher pay reforms, affected schools differently by the letter grade their received, so they are unlikely to cause the observed changes in choice.

other observable time-varying school attributes:

$$s_{cjt} = \sum_g \beta_g D_{jtg} + X'_{jt}\gamma + \mu_{ct} + \alpha_{cj} + \epsilon_{cjt} \tag{1.4}$$

.

D_{jtg} indicates that school j received a grade of g in year t. The rest of the notation remains the same as in equation (1.3). The identifying assumption that allows to interpret β_g as the causal effect of receiving a letter grade g on demand for a school, relative to not being graded, is a standard conditional independence assumption. Conditional on school and time fixed effects and on observable time-varying characteristics, letter grades (changes) are independent of unobserved (changes in) preferences for schools.

Table 1.5 shows the estimates of letter-grade effects, β_g, on school shares using choices of the 2010-2014 cohorts, who see letter grades on the school directories. Each school page in the directory typically reports two separate letter grades, from the reports of the two previous school years. In this table I only report estimates of the letter grade premia for the most recent quality report and leave to robustness checks the analysis of the effects of the two letter combination. The omitted category in columns (1) and (3) is school-years not receiving a letter grade, while columns (2) and (4) restrict the subset of school-years to those with a letter grade, which comprises older and larger schools.[35]

Letter grades shift demand for schools substantially: receiving an A increases the probability that a school is picked as a top choice by 0.15 p.p. on average, an increase of about 25% with respect to the average school share. Receiving an F reduces the probability a school is ranked as a top choice by 0.21 p.p., or 34% of the average school share. Receiving a grade of C is approximately equivalent to receiving no grade. Column (3) uses school log shares as left hand side variable, which yields consistent estimates. The effect of year-to-year changes in letter grades are consistent with the magnitude of the pooled pre-post changes following the introduction of high letter grades, but smaller than those following the introduction of persistently low grades. This suggests that bad reputation following a low grade may be more sticky over time than positive publicity from a high grade.

The table also reports the effect of graduation and college rates on demand, which is positive only when these are printed on the school directories, suggesting the existence of significant costs of searching for information about schools when this is not made easily available by institutional sources. The magnitude of visible school graduation rates is large and comparable

[35] Not all schools received letter grades in all years. This happened if the school had recently opened and/or the student sample size with achievement data was deemed too small to compute reliable quality score estimates.

to that of receiving a grade of A, at least when focusing on the subset of graded schools that are larger and more established (column 2). Visibility of college rates, on the other hand, has little to no effect on the demand for schools.

The effect of letter grades on choice behavior is informative of families' preferences for and beliefs of school quality under the assumption that applicants interpret letter grades as signals of school effectiveness. Features of the institutional setting support the validity of this assumption, because letters were framed as measures of causal school value-added. Sophisticated families, however, might have realized that letter grades reflected in part student selection or other school features different from value-added. Their responses to changes in letter grades might therefore not only be indicative of preferences for school quality and uncertainty in beliefs about it, but rather of preferences and beliefs for a mix of school attributes. While I cannot perfectly rule out this hypothesis, I argue that it is not very plausible by running a couple of tests. If families are sophisticated and know what goes in the quality score, they would be aware that it is made of different components. Therefore they should respond to changes in the score sub-components, potentially with different weights depending on their taste for these different attributes. In Appendix Table A.1.7 I show that, while within-school changes in the two largest components of the quality score are correlated with changes in demand for schools, the correlation is no longer statistically significant after controlling for letter grades. In fact, choices do not seem to react even to changes in the main underlying quality score: the positive correlation between demand and the score goes to zero (or if anything is slightly negative) after controlling for grades, suggesting that applicants only paid attention to the letter grades.

In appendix Tables A.1.5 and A.1.6, I extend the model in equation (1.4) to consider also the effect of the second letter grade reported on the school directory. Columns (4)-(6) of A.1.5 use a model that controls separately for the two letter grades received, showing they both have a separate effect on demand. The regression estimated in columns (7)-(8) substitutes letter grade fixed effects with an indicator for having two As and an indicator for receiving an A in only one out of the two quality reports considered in the school directory of the applicant's cohort. Receiving an A for two consecutive years has an effect on demand that is twice as large as the effect of receiving A only in one of the two years, suggesting that the two grades have an additive effect. Table A.1.6 compares this additive model to a more flexible one where demand may vary with each possible combination pair of letters. Estimates suggest that a model with additive effects yields estimates that are very similar to those of the more flexible model and that overall families put more weight on the most recent signal.

1.3.2 Black and Hispanic and Less Informed Students Respond More

Some studies, including in education, have found larger responses to information provision among richer households, suggesting that barriers to information take-up may be larger among the least affluent or least well connected (Corcoran et al., 2022; Bhargava and Manoli, 2015; Bergman, 2020). A standard model of Bayesian updating would instead predict larger responses among less connected and less well informed students. This discussion motivates studying heterogeneity in responses to letter grades across student demographic and socio-economic background.

Panel (b) of figure 1-3, plots raw trends in shares by race and school letter type. White and Asian students choices appear already more aligned with letter grades in the years prior their introduction, ranking Type A schools (Type low schools) among top choices 50% more often (60% less often) than minority students. These patterns suggest that Black and Hispanic students may have been less informed before the introduction of grades. This could also explain why school shares among Black and Hispanic students appear to vary substantially more following both the introduction and the removal of letter grades.

To measure whether these differences in responses are statistically significant, I extend equation (1.3) and estimate a triple difference model in which I interact a dummy M_c indicating Black or Hispanic student covariate cells with year and school category indicators to estimate the differential effect of the introduction and removal of quality signals by race:

$$s_{jct} = \sum_{L} \sum_{\tau=2006}^{2014} \left(\delta_L^{t=\tau} (D_{jL} \times \lambda_{t=\tau} \times M_c) + \beta_L^{t=\tau} (D_{jL} \times \lambda_{t=\tau}) \right) + X'_{jt}\gamma + \mu_{ct} + \alpha_{cj} + \epsilon_{jct} \tag{1.5}$$

.

I normalize the baseline difference in share by race $\delta_L^{t=2007}$ to zero for all school categories L. Figure 1-4 plots δ_L^t separately for the two natural experiments of introducing and removing grades.[36] Regression estimates confirm that average responses to changes in information mask substantial racial heterogeneity: Black and Hispanic choices respond substantially more to both the introduction and the removal of letter grades than those of white and Asian students, suggesting that information frictions may be larger among the former group of students.

Figure 1-6 and Table A.1.4 explore heterogeneity in responses to year-to-year changes in grades estimating equation (1.4) on race-specific school shares, which yields consistent findings. Receiving an A increases demand by 30% (of the average school share) among Black and Hispanic students, but only by 14% among white students. Symmetrically, receiving an F grade

[36] Appendix figure A.1.4 plots event study estimates of coefficients β_L^t in equation (1.3) separately by race.

decreases a school share among Black and Hispanic choices by 48% while it has non statistically significant effect on the choices of white and Asian students. Cross-race differences in letter grade premia on log shares (columns (7)-(12)), however, are much smaller compared to the effects on share levels. This means that white and Asian student choices are more responsive to changes in letter grades when focusing on the set of schools that they choose at higher rates. In other words, white students may hold stronger preferences for the school attributes of a small sample of schools, which makes them less responsive to changes in perceived school quality on average.

Even within students of the same race, some may be better informed and respond less to information disclosure. If applicants used Bayes' rule to update their priors about schools, we should expect larger responses among the ones most surprised by the quality signals embedded in the letter grades. I use a lasso regression to classify students as more or less exposed to new information based on covariates related to their middle school and neighborhood of residence.

Using student choices before the introduction of grades, I estimate which student covariates predict concordance of choices with school grade types. Alignment of choices with letter grades is measured by the variable $Information\_index_i$, which takes values in {-1, 0 , 1}. Each applicant starts from a value of 0 and gets a point if she ranked in her top three choices a Type A school, and is subtracted a point if she ranked a Type low school. Differences in $Information\_index_i$ across students may stem from unequal access to information, but may also indicate differences in informed preferences for schools. If they were only due to differences in preferences, however, the introduction and removal of grades should not affect the choices of families with covariates predicting their preference for low grade schools. Finding that information affects students with lower values of $Information\_index_i$ is another test of the existence of imperfect information about school quality and one that may suggest larger misinformation among certain students.

I estimate which student covariates are most strongly predictive of this information index using a lasso estimator and a random 50% subsample of applicants in 2007:

$$Information\_index_i = \alpha + \lambda_{MS(i)} + \lambda_{z(i)} + \lambda_{d(i)} + X_i'\beta + \epsilon_i \tag{1.6}$$

The right hand side covariates capture potential reasons why households before the introduction of letter grades might have been more or less informed about school quality, such as exposure to different social networks. They include middle school fixed effects $\lambda_{MS(i)}$, residential school district fixed effects $\lambda_{d(i)}$, and zip code fixed effects $\lambda_{z(i)}$. The vector X_i also includes a gender indicator, an indicator for subsidized lunch eligibility and one for English language learners, the share of students in the same middle school from the previous cohort ranking Type A schools and Type Low schools among their first choices.

I use the estimates to predict $Inform\hat{ation\_index_i}$ for all remaining students in the applicant sample. I split the sample of applicants in a control group and a treated group, the latter including students whose predicted information index is below the median predicted value in 2007. This group, denoted with a dummy $Treated_i$, should be more surprised by the introduction of letter grades. I use this dummy to construct school shares at the level of demographic cells defined by combinations of $Treated_i$, race, baseline achievement and residential borough.

Appendix table A.1.8 reports summary statistics by treatment status. Treated students are more likely to be Black or Hispanic, which is not surprising in light of the evidence presented above. However, there is substantial heterogeneity in exposure to the grades information even within races. Treated students are more likely to be English language learners, eligible for subsidized lunch, and have lower baseline test scores. They are more likely to live in the Bronx, and less likely to live in Manhattan. They live in neighborhoods with more Type low schools seats and attend middle schools were students in 2006 were more likely to apply to Type low schools and less to Type A schools.

Appendix Figure A.1.6 plots estimates of δ^t_L for a version of equation 1.5 that considers heterogeneity along treatment status rather than across race.[37] The results confirm that demand responses to the introduction and removal of letter grades were larger among students who appeared less informed, as predicted on the basis of covariates related to exposure to different social networks. These patterns are similar for minority and non-minority students. Importantly, regression to the mean is not a concern here because I am using a different sample for prediction and for estimation and because I can also look at responses right after the removal of letter grades, when we should expect a reversal of demand back to the status quo.

Choice responses to information about school quality imply that families value school value-added but hold imperfect information about it. That is, they have uncertain and possibly inaccurate beliefs about the quality of schools. Differences in responses to information across race or other demographic characteristics, however, could be explained by multiple factors. The first is differences in prior information about school quality, in the form of larger uncertainty or larger differences between average prior beliefs and quality ratings. Information disparities by income and race may play an important role in this context. With more than 700 programs to choose from, parents complaint that the high costs of searching for information about schools

[37]Appendix Figure A.1.5 instead reports estimates of β^t_L for versions of equation (1.3) that separately use students more or less exposed to new information within race.

often results in unequal access to information.[38] The larger responses among less informed students, as predicted on the basis of covariates related to exposure to different social networks, suggest a role for information gaps.

The second is heterogeneous preferences for school quality relative to other school attributes, such as distance or the demographic composition of the school student body. Informing families that a school has high value-added may not be enough to convince them to apply if they value school effectiveness little or if the school is undesirable for other reasons. In the next sub-section, I consider more in depth how differences in beliefs and preferences across schools might have influenced choice responses to information. This discussion further motivates the need to estimate the joint distribution of school preferences and beliefs, along with its heterogeneity across students, an exercise I undertake in section 1.5.

1.3.3 What School Attributes Influence Reactions to Quality Signal?

Preferences for Other School Attributes Mediate Intensity of Responses To gain insight on how information about quality interacts with preferences for other school attributes, I consider how the magnitude of demand responses varied across schools with different observable characteristics. Rather than estimating separate heterogeneous coefficients for each letter grade, I map letter grades into a discrete quality signal index S_{jt} that varies from 1 (corresponding to an F) to 5 (corresponding to an A) and I interact this index with school attributes X_j. To test whether cross-race differences in responses to information for different schools are statistically significant, I also interact these regressors with a minority dummy M_c. The resulting estimating equation is:

$$s_{cjt} = \beta_0 S_{jt} + \delta_0(S_{jt} \times X_j) + \beta_1(S_{jt} \times M_c) + \delta_1(S_{jt} \times X_j \times M_c) + \mu_{ct} + \alpha_{cj} + \epsilon_{cjt} \qquad (1.7)$$

. As before, I control for school fixed effects, so that demand responses are identified off of within-school changes in letter grades. X_j indicates that school j enrolls a high share of white students, has high peer quality or high average Regents test scores, depending on the specification.[39] Table 1.7 shows that white and Asian student demand for schools responds to changes in letter grades, S_{jt}, only within schools enrolling high shares of white and high achieving peers. Choices of Black and Hispanic students are more responsive on average and equally responsive

[38] For instance, many parents notice how having the means to afford taking days off from work to attend information sessions or open houses or hire admission consultants may result in unequal access to information (Corradini and Idoux, 2023).

[39] These are indicators for being in the top third of the city distribution of these three dimensions (share white, peer quality, achievement levels).

to changes in the signalled quality of schools with different attributes.

These findings also shed light on the reason why heterogeneity in log share responses to letter grades across applicant race was much smaller than in the specification using share levels in table A.1.4. They align with the view that white and Asian student choices are concentrated on a small subset of schools and that they respond to information only within these schools. Preferences for the demographic composition of the students at a school might be responsible for these patterns, limiting choice changes within a specific set of schools that also have these other desirable characteristics.

Beliefs Mediate Choice Responses to Positive and Negative Signals The survey data presented in section 2.2.2 indicates that families' school quality beliefs are positively correlated with the mean Regents achievement level of students at a school. If applicants updated their beliefs according to Bayes' rule, we should observe larger increases in the demand of schools with lower achievement levels after they receive a high grade, relative to receiving none, because applicants believed they were of lower quality. And similarly, larger decreases in the demand of schools with higher achievement levels after they receive a low grade.

Table 1.6 presents estimates of equation (1.4) separately for schools with above and below median achievement levels at baseline, showing they are consistent with these predictions.[40] Positive responses to higher letter grades (A or B) are larger among lower performing schools, while the negative responses to letters D and F are larger for higher performing schools. Moreover, while the estimate of β_C is positive and that of β_D is non significantly different from zero for lower performing schools, they are negative for higher performing schools, suggesting that receiving a letter grade of C or D is perceived as a negative surprise only if the school had high achievement levels. Regressions using log shares on the left hand side (columns (3) and (4)) yield qualitatively similar results.

Appendix figure A.1.7 explores the same type of heterogeneity in responses to quality signals using the introduction and the removal of letter grades. It plots the average percent change in demand for Type A schools, distinguishing between schools with high or low achievement levels. Once again, the increase (decrease) in demand following the introduction (removal) of positive letter grades is larger for lower (higher) performing schools. There is no differential change, however, in responses to the introduction or removal of a persistent negative signal. Taken together, these estimates suggest that taking into account how households form and update beliefs seems important to predict the effects of counterfactual information disclosure

[40] I define a school baseline achievement level as the average yearly performance on Regents math exams of its students, taking a cross-year average for years before the introduction of letter grades on directories (2006-2009).

policies.

1.4 Consequences for Racial Inequality

This section studies the consequences of the larger response to grades of minority students' choices on racial inequality in education outcomes. I estimate the following event study regression:

$$Y_i = \sum_{\tau} \delta^{t=\tau} (M_i \times \lambda_{t=\tau}) + \mu_t + X_i'\gamma + \epsilon_i \tag{1.8}$$

.

The strategy is similar to that in section 1.3.1, but is adapted to student-level outcomes. M_i indicates Black and Hispanic students and $\lambda_t = \lambda_{t(i)}$ are, as before, cohort indicators. The vector of controls X_i includes ethnicity, gender, ell status, subsidized lunch status and fixed effects for combinations of student borough and baseline test score terciles to account for potential changes in the composition of students over time.

I first consider changes in attributes of top choices and compare them to those of school offers. Figure 1-7 plots δ^t estimates for four different regressions. Dependent variables are dummies indicating applying to at least one Type A or Type low school among the first three choices, or receiving an offer to these schools. Choices and offers of minority students improve after the introduction of grades relative to those of non-minority applicants. Minority students were up to 10 p.p. more likely than white students to rank Type A schools, and up to 15 p.p. less likely than white students to rank low grade schools among their top 3 choices, after the introduction of grades.[41] These large relative changes are not due to lack of room for improvement in white choices: only 64% of white students were ranking Type A schools before the introduction of letters (compare to 48% of minority students). After the removal of letter grades, racial gaps in choices partially return to their pre-letter grades levels.

The effect on racial gaps in offers, however, is smaller than that on choice gaps. Minority students were only 3 p.p. more likely to receive a Type A offer and only 4 p.p. less likely to receive a Type Low offer. In summary, while the racial gap in Type A choices closed by 43%, the corresponding enrollment gap was reduced only by 23%.[42]

[41] These gaps are somewhat smaller when focusing only on first choices, as shown in Panel A of Table 1.8. If everyone had been assigned to their first school choice, Black and Hispanic students would have been 5 p.p. more likely to enroll in a type A schools after the introduction of letter grades, compared to white students.

[42] Effects on enrollment schools follow closely the effects on offers and are not shown to improve the readability of the figure. Compliance is high and 73% of those offered a school in the first round of the match are enrolled at

Type A schools are more likely to use admission rules based on middle school test scores, which could explain why Black and Hispanic students had relatively lower chances of receiving an offer at these schools conditional on applying. Appendix Table A.1.9 compares the effect on the probability of receiving an offer to Type A or Low schools to the corresponding effect in simulated assignments that use student reports but assume that schools prioritize students only in order of their random lottery number. The relative improvement of minority students' simulated offers is larger by 1 p.p. than in reality, but still smaller than the relative change in choices.[43] Higher congestion in the grade-A schools chosen by Black and Hispanic students, as compared to the grade-A schools white students chose, must explain the remaining discrepancy. These results highlight how in markets with binding capacity constraints information interventions may shift everyone's choices but need not translate into large average achievement boosts. Understanding how assignment rules and capacity constraints clear the market in the presence of increased demand for high quality schools becomes important to gain insight on who are the winners and losers of information disclosure policies.

The first two columns of Table 1.8 summarize the findings of figure 1-7 in pooled diff-in-diff coefficients, while the remaining ones focus on different attributes of students' choices, school offers or enrollment schools. Everyone chooses and is matched to schools with higher value-added after the introduction of letter grades. In addition, the school quality of ranked and offered schools for minority applicants improves by an additional 4.5 percentiles (or 0.03σ) relative to white students, which corresponds to a 25% reduction in the baseline cross-race gap. Students also choose schools enrolling more white and Asian students and of higher peer quality, which are attributes positively correlated with grades. Figure A.1.8 plots regression estimates of trends in some of these attributes of top three choices and shows that the timing of the changes coincides with the introduction of grades on the school directory.[44]

Because the value-added of schools offered to Black and Hispanic students improved relative to that offered to white students, we should mechanically observe a reduction in achievement inequality. Appendix Table A.1.10 confirms that this is the case by reporting pooled diff-in-diff estimates of equation (1.8) on student achievement. Panel A compares changes in achievement across race, while Panels B and C do the same within race groups, across levels of the information exposure dummy introduced in section 1.3.2.

that school in June of their 9th grade.

[43]Both the removal of academic screening and priorities based on residential address could explain the discrepancy between realized and simulated offers.

[44]The bottom right panel also shows that, when letter grades were in place, students were more likely to rank schools outside their borough, suggesting that students may be willing to travel further to attend schools they perceive are of higher quality.

Minority students' Regents Math test scores improve relative to those of non-minorities after the introduction of grades, while there is no differential effect on SAT Math test scores. This is consistent with the higher correlation of the quality score with Regents VA than with SAT VA.[45] On-time graduation and college enrollment rates of minority students also improve by 5 p.p. and 7 p.p. respectively relative to non-minority students. Similar patterns are also present when looking at the diff-in-diff effects within race by heterogeneity in exposure to information: Regents scores, graduation and college rates of less informed students improve relative to the more informed.

I am cautious in interpreting the diff-in-diff estimates on racial inequality in achievement outcomes as being entirely driven by the introduction of school letter grades as this was a time of major reforms in NYC schools. These other reforms might have affected achievement outcomes differently for students of different demographics.[46] The magnitudes are in fact larger than the effect on offered value added: the relative improvement of Black and Hispanic Regents math is 0.06σ, twice as large as the relative change of offered Regents VA. This discrepancy indicates that the reduction in racial inequality is likely not entirely driven by relative improvements in reallocation to schools. One possible explanation is that the observed reduction in inequality could be the combination of a reallocation effect and a competitive pressure effect which may vary in intensity across schools within the same market. I leave the study of changes in supply-side incentives to future work.

1.5 A Model of School Choice with Imperfect Information About School Quality

This section presents and estimates a model of school choice with imperfectly informed students that leverages the variation in quality signals provided by the letter grade system. I use the model to separately measure student preferences for quality and their beliefs about it (Vatter, 2022) and to study the extent to which information design can reduce achievement gaps. More broadly, the model can shed light on which features of supply and demand for schools, including distaste for commuting, admission rules, and capacity constraints, interact with information provision to determine the equilibrium allocation of students.

[45] Effects on racial differences in offered SAT VA, which are not reported here, are in fact null.

[46] After mayoral control of the schools was authorized in 2002, mayor Bloomberg implemented a broad agenda of market-based reforms that included a re-organization of the school district management, teacher pay reforms, changes in the curriculum, large emphasis on math and English test scores and the opening of smaller high schools to replace low performing large ones.

1.5.1 Set Up

Applicant i's indirect utility from attending school j is additive separable in distance and linear in school characteristic. Students' choices depend on their expectation of school j's quality, which I denote with $E_{f_{c(Z_i)j}}[q_j|s_{jt}]$, and other known school characteristics. Students form expectations about the quality of schools using their prior belief $f_{c(Z_i)j}(q)$ and a quality signal s_{jt}, if available. This departs from standard models of school choice that assume students are perfectly informed about the quality of schools. The indirect utility from attending school j is:

$$u_{ij} = \underbrace{X'_{jt}\beta_{c(Z_i)} + \gamma_{c(Z_i)}E_{f_{c(Z_i)j}}[q_j|s_{jt}] + \xi_{c(Z_i)jt}}_{\delta_{cjt}} - \lambda_{c(Z_i)}d_{ij} + \epsilon_{ij} \tag{1.9}$$

, where δ_{cjt} denotes the average utility from attending school j for students of demographic cell $c(Z_i)$, applying to high school in year $t = t(i)$. Preferences and beliefs may vary across student demographic cells, which are defined based on the vector of covariates Z_i. In the empirical estimation, cells correspond to combinations of student baseline test score terciles and race (Black, Hispanic and white).[47] X_{jt} denotes observable characteristics of school j in year t, ξ_{cjt} preferences for unobserved school characteristics, and d_{ij} is the distance between student i's home and school j. ϵ_{ij} captures idiosyncratic tastes for schools. As it is standard in this type of models, I assume that unobserved tastes for schools ($\epsilon_i = (\epsilon_{i1}, ..., \epsilon_{iJ})$) are conditionally independent of distance to school given observed student and school characteristics as well as school unobservables: $\epsilon_i \perp d_{ij}|Z_i, \{X_{jt}, s_{jt}, \xi_{jt}\}_{j=1}^{J}$. This assumption rules out the possibility that students systematically reside near the schools for which they have idiosyncratic tastes.

The coefficient $\lambda_{c(Z_i)}$ is normalized to 1 for all students to specify a distance-metric utility function, an approach often adopted in the school choice literature (Agarwal and Somaini, 2020). This normalization does not impose restrictions on how students trade-off different school characteristics because the scale parameter of the structural error distribution can vary across cells $c(Z_i)$.[48] The utility of the outside option is normalized to zero $u_{i0} = 0$ for all students.

Following the literature documenting the existence of preferences for attending schools with students that are higher achieving and of the same race, the vector of observable school at-

[47] For the purposes of modeling heterogeneity in preferences and beliefs, I always pool Asian students with white students.

[48] That is, I assume ϵ_{ij} is distributed according to a type-1 extreme value distribution with location parameter equal to 0 and scale parameter $\sigma^{\epsilon}_{c(Z_i)}$. This assumption simply allows to express utility parameters in terms of the willingness to travel of each covariate group. It is effectively equivalent to normalizing the scale of the structural error across covariate groups, $\sigma_{c(Z_i)} = 1$, $\forall c$, and rescaling all preference parameters by the cell-specific distaste for commuting λ_c.

tributes X_{jt} includes peer quality and the share of white and Asian students enrolled at the school (Abdulkadiroğlu et al., 2020; Ainsworth et al., 2023; Rothstein, 2006; Hailey, 2022; Corradini and Idoux, 2023). Both variables are based on averages for the cohort enrolling in school j the year before applicants submit their applications. I further assume that preferences for school unobserved characteristics can be decomposed into a component that is fixed over time and one that varies over years, $\xi_{jt} = \tilde{\xi}_j + e_{jt}$, and that changes in the time-varying component are orthogonal to changes in letter grades. That is, re-writing δ_{cjt} as:

$$\delta_{cjt} = X'_{jt}\beta_c + \underbrace{\gamma_c E_{f_{cj}}[q_j|s_{jt}]}_{\eta_{cjs}} + \tilde{\xi}_{cj} + e_{cjt}$$

I am assuming, $E[e_{cjt}|X_{jt}, \eta_{cj}, \tilde{\xi}_{cj}] = 0$, where η_{cjs} are fixed effects for combinations of schools and quality signal (letter grades or their absence), and η_{cj} stacks the fixed effects relative to one school.

Students form expectations about the quality of each school based on their priors $f_{c(Z_i)j}(q)$ and on the quality signals s_{jt} provided by the school district, when these are available. These signals depend on the letter grades and therefore may vary depending on the year when i applies to high school. I assume that students applying to high school between 2010 and 2014 observe the letter grades and know the quality quantile cutoffs used to partition the quality space.[49] When observing letter grades, students update their beliefs according to Bayes rule.

In the main specification, I assume that priors are distributed as a truncated normal with mean μ_{cj} and standard deviation σ_{cj} over the space $[q, \bar{q}]$ of value-added in the city. Given this functional form, the expected quality of school j given the letter signal L is conveniently the mean of a twice truncated normal given by: $E_{f_{cj}}[q_j|s_{jt} = L] = \int_{\underline{c}_L}^{\bar{c}_L} \frac{f_{cj}(q)}{F_{cj}(\bar{c}_L) - F_{cj}(\underline{c}_L)} dq$, where $s_{jt} = L$ denotes that school j received a grade of $L \in \{A, B, C, D, F\}$, which implies that $q_j \in [\underline{c}_L, \bar{c}_L]$. Because the empirical distribution of value-added is bell-shaped and approximately normal, as shown in appendix figure A.3.15, this functional form restriction is equivalent to saying that students are more or less optimistic about the quality of a specific school with respect to the average school, and more or less certain about their opinion.

I restrict belief heterogeneity across schools by specifying prior moments as a function of school characteristics. In the main specification, priors are a function of the school average performance on Regents exams. This choice is motivated by the reduced form results presented in

[49] The cutoffs in terms of percentiles and quantiles of the quality score distribution were clearly communicated, as can be seen from the progress report in figure A.1.1. I am assuming that students knew these cutoffs, an assumption often adopted in papers using similar methods (Vatter, 2022; Barahona et al., 2023b; Dranove and Sfekas, 2008; Chernew et al., 2008).

section 1.3.3 and the positive correlation between the beliefs elicited with the survey and school achievement levels. I relax this assumption in alternative specifications, allowing priors to be a function of school value added, in addition to school average achievement levels. In the main specification, prior moments vary across discrete school types defined based on achievement levels.[50] In robustness checks, I instead let the prior mean and variance parameters be a linear continuous function of achievement levels or of achievement levels and school value added, rather than a non-parametric function of discrete school types. These restrictions do not require the prior to be accurate but they reduce the dimensionality of the parameters that I need to estimate.

This model can explain patterns observed in the reduced form analysis. As long as students hold non-degenerate priors and positive preferences for quality, their choices would respond positively to high letter grades and negatively to low ones. Following Bayes' rule, beliefs and choices change more if signals of school quality are more surprising, which could explain the larger positive responses to high grades for schools with low achievement levels. Differences in responses to grades across students of different backgrounds could be explained by differences in preferences for quality relative to other school attributes, or differences in belief precision and bias.

Microfounding Heterogeneity in Prior Beliefs Across Students Why would different students hold different priors? One source of heterogeneity across families of different backgrounds is differences in access to information through sources like social networks, school admission consultants, guidance counselors or different use of online websites. For simplicity, I will refer to these additional sources as social networks. Formally, assume that before receiving signals from their social networks, all students hold similarly uninformed priors $f_{ij}(q) = f^U(q)\ \forall i, j$, identical to the distribution of quality in the city. Each student i receives a signal from her social network about the quality of school j, n_{ij}, before observing any rating of quality provided by the policy maker. The parameters governing the distribution of students' beliefs will depend on the mean and precision of the signals they receive from their social network. Considering the simple case in which the distribution of school quality in NYC is well approximated by a normal $f^U(q) \sim N(\mu_q, \sigma_q)$, and students receive social network signals that are also normally distributed, $n_{ij} = \tilde{\mu}_{ij} + e_{ij}$, $e_{ij} \sim N(0, \sigma_{e_{ij}})$, the resulting belief about the quality of school j, f_{ij}, will be also normally distributed. Its mean μ_{ij} and variance σ^2_{ij} depend on the social network

[50] That is, $f_{ij} = f_{ij'}, \forall jt, j't'\ s.t.\ h(R_{jt}) = h(R_{j't'})$, where $h(R_{jt})$ is a function defining a school type based on its achievement level R_{jt}.

signal as follows:

$$\mu_{ij} = \mu_q + \frac{\sigma_q^2}{\sigma_q^2 + \sigma_{e_{ij}}^2}(\tilde{\mu}_{ij} - \mu_q)\,, \qquad \sigma_{ij}^2 = \frac{1}{\frac{1}{\sigma_q^2} + \frac{1}{\sigma_{e_{ij}}^2}}$$

. Students receiving more precise social network signals (smaller $\sigma_{e_{ij}}$) will have less uncertain beliefs about the quality of school j. The social network signal mean and precision will also govern how far students believe the quality of j is from the average school in the city.

1.5.2 Estimation and Identification

I adopt a two-step estimation procedure similar to the one in Goolsbee and Petrin (2004). In the first step I use students' rank-order lists to estimate δ_{cjt} with maximum likelihood. In the second step, I use a minimum distance estimator to decompose the mean utility $\hat{\delta}_{cjt}$ in its main components. I only leverage time-series variation within utility for attending a school to remove systematic preference for specific schools that are fixed over time. The resulting estimator is:

$$\min_{\theta_c} \sum_j \sum_t \sum_{\tau > t} (\Delta\hat{\delta}_{cjt,\tau} - \Delta X_{jt,\tau}\beta_c - \gamma_c \Delta E_{f_{cj}}[q_j | s_{jt}, s_{j\tau}])^2$$

The parameter vector $\theta_c = \{\beta_c, \gamma_c, \boldsymbol{\mu_c}, \boldsymbol{\sigma_c}\}$ varies with student demographic cell c and includes the parameters governing preferences for school characteristics and the vector of prior moments. I recover estimates of $\tilde{\xi}_{cj}$, the time-invariant component of preferences for school unobserved attributes, from average residuals: $\hat{\tilde{\xi}}_{cj} = (\sum_t \hat{\delta}_{cjt} - X_{jt}\hat{\beta}_c - \hat{\gamma}_c \widehat{E[q_{jt,\tau}]})/T$.

In the estimation, I focus on applicants enrolling in 9th grade between 2011-2015, relying on variation in letter grades within years (2011-2014) and on the removal of letter grades in 2015. I use 2016 applicants to assess model fit out-of-sample and as the basis to simulate counterfactuals.[51]

Identification of first step estimates of the mean utility δ_{cjt} rely on standard revealed preference arguments valid under the assumption that students rank schools in order of true preference and that the ranked schools are preferred to the outside option. The truthful reporting assumption is often maintained when DA is used to allocate students to schools, because it is

[51] I could also rely on the introduction of letter grades as an additional source of variation, but choice shifts following the grades removal in 2015 are more informative of student beliefs in 2016. Differences in choice responses to the introduction and the removal of grades suggest that some learning occurred, potentially due to some stickiness in reputation. Using 2006-2014 to estimate the model yields similar estimates, except for prior means. Minority students' beliefs before the introduction of grades were negatively correlated with the school achievement levels. This is no longer the case today, as validated by the survey evidence. I also exclude 2010 because the directory in 2010 does not show graduation rates and I want the information environment to be the same in all years, except for changes in grades. Nevertheless, adding 2010 to the sample changes estimates very little.

strategy-proof when applicants are allowed to rank every school. Even if the number of school choices is capped at 12 in NYC, most students do not fill their list, and truthful ranking is a dominant strategy also in this situation. I relax the trutful reporting assumption in a robustness check in table A.1.17.

The identification of preferences for school quality and prior beliefs in the second step requires instead a more careful discussion. Letter-grade demand premia for different schools can be estimated as part of fixed effects for combinations of grades and schools, which are identified from willingness to trade distance for higher letter-grades, all else equal. However, simulating the equilibrium effects of counterfactual information disclosure policies demands additional structure in order to understand how families update beliefs under scoring designs that use different quality cutoffs. The challenge is to tell whether students are willing to travel further to enroll in schools receiving higher letter grades because they believe the quality difference is small but very valuable (i.e., γ is large) or because they value quality little but they are updating their quality belief a lot, for instance due to large uncertainty or large biases (i.e., γ is small). I adapt the argument used in Vatter (2022) to my setting and show in appendix C.4 that within-school changes in letter grades over the years would lead these two configurations to generate systematically different choices. As in Vatter (2022), I maintain the assumption that the payoff from quality enters the utility linearly and, as already discussed, that students understand the school letter-grade cutoff structure.

The intuition behind this argument is that the assumption that families know the letter grade cutoffs implies bounds on belief updating. In turn, this implies bounds on preferences for quality, given students' willingness to commute for increments of letter grades. For instance, the quality score cutoffs used to assign letter grades imply that the quality of B-schools is bounded between $[-0.4\sigma_q, 0.5\sigma_q]$ while the quality of D-schools is in $[-1.4\sigma_q, -1\sigma_q]$, where σ_q denotes standard deviations of the distribution of quality across schools. If the willingness to commute for a B-school, relative to a D-school is 5 minutes, simple algebra shows that the change in expected quality is bounded between $\Delta E[q] \in [0.6\sigma_q, 1.9\sigma_q]$. This implies that γ can be bounded between $[2.6, 8.3]$ minutes. Variation in letter grades (and their absence) for the same school (type) generates additional bounds. With sufficient variation in the quality signals that a school receives, priors would be non parametrically identified. In practice, because the variation is limited, assuming a functional form for the priors helps identification. Finally, identification of preferences for time-varying school characteristics, namely peer quality and the school racial composition, comes from within-school changes in these characteristics over the years.[52]

[52] Peer quality is highly correlated with a school average achievement level, R_{jt}. If peer quality and achievement levels change together from year to year, both school preferences and beliefs about its value-added change. Thus

1.5.3 Estimates and Model Fit

Table 1.9 summarizes the model estimates. It shows weighted averages of cell-specific estimates across cells sharing the same covariate (race or baseline test scores), using weights proportional to cell size. The cell-specific estimates and their asymptotic standard errors are reported in Appendix Table A.1.11.[53] Panel A reports summary statistics of the first step estimates of the mean school utility δ_{cjt}. Mean school utilities are positively correlated with peer and school quality, more so for white and higher achieving students. Their within-cell standard deviation ranges from 21 to 28 minutes of public transport commute.

Panel B reports the second step estimates of the preference parameters $\gamma_c, \beta_c, \tilde{\xi}_{cj}$ and panel C of the prior means μ_c and precisions σ_c^{-1}. These estimates are for the simple case in which priors vary across a binary school type, which depends on whether the school is in the top third or bottom two thirds of average Regents performance. I describe estimates of more complex models where beliefs depend on more discrete school types or on a continuous function of school characteristics below. All models yield largely consistent estimates.

Estimates offer two explanations behind differences in responses to information across students. The first is differences in beliefs, in the form of smaller bias and lower uncertainty in white students' priors. Estimates are overall consistent with the survey evidence of section 2.2.2. Everyone believes that schools with higher achievement levels are of higher quality, but the average beliefs of white, Hispanic and higher achieving students are more strongly correlated with mean achievement levels, and therefore with value added. White students also appear to be more certain that schools with higher achievement levels are of higher quality. These findings align with the interpretation that white and higher achieving students receive more precise signals about school quality from their social network and these signals are based on student achievement levels at the school. Overall, however, all students tend to be quite uncertain and mis-informed about school quality. Average priors are indeed close to the mean quality in the city for schools with both high and low achievement levels.

The second explanation, quantitatively more important, is that white students hold strong preferences for the school-specific attributes of a selected sample of schools. The willingness to

one might be worried about the separate identification of the two. However, preferences are identified from within-school changes in X_{jt} over time, while beliefs thanks to changes in grades. Even if X_{jt} and R_{jt} were perfectly correlated, changes in preferences due to changes in the peer quality of two schools with the same change in X_{jt} would be identical, but changes in beliefs of their quality would be different if the two schools receive different letter grades.

[53] Asymptotic standard errors of the minimum distance second-step estimates take into account the sampling error of the first stage estimates. They apply the delta-method to the first-step estimates of the variance-covariance matrix of δ_{cjt} and rely on numerical approximations when necessary.

travel for an additional cross-school standard deviation in school quality is similar across student demographics and ranges between 3 and 7 minutes. However, white and Asian students' preferences for unobserved school-specific attributes $\tilde{\xi}_{cj}$ are right-skewed and concentrated on few schools with high peer quality, value-added and enrolling more white and Asian students. These school fixed-effects explain 79% of the variation in mean school preferences of white and Asian students, and only 66-68% of that of Black and Hispanic students. This means that, relative to other school-specific attributes, school quality is less important for white and Asian students. As a consequence, changes in their expectations of school quality change their choices primarily within the few schools that are majority white and Asian and have higher peer quality.

Different races have also similar preferences for changes in peer quality over time, comparable in magnitude to preferences for school quality. Preferences for changes in white and Asian enrollment are instead smaller (between 1 and 2 minutes for an extra standard deviation, corresponding to 25 p.p. more white and Asian students). Preferences for both peer and school quality are increasing in applicant baseline achievement and this test score gradient is steeper among Black and Hispanic applicants.

Table A.1.12 assesses the model fit out of sample using the choices and offers of the 2016 cohort. Panel B also uses data from the 2014 cohort to compare real and model-based changes in the probability of ranking schools receiving an A or a low grade after the removal of grades. Overall, the model predicts well heterogeneity in choices, choice changes, and offers across applicants' demographics.[54]

I consider robustness of my estimates to alternative functional forms of student priors. Appendix Table A.1.16 shows second step estimates when priors are distributed either as a log-normal or as the empirical distribution of quality with a location and a scale shifts that vary across school discrete types. In Appendix Table A.1.13, I increase the number of discrete school types considered in the simple binary case to four, corresponding to quartiles of average Regents performance. Finally, rather than dividing schools into discrete types, in Table A.1.14 I let the prior mean and precision be a continuous linear function of the school Regents achievement levels R_{jt} as follows: $\mu_{jt} = \mu_0 + \mu_1 \cdot R_{jt}$, $\sigma_{jt}^{-1} = \sigma_0 + \sigma_1 \cdot R_{jt}$. Similarly, the model in Table A.1.15 lets prior mean and precision be a linear function of both the school average Regents achievement levels R_{jt} and the school value-added Q_j: $\mu_{jt} = \mu_0 + \mu_1 \cdot R_{jt} + \mu_2 \cdot Q_j$,

[54] School offers in panel C are simulated using the model-based rank ordered lists, real school capacities and the student priorities assigned based on admission rules announced for the 2016 admission cycle. The equilibrium simulations require some restrictions and assumptions that make the fit of offered school characteristics worse than that of choices. I discuss the details in appendix A.3.2.

$\sigma_{jt}^{-1} = \sigma_0 + \sigma_1 \cdot R_{jt} + \sigma_2 \cdot Q_j$. Because this specification allows priors to be a function of school quality, it allows for the case in which students have perfectly accurate beliefs about school quality (namely $\mu_0 = 0, \mu_1 = 0, \mu_2 = 1$) and it can be used to more directly test for bias. Regardless of the specification, white, Hispanic and higher achieving student mean beliefs are more strongly correlated with achievement levels, and belief precision is higher on average for white students. Nevertheless, as in the simple binary school type case, prior means are close to the mean quality in the city and differences across race are not very large. The other preference parameters are virtually unchanged and always show a small degree of heterogeneity in willingness to travel for quality across race on average, and a larger degree of heterogeneity across baseline achievement.

Finally, I relax the assumption that students report preferences truthfully in table A.1.17. The estimates in that table are obtained under the assumption that students do not consider schools where they have a zero probability of admission but report preferences truthfully among the remaining schools.[55] While Black and Hispanic applicants' preferences for peer quality and for the share of white and Asian students in a school are slightly larger in the strategic reporting scenario than under truthful reporting, estimates of preferences for quality are unchanged. The main heterogeneity patterns in preferences and beliefs across students and schools are also largely unchanged.[56]

1.6 Counterfactuals

In this section, I use my model to evaluate the effects of counterfactual information disclosure policies on student welfare. My definition of welfare is an average of student test scores Y_i, possibly weighted by welfare weights ω_i.[57] According to the model of student achievement in

[55]To compute probabilities of admission I bootstrap each school match 100 times redrawing each time a sample of applicants and a sequence of tie-breakers. Applicants are sampled with replacement independently. For each assignment and school, I obtain admission cutoffs from the priority and tiebreaker of the marginal student admitted to each school. The relevant tiebreaker is the largest lottery number among admitted applicants for programs that rank applicants based on lottery number, or as the lowest score among admitted applicants for programs that rank applicants based on prior academic performance. The admission probabilities are estimated based on these bootstrapped cutoffs, which capture the uncertainty in admission due to variation in the lottery draw and year-to-year variation in the applicant population.

[56]More sophisticated approaches to school demand estimation under strategic reporting rely on stability as in Fack et al. (2019) or view applicants' rank ordered lists as the outcome of an optimal portfolio problem as in Agarwal and Somaini (2018); Larroucau and Rios (2020); Idoux (2021); Calsamiglia et al. (2020). These methods could also be applied to estimate my model in the future.

[57]In the education literature, evaluating interventions and changes in market designs using a notion of welfare that depends directly on student outcomes is often the standard (Kapor, 2020; Barahona et al., 2023a), although it is also possible to prefer revealed-preference measures of student utility (Abdulkadiroğlu et al., 2017a; Kapor et al., 2020).

equation (1.1), welfare depends on the allocation of students to schools, μ, through school value added:

$$W(\mu) = \sum_i \omega_i Y_i(\mu) = \sum_i \omega_i (\alpha_{\mu(i)} + X_i'\Gamma + \epsilon_i)$$

where $\alpha_{\mu(i)}$ is the value added of the school that student i gets under allocation μ.

I evaluate welfare gains or losses on average and for student subgroups with respect to the simulated status-quo scenario in which the policy maker provides no information about school effectiveness, which I have shown replicates well realized choices and offers in 2016. Denoting the status-quo allocation with μ^0, the total change in welfare associated with an information policy that induces the allocation μ is simply given by the average change in the value-added of enrollment schools:

$$\Delta W(\mu, \mu^0) = \sum_i \omega_i (\alpha_{\mu(i)} - \alpha_{\mu^0(i)})$$

.

Achievement gains in this model are realized when students reallocate to vacant school seats that are of higher quality than their current allocation. A slack in the capacity constraint of high-quality schools is necessary to obtain improvements in average student welfare. To benchmark welfare gains in my simulations against what is feasible given school capacities in 2016, I quantify the maximum attainable gains as the difference between the average student achievement under the allocation that matches students to the best available school and average achievement in the status-quo allocation. I call this difference “first-best" achievement gains. They would be realized in the student-proposing DA allocation if students only valued school quality and ranked schools in order of value-added.[58] Reallocating students to vacant high quality school seats can increase average test scores at most by 0.039σ. In what follows, I often express welfare gains under different allocations as a percentage of this number.

1.6.1 Effects of Providing Perfect Information

Full Information Benchmark This section studies the effects of providing perfect information about the value added of each school on choices and offers. Panel A of figure 1-8 compares the average quality of top three choices under full information and in the simulated status quo. On average, chosen quality increases from the 68th to the 74th value-added school percentile ranking. These simulated changes are larger than those observed in the Bloomberg era, when letter grades were only an imperfect proxy for value-added. Students, however, do not “max

[58] In the first best simulation, students rank schools in order of value-added and schools rank students using the admission rules, priorities, and tie-breakers used in the 2016 admission cycle.

out" on value-added even under perfect information because of preferences for other school attributes. In the absence of capacity constraints, these changes in choices would result in average achievement gains of 0.07σ. Thanks to the larger response of Black and Hispanic student choices, information entirely closes the racial choice gap conditional on test scores and could reduce the mean cross-race achievement gap by 62% or 0.035σ.

Panel B of figure 1-8 compares changes in chosen quality to changes in offered quality for different groups of students, defined on the basis of the student race (minority or non-minority) and baseline achievement (above or below median). Due to binding capacity constraints in high-quality schools, the average offered VA improves only by 0.01σ.[59] Therefore both capacity constraints and preferences for school attributes other than school quality reduce the extent to which information provision can improve allocative efficiency and boost test scores.[60] Nevertheless, perfect information yields 23% of the first best average achievement gains, which are depicted in the last group of bars on the right in figure 1-8. Information also disproportionately improves Black and Hispanic students offers, not just their choices.

Appendix Table A.1.18 summarizes changes in other characteristics of school choices and offers. Because higher quality schools tend to enroll higher achieving peers, the average peer quality and share of white students in applicants' top choices are also higher but differences are small, suggesting that students respond to information about school quality only within a subset of schools with a demographic composition that is more similar to theirs.

Why Do Black and Hispanic Students Benefit More From Information? Model estimates indicate that information can disproportionately affect Black and Hispanic students both because this group is relatively less well informed than non-minorities and because their preferences for school quality are stronger relative to those for other school traits. To quantify the importance of each channel, I compare the simulated effects of providing perfect information about school VA on school choices and offers in three hypotetical scenarios. In the first, called "Uninformed Priors" (UP), I remove differences in prior information about school quality. I assume everyone holds the same uninformed prior for all schools, equal to the empirical distribution of VA in NYC. In the second, "No preferences for Peers" (NP), I assume applicants have no preferences over school specific traits and the composition of students at a school. The third (UP+NP) combines the first two.

[59] 72% of the best 20% of schools are oversubscribed in the status quo, while only 21% of the worst 20% are.

[60] The fact that seats would be filled regardless of school efforts to improve may undermine their competitive incentives to increase quality even under full information. I leave to future work the analysis of these potential supply-side implications.

Panel (a) of figure 1-9 shows the effect of providing information on choices in these scenarios. The first bar in each subgroup benchmarks the information effects against those estimated under the actual preferences and beliefs. Information effects in the uninformed-prior simulation are similar to the actual full-information benchmark across student demographics. In contrast, White and Asian students would choose much higher quality when informed about it in the no-peer-preferences scenario than in the benchmark. Information effects on minority choices remain similar across all counterfactuals. These results indicate that student priors are overall quite misinformed and that differences in preferences are more important than differences in beliefs in explaining the larger response to information of minority students.

Panel (b) reveals that the larger response of white and Asian students' choices changes the distribution of welfare gains in the simulations that remove preferences for school traits different from quality and distance. In these hypothetical scenarios, their stronger reaction to information would displace low achieving Black and Hispanic students out of high-quality seats, even in the absence of changes in minority choices. This exercise reveals that Black and Hispanic students disproportionately benefit from information thanks to white and Asian students' strong preferences for school traits different from quality.

Finally, I show that differences in distance to high quality schools across race play no role in explaining the choice gap and the effects of information on choices and offers. Appendix figure A.1.9 compares the benchmark effects of providing information (FI) on choices and offers to the effects of removing distaste for commuting (ND) and the effects of providing information in this hypothetical scenario (FI if ND). Removing any role for differences in distance to schools has essentially no effect on student choices and does not change the effects of providing information on chosen and offered school quality. This also shows that the smaller information effect on offers compared to choices is not explained by the spatial distribution of schools. That it, the discrepancy is not explained by a lack of good schools nearby where students live that creates congestion only in a small number of high quality schools.

Information About Value-Added vs. Achievement Levels Information interventions in education often inform families of the average achievement of students enrolled at a school rather than of causal school value-added (Hastings and Weinstein, 2008; Cohodes et al., 2022; Corcoran et al., 2022; Allende et al., 2019; Andrabi et al., 2017). On the one hand, if achievement levels and value-added are positively correlated, naive policies providing information about the former can still induce students to reallocate to better schools. On the other, this information might be mostly redundant and fail to shift households' choices towards higher quality options because families' perceptions of quality are already based on school achievement levels.

Figure A.1.10 compares the welfare gains obtained by information about school value-added relative to information about school achievement levels. In these simulations, information interventions either perfectly disclose differences across schools (denoted by “FI" or Full-Information) or take the form of coarse ratings corresponding to quintiles of value-added or achievement levels (denoted by “5L" or 5 Letters).[61] Information interventions based on achievement levels obtain half of the gains of those based on causal value-added for Black and Hispanic students' test scores, and produce no gains for white and Asian students. White and Asian quality beliefs are more strongly correlated with achievement levels and their preferences concentrated on the subset of schools with high performing students, therefore information about school performance levels does not change their choices much. By virtue of the positive correlation of VA and achievement levels, however, information about the latter can still partly re-direct Black and Hispanic students to choose better schools.

Comparing Information Provision to Changes in Admission Rules The previous results suggest that information disproportionately improves Black and Hispanic student achievement. An alternative set of policies currently considered to help Black and Hispanic students match to high quality schools are changes in school admission rules ranking students based on their residential address or their middle school test scores.[62] These admission rules may also substantially limit the beneficial effects of information interventions, which motivates not only comparing the effects of information provision to those of admission reforms but also studying their combined effects.

Figure 1-10 benchmarks the welfare changes of providing full-information against the effects of removing all geographic priorities and academic screens in admissions, denoted by “NS" in the figure. The effects of combining changes in admission rules with perfect information about value added are denoted by “FI+NS".[63] This exercise offers three main insights. First, changing admission rules redistributes school quality from high to low achieving students within student race, while providing information primarily redistributes across race and benefits Black and Hispanic students across all achievement levels. Information displaces low

[61] The first bar for each student group shows gains under the full-information benchmark. The second bar corresponds to gains realized under the disclosure intervention that assigns schools letter grades based on VA quintiles. The third and four counterfactuals simulated the effects of providing information about achievement levels in these two forms while presenting it as if it were about school VA. That is, if the difference in achievement levels of two schools is one school-level standard deviation, households are told that the difference in quality between the two schools is one school-level standard deviation.

[62] These rules are often thought to disproportionately favor white and Asian students and are often at the center of debates about the equity of the school match (Cohen, 2021; Idoux, 2021; Park and Hahm, 2023).

[63] Removing geographic priorites and academic screens ensures that any two students with the same rank ordered list have the same admission chances to any school before uncertainty in their lottery number is resolved.

achieving white and Asian students out of high-quality schools under the current admission rules and high achieving white and Asian students when removing screens. Second, combining information and leveling the playing field in admission rules are not substitute policies. Their effects are cumulative and, if anything, they seem to act as complements in raising Black and Hispanic student test scores. Intuitively, not only minority students would know where to find higher quality schools, but they would also have fair chances to get in. Combining the two does not help white students reallocate to better schools when they lose their priority advantage, but it further displaces them out of high-value added schools. Third, providing information yields achievement gains among Black and Hispanic students comparable to those obtained by reforming admission rules, which is a more politically controversial policy.[64] Average Black and Hispanic student test scores improve by 80% of what would be possible under status-quo information if school admission rules treated all students equally. Even just focusing on minority students with achievement levels below the median, information provision allows to achieve a third of the de-screening gains. Moreover, average achievement is higher under full information as white students also benefit from information on average. On the contrary, their average test scores are 0.02σ lower in the NS counterfactual compared to the status quo.

Targeted Outreach Redistribution of high-quality school seats in favor of lower achieving students might be in line with the policy maker's objective, for instance if there are critical levels of achievement that need to be reached (e.g. failing to graduate is more costly than failing to graduate with the highest honors). Due to capacity constraints, one strategy to achieve this objective could involve selectively providing information to a targeted group of students to mitigate congestion effects that can arise if information is disseminated to everyone.

To mimic the effects of a realistic policy, I simulate the effects of an outreach intervention that provides information to all students at middle schools with average test score levels below the median, rather than differentiating students within middle schools. This targets 30% of students, coming from a disadvantaged population: targeted students' 7th grade Math test scores are lower by 0.9σ on average, 84% are eligible for free and reduced price lunch (compared to 64% among the non-targeted) and 90% are Black or Hispanic (compared to 51% among the non-targeted). Table A.1.19 shows that targeted students choose schools with 0.074σ higher value added and their offers would improve by 0.033σ, 85% of the average first-best gains, substantially more than if everyone were informed. The cost of preventing targeted outreach on

[64]The assumption I maintain here is that application behavior does not change in response to changes in admission rules. Past research finds that this feedback effect might amplify the effects of these policies, suggesting that the changes I simulate in the no-screen counterfactual provide a lower bound for the achievement gains of low achieving students (Idoux, 2021).

the outcomes of the most disadvantaged students can be quantified as the welfare change with respect to the full-information benchmark, corresponding to achievement losses for targeted students of 0.02σ on average.

Panel B of table A.1.19, instead, presents the effects of supplying quality information exclusively to Black and Hispanic students whose test scores fall within the top tercile of the city's distribution. Such outreach initiatives are currently under consideration as policy alternatives to affirmative action, with the aim of enhancing the representation of racial minorities in high-quality schools. Due to the large preferences for quality of this subgroup of students, and because few students receive information, achievement gains for targeted students are three times as large as when everyone is fully informed. However, the share of Black and Hispanic students in the best 20% of schools increases only by 2.3 p.p. Providing full information increases this share by 2.6 p.p. and removing geographic priorities and screening by 9 p.p., suggesting that targeted outreach is less effective than other policies to increase representation of non-white students in top quality schools.

1.6.2 Optimal Information Design

Providing information only to a selected group of students is one obvious way to redistribute high quality school seats in the presence of capacity constraints, but it may be politically unfeasible. In this subsection, I consider how the policy maker could instead design school quality ratings visible by all applicants to improve achievement among minority or low achieving students. The effects of coarse quality ratings may also offer a more realistic reference point for the effects of information disclosure, as this is often the prevailing format in real-world settings, such as healthcare, education, nutrition, and finance. My model does not take into account the cognitive cost of information processing, often cited as the primary reason coarser information may be more effective than finer, more detailed information. However, the model can justify why coarsening information may be preferred to full information disclosure through two main channels.

First, when the social planner objectives are different from the maximization of participants' utility, as is the case if it only cared about education achievement, coarsening information may persuade agents to take actions that would not be optimal from the agent's point of view but may be from the point of view of the sender of the signal (Kamenica and Gentzkow, 2011). In this setting, coarsening information could induce students to choose schools of higher quality compared to the full-information benchmark, and translate into higher average achievement. Second, coarser information may allow to redistribute quality to less advantaged students more than precise information in equilibrium.

To build some intuition, consider the following two stylized examples.

Example 1 - Coarser information is Pareto improving from the planner's perspective. There are 4 schools a, b, c, d each with 1 seat and 2 students i_1, i_2. Student i_1 always has higher priority than student i_2 at all schools, for instance because she has higher test scores. Students care about school quality q and other school attributes p. School characteristics and student utilities are as follows:

	a	b	c	d
q	3.5	1.5	1	0.5
p	3	2	4	3.5

$$u_{1j} = E[q_j] + p_j$$
$$u_{2j} = E[q_j] + \frac{1}{2}p_j$$

Students hold uninformed priors that correspond to the distribution of quality in the city. Denote with F the state in which students know the quality of each school, with N the state in which they receive no information, and with C the state in which they receive information about which of the four schools are the best two and which are the worst two. Student preferences in these three states are:

$$c \succ_1^N d \succ_1^N a \succ_1^N b \qquad c \succ_2^N d \succ_2^N a \succ_2^N b$$
$$a \succ_1^F c \succ_1^F d \succ_1^F b \qquad a \succ_2^F c \succ_2^F b \succ_2^F d$$
$$a \succ_1^C b \succ_1^C c \succ_1^C d \qquad a \succ_2^C b \succ_2^C c \succ_2^C d$$

And the resulting allocations from student-proposing DA are:

$$\mu^N = \begin{pmatrix} i_1 & i_2 \\ c & d \end{pmatrix} \quad \mu^F = \begin{pmatrix} i_1 & i_2 \\ a & c \end{pmatrix} \quad \mu^C = \begin{pmatrix} i_1 & i_2 \\ a & b \end{pmatrix}$$

Under coarse information, i_2 receives strictly higher value-added than under full information, and i_1 is no worse-off. Intuitively, because students care about quality, pooling the best two schools convinces them to rank b higher, which has higher q but not high enough to be preferred to c under full information.

Example 2 - Coarser information is redistributive. Now instead, let student i_1 care even more about school attributes p, so that her utility becomes $u_1 = E[q]+2p$. Preferences and allocations now are:

$$c \succ_1^N d \succ_1^N a \succ_1^N b, \quad c \succ_2^N d \succ_2^N a \succ_2^N b$$
$$a \succ_1^F c \succ_1^F d \succ_1^F b, \quad a \succ_2^F c \succ_2^F d \succ_2^F b$$
$$c \succ_1^C a \succ_1^C d \succ_1^C b, \quad a \succ_2^C c \succ_2^C b \succ_2^C d$$

$$\mu^N = \begin{pmatrix} i_1 & i_2 \\ c & d \end{pmatrix} \quad \mu^F = \begin{pmatrix} i_1 & i_2 \\ a & c \end{pmatrix} \quad \mu^C = \begin{pmatrix} i_1 & i_2 \\ c & a \end{pmatrix}$$

The average test scores under both the full and the coarse information scenario are the same, and are higher than under no information. However, who gets the highest value-added school, a, differs. Because student i_1 cares a lot about other school attributes, coarse information does not provide a strong enough signal to induce her to choose the highest quality school. Coarsening information therefore removes the competition for school a, which student i_2 prefers, but that gives priority to i_1.

Similar patterns could be observed in reality due to heterogeneity in preferences for school attributes across students. Intuitively, information about the quality of schools that are considered non-desirable for other reasons is not much valuable. Therefore, precise information about schools with lower peer quality or enrolling higher shares of minority students will not induce large responses among white and higher achieving students. Conversely, precise information about schools that white and higher achieving students like increases their sorting to high-quality seats, displacing disadvantaged students. Moreover, the precision of priors of white and high achieving students is higher, which may induce them to respond less than other students when quality signals are coarse but not when they are precise. The policy maker, therefore, may face a trade-off between providing information and convincing students to rank less preferred but higher quality schools (as in example 1) or discouraging some particular students from applying to higher quality schools (as in example 2).

I first consider where to place the quality cutoffs of 5 letters grades. Table A.1.20 reports the best and worst 5 letter policy for the welfare of different groups of students and the associated average welfare change, as a share of first best mean achievement gains. The best 5 letter policy in terms of average welfare can achieve 20% of the first-best gains, and a naive policy that places cutoffs evenly distributed along the value-added distribution around 17%.

The different spacing of letter cutoffs, however, determines who gains the most from information. The policy that helps lower achieving students the most, even more than full-information, is one that provides very precise signals at the bottom of the quality distribution. On the contrary, the worst policy for this subgroup is one providing precise signals at the top. The opposite is true for high achieving students, who would benefit from more precise signals at the top than at the bottom. To build some intuition for why this is the case, I consider the simple case of deciding how to design 2 quality ratings, high and low. Figure A.1.11 plots quality in choices and offers as a function of the high rating cutoff. As the cutoff increases, all students choose higher quality schools. Chosen quality, however, increases faster among higher achieving students as the signal at the top gets more precise, resulting in tougher competition for high quality seats. Intuitively, because high achieving students hold strong preferences for the attributes of schools that on average have higher quality, increasing the cutoff provides them more precise information about the set of schools they like. They find this information more valuable than precise information about low quality schools and react to it more strongly. As a result, offered value added of low achieving students is maximized in this simple two-rating case when the cutoff is placed at the 30th percentile. As the cutoff increases they are displaced out of higher quality schools at higher rates.

These results indicate that policies providing students with a list of top performing schools need not help equally everyone (Cohodes et al., 2022). As in example 2, such policies might increase competition for high quality schools from high achieving students who react more when seeing precise signals of quality at the top. Disadvantaged students might be screened out of top performing schools based on their achievement, and might be at risk of ending up in worst schools if they cannot distinguish the bad schools from the average ones.

Next, I vary the precision of the quality signal along the slackness of the capacity constraint. In this model aggregate gains are possible through the reallocation of students from low quality schools to higher quality schools with empty seats. The ultimate goal for the policy-maker is to convince students to rank higher in their list good schools that are not yet full. I simulate the effects of providing students with a less precise quality signal (above or below median) about schools oversubscribed in the status-quo and an infinitely precise signal for schools that are undersubscribed. This mimicks an advertising campaign that provides exact information only about undersubscribed schools. I call this counterfactual “pooling", because the quality of oversubscribed schools is pooled, while that of undersubscribed schools is not.

Figure 1-11 compares welfare gains from the pooled policy with those of the full-information benchmark and of the best and worst five letter grades policies for low achieving students. As in the previous counterfactual, low achieving students strictly benefit in equilibrium from the policy with less information compared to the full information scenario. High achieving students still fare better than under the status quo, but less than under full information. The intuition is again the same as in Example 2. The coarse quality signals for oversubscribed schools are not informative enough to induce a strong sorting of high achieving students to the best oversubscribed schools, limiting displacement of low achieving students out of high quality over-subscribed schools. At the same time, high achieving students are relatively less responsive to precise signals about quality of under-subscribed schools because they like the characteristics of these schools relatively less than low achieving students. The results of both counterfactual simulations presented in this sub-section thus suggest that the policy maker can partly leverage information design to redistribute value added across student groups, less so for pushing the Pareto frontier of test score achievement.

1.7 Conclusions

School choice can achieve equity, allocative, and efficiency gains provided that families reward school effectiveness. This assumption, however, has proven to fail in many settings. This paper shows that a lack of accessible information about school quality is partly to blame and explains a portion of the disparities in access to high-quality education across races. To do this, the paper leverages a natural policy experiment that varied the information about school quality available to students in NYC. Black and Hispanic students are more responsive to school ratings, allowing information to reduce achievement inequality.

Based on a structural model of demand for schools, choice responses to information reveal differences in both beliefs and preferences for quality across students of different races and with different baseline achievement. Everyone is misinformed about which schools are of higher quality, and minority students more than non-minority students. Even if misinformed, all applicants care about value-added, separately from the composition of the students at a school. White and Asian families, however, value other school attributes relatively more than minority families. Their strong preferences for schools enrolling white and high-achieving students limit their responsiveness to quality information. As a consequence, information interventions and their design can partly redistribute school quality even under a fixed supply of school seats.

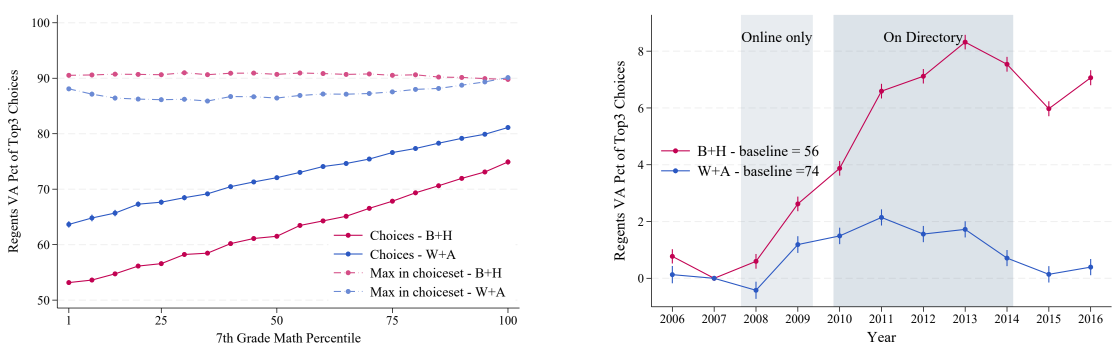
The findings of this paper become even more relevant in light of recent developments in school accountability policy. Following the passage of the No-Child Left Behind Act, school accountability received significant attention. Accountability reforms aimed not only to incentivize schools to improve student achievement but also to empower families with informed decision-making tools. In many school districts, these objectives were achieved through the provision of easily understandable school performance ratings, such as letter grades. However, in recent years, school performance measurement and accountability appear less prominent on the education policy agenda. Many school districts have shifted away from summative assessments based on student achievement. In the specific context of my study, letter grades have been replaced with multi-dimensional school quality measures that may be less visible and harder to parse.

While school letter ratings in NYC were far from perfect and, to some extent, reflected student selection rather than true causal value-added, they had a significant impact on the choices of less advantaged families. This underscores the importance of providing accessible information about school quality to all. Designing ratings that more accurately represent causal estimates of school effectiveness and can be tweaked to help the most disadvantaged could be a more effective policy approach to reducing achievement inequalities. This approach may prove superior to both the earlier simplistic ratings and current policies that are veering away from ratings altogether.

1.8 Figures

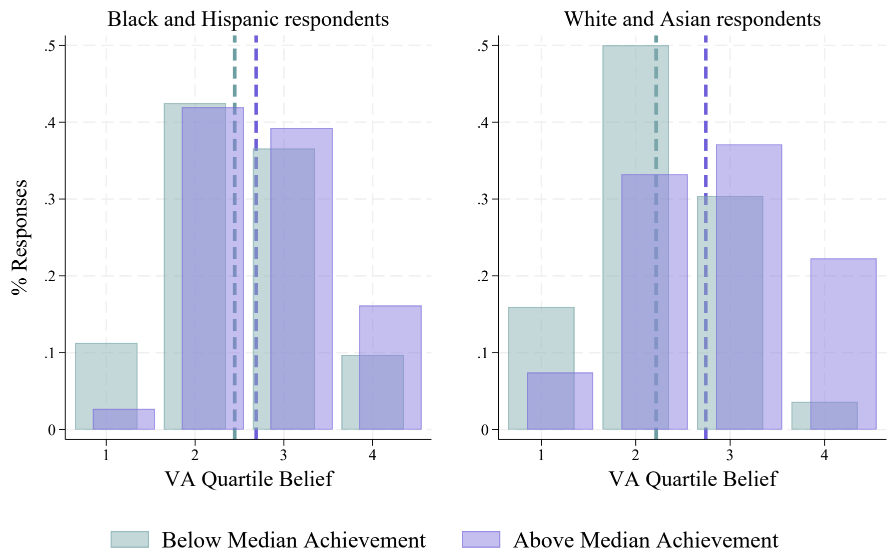
Figure 1-1: The Racial School Quality Choice Gap

(a) By Baseline Achievement (b) Across Cohorts



Notes: This figure describes cross-race differences in chosen school quality, as measured by Regents value-added. Panel (a) plots the relationship between the percentile rank of the student's baseline score and two variables: the average value-added of students' first three choices (solid lines) and the average value-added of the best three school options in the student's feasible set (dashed lines). Blue lines are averages for white and Asian students, pink lines are averages for Black and Hispanic students. Each line is a raw average computed within student cells defined by combinations of race and 20 baseline test score bins. Panel (b) plots the difference in average school value-added in the first three choices by race and cohort with respect to 2007. Race differences in choices are normalized to zero in 2007.

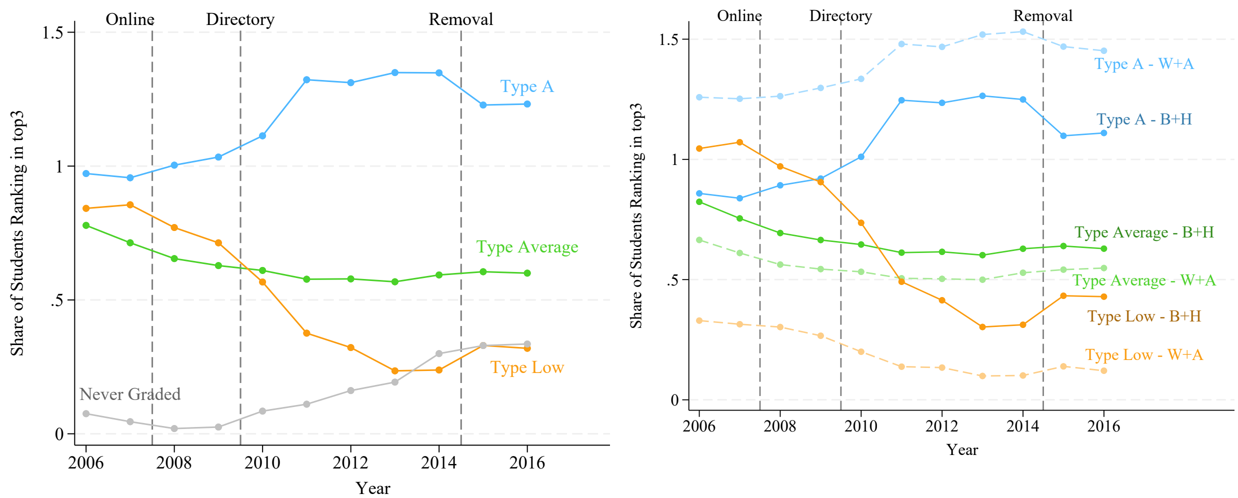
Figure 1-2: Distribution of Beliefs About School Quality by Race and School Achievement



Notes: This figure shows the distribution of responses to a survey question eliciting beliefs about school quality, separately by respondents' race and the school average achievement level. The question asked: "How well does *school name - (school code)* prepare students for their Regents exams compared to other schools in your borough?". Possible responses ranged from 1 (corresponding to the bottom 25% of school quality) to 4 (best 25% of schools). Violet bars are responses to questions asking beliefs about schools with above median average Regents levels and green bars are for schools with below median Regents levels. The panel on the left shows the distribution of answers for Black and Hispanic respondents, the one on the right that of white and Asian respondents.

Figure 1-3: Trends in School Shares by School Letter Category

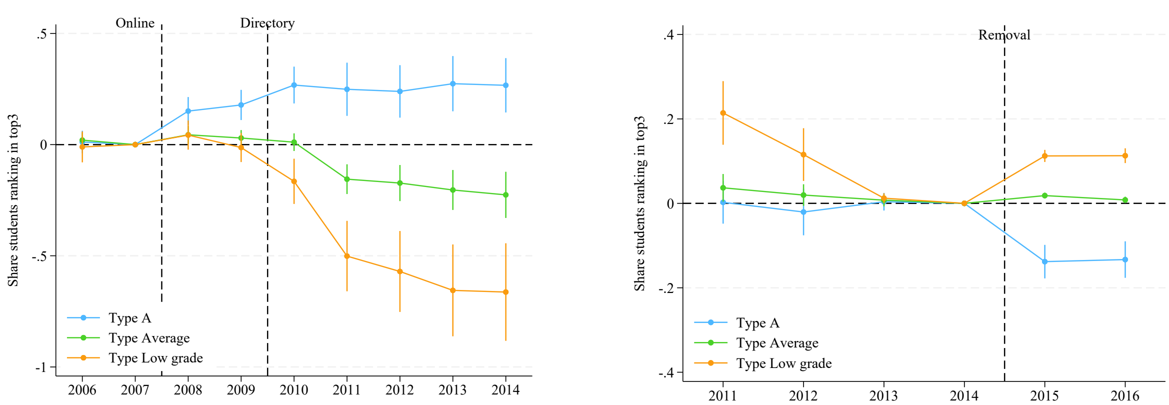
(a) All Students (b) Separately by Race



Notes: The figure shows trends in demand for schools. Demand is measured as a school share among student first three choices. The graph shows the average school share by school grade type and year. Type A schools receive an A in 5 out of the 7 years, Type Low schools receive a grade of C,D, or F in 5 out of the 7 years, Never graded schools never received a grade and Type Average schools are all remaining schools. Vertical lines indicate, in order, the introduction of letter grades online, their introduction on the directory, and their removal. Panel A pools the choices of students of all races, while panel B separately shows school shares by student race.

Figure 1-4: Event Study Estimates of Demand Responses to Introduction and Removal of Quality Signals

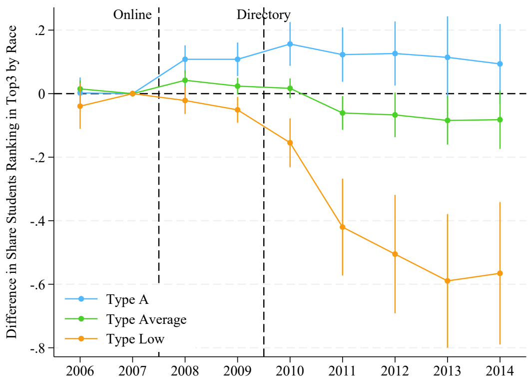
(a) Introduction of Letters (b) Removal of Letters



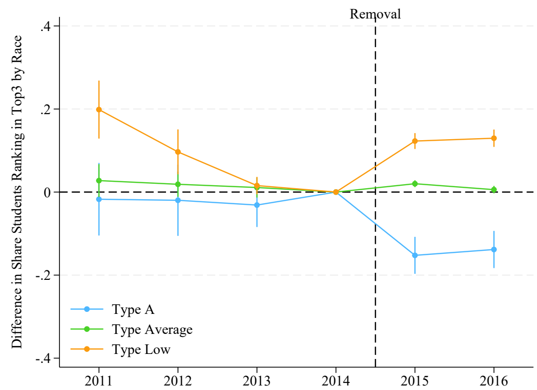
Notes: The figure plots event study estimates of the coefficient β_L^t of equation (1.3). Panel (a) considers changes relative to 2007, the year before the introduction of letters, using applicant cohorts of 2006-2014. Panel (b) considers changes around the removal of letters, normalizing shares differences to 0 in 2014, and using cohorts of 2011-2016. The blue line is for changes in shares of Type A schools, the orange for shares of Type Low schools and the green line is for Type Average schools.

Figure 1-5: Event Study Estimates - Heterogeneity by Student Race

(a) Introduction of Letters

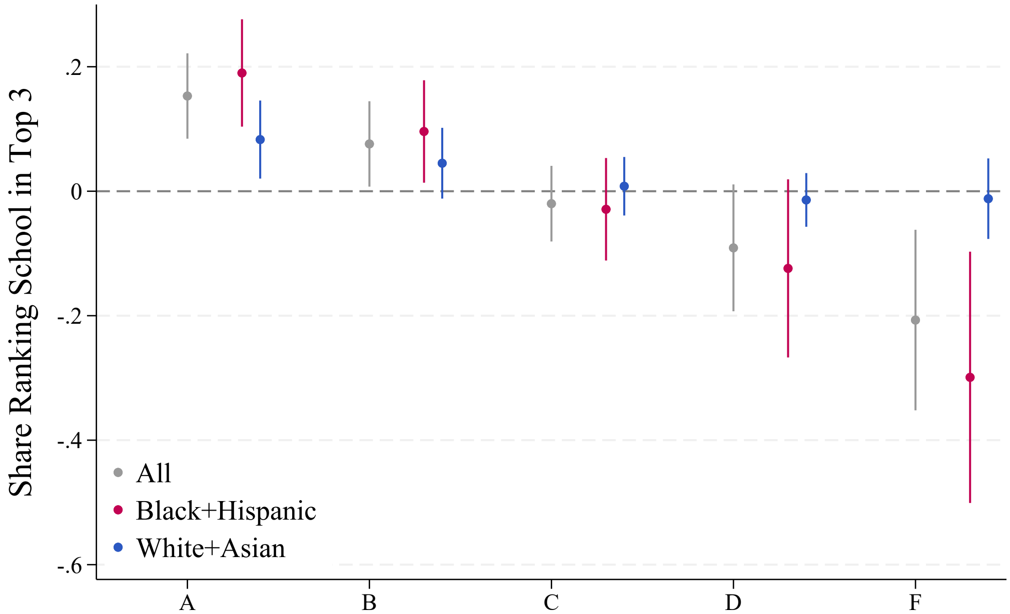


(b) Removal of Letters



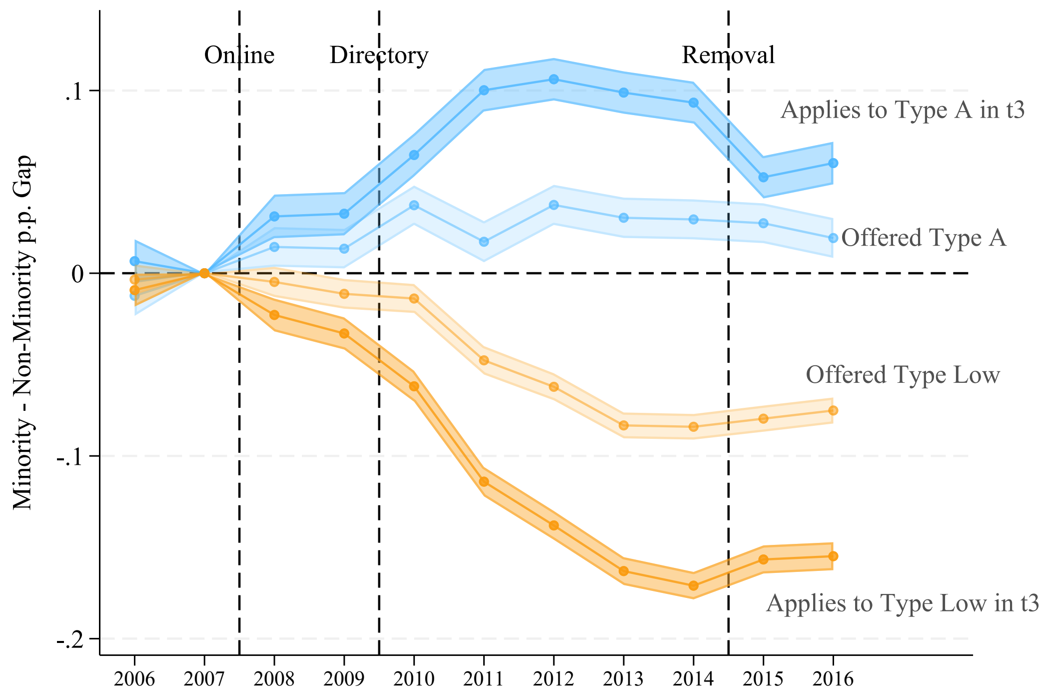
Notes: The figure plots event study estimates of the coefficient δ_L^t of equation (1.5), capturing cross-race differences in choice responses to the introduction and removal of letter grades. Panel (a) considers differential changes by race relative to 2007, the year before the introduction of letters, using applicant cohorts of 2006-2014. Panel (b) considers changes around the removal of letters, normalizing shares differences to 0 in 2014, and using cohorts of 2011-2016. The blue line is for changes in shares for Type A schools, the orange for shares of Type Low schools and the green line is for Type Average schools.

Figure 1-6: Year-to-Year Demand Responses to Letter Grades - Heterogeneity by Race



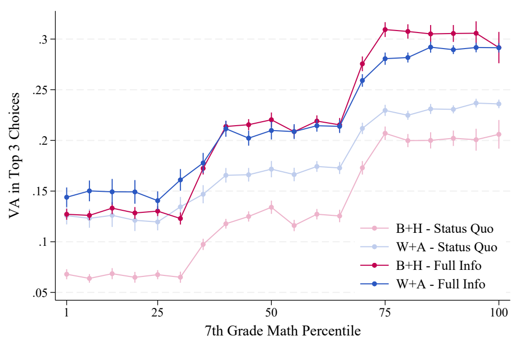
Notes: The figure plots letter grade effects on demand for schools, measured by estimates of the coefficients β_g in equation (1.4). The dependent variable is the share of students in demographic cell c and application cohort t ranking the school among their first three choices. Demographic cells are defined by the interaction of a student residential borough and baseline test score tercile (gray estimates) supplemented with student race (pink and blue estimates). Controls are those used in table 1.5 and always include school-cell fixed effects and year-cell fixed effects. The sample includes applicant cohorts from 2010 to 2014 included. Standard errors are clustered at the school-year level.

Figure 1-7: Event Study Estimates of Changes in Choice and Offers Probabilites by Race

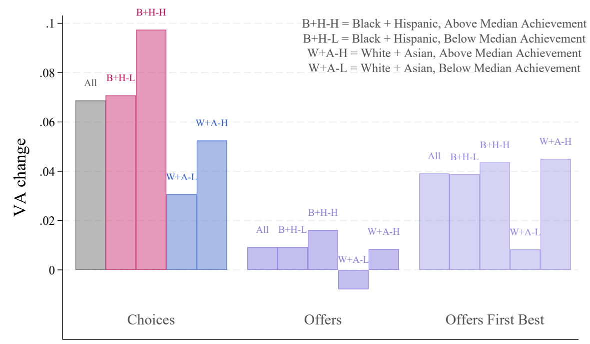


Notes: The figure plots event study estimates of the coefficient δ^t of equation (1.8) for regressions using four different dependent variables, indicating ranking a Type A school among a student first three choices, receiving an offer to a Type A school, and similar events for Type Low schools. The coefficient δ^{2007} is normalized to zero. The sample includes students applying to enroll in 9th grade between 2006 and 2016. Controls include gender, ell status, subsidized lunch status and fixed effects for combinations of student borough and baseline test score terciles.

Figure 1-8: Change in VA in Top 3 Choices and Offers Under Full Information



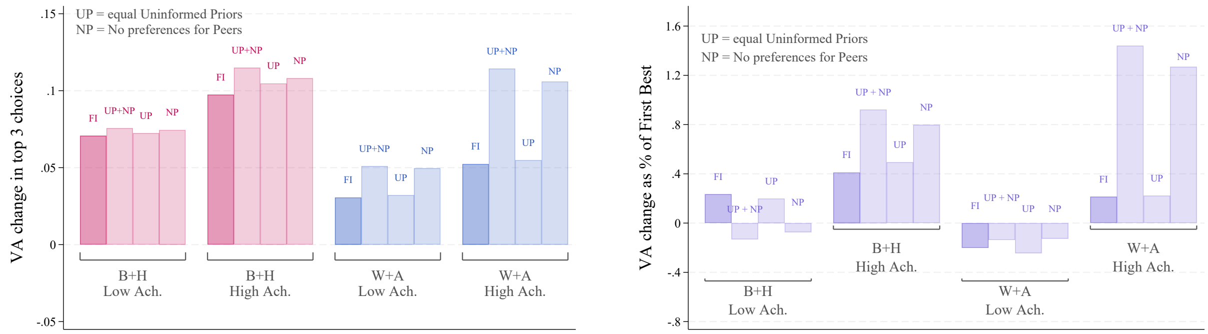
(a) ΔVA in Choices by Race and Bl. Achiev.



(b) Average ΔVA in Choices and Offers

Notes: Panel (a) plots the relationship between the percentile rank of the student's baseline score and the average value-added in the student's simulated first three school choices. The darker lines correspond to the full-information counterfactual, while the lighter ones are for choices in the (simulated) status-quo. Blue lines are averages for white and Asian students, pink lines are averages for Black and Hispanic students. Each line is made of raw averages computed within student cells defined by combinations of race and 20 baseline test scores bins. Panel (b) compares changes in VA of choices to changes in offered VA. The first group of bars plots the average change in VA of students' top 3 choices between the full-information benchmark and the status quo by student subgroups defined by combinations of race and baseline achievement (above or below median). The second group of bars plots the corresponding changes in offered VA. The last group of bars plots the average change in offered VA between the first best and the status quo.

Figure 1-9: Role of Beliefs and Preferences in Explaining Effects of Information

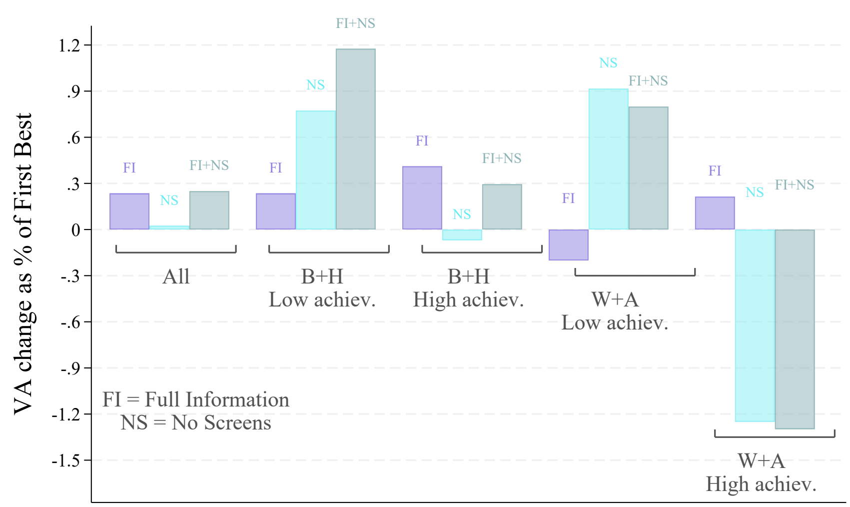


(a) Average ΔVA in Choices

(b) Welfare Gains

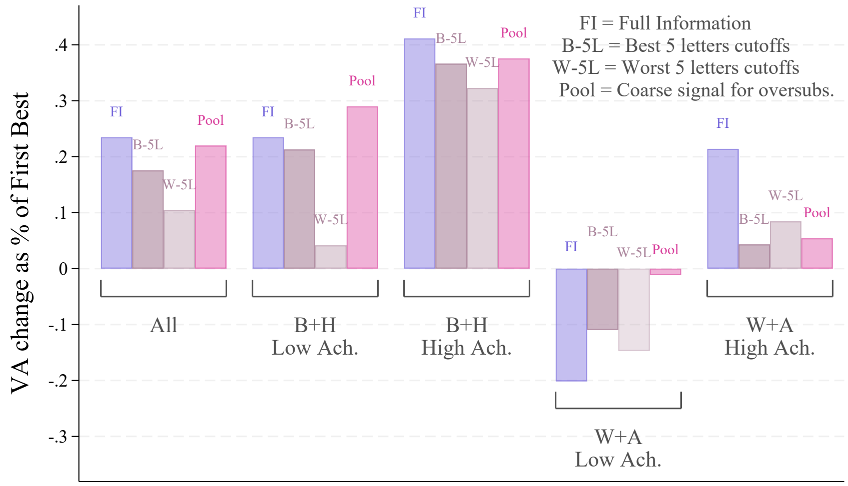
Notes: Panel (a) shows changes in VA of top 3 choices across four different simulations, taking averages within student groups defined by race and baseline achievement (above or below median). Panel (b) does the same thing for the resulting change in offered VA, expressed as a percentage of the average first-best achievement gains. Within each students subgroup, the first bar corresponds to the full-information benchmark that uses the actual model estimates. The second bar corresponds to a simulation that changes both priors and preferences: students' priors are equally uninformed and students judge schools only on the basis of their quality and their distaste for commuting. The third bar simulates choice changes only assuming students' priors are equally uninformed while the fourth bar assumes students only care about quality and distance but may hold different priors.

Figure 1-10: Welfare Changes Under Full Information and No Screening in Admissions



Notes: This figure plots the average welfare change from the status quo by student subgroups for three different counterfactual simulations of student assignment. Welfare gains are VA changes, expressed as a percentage of the average first-best achievement gains. Student subgroups are defined by combinations of race and baseline achievement. FI denotes student assignment under full-information; NS student assignment if admission rules had no academic screens or geographic priorities under the status-quo information environment; FI+NS combines full information with the removal of academic screens and geographic priorities.

Figure 1-11: Welfare Changes Under Coarser Information



Notes: This figure plots the average welfare change from the status quo by student subgroups for four different counterfactual simulations of student assignment. Welfare gains are expressed as a percentage of the average first-best achievement gains. Student subgroups are defined by combinations of baseline achievement and race. FI denotes student assignment under full-information, “B-5L" assignment under the best 5 letter grade rule for students with below median baseline achievement, “W-5L" assignment under the worst 5 letter grade rule for students with below median baseline achievement, while “pooled" denotes a counterfactual in which students receive a coarse signal about the quality of schools oversubscribed in the status quo and an infinitely precise signal about the quality of under-subscribed schools.

1.9 Tables

Table 1.1: Applicants Descriptive Statistics

	All	Minority	Non-minority	Black	Hispanic	White	Asian
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: student demographics							
N	625,868	425,579	200,289	185,658	239,921	91,272	104,405
Black	0.30	0.44	0.00	1.00	0.00	0.00	0.00
Hispanic	0.38	0.56	0.00	0.00	1.00	0.00	0.00
White	0.15	0.00	0.46	0.00	0.00	1.00	0.00
Asian	0.17	0.00	0.52	0.00	0.00	0.00	1.00
Subsidized lunch	0.78	0.84	0.64	0.81	0.87	0.48	0.78
Ell	0.09	0.09	0.08	0.02	0.15	0.04	0.12
7th grade Math	0.18	-0.07	0.72	-0.10	-0.05	0.64	0.82
7th grade English	0.13	-0.05	0.52	0.01	-0.10	0.64	0.43
Bronx	0.22	0.30	0.06	0.23	0.35	0.06	0.06
Brooklyn	0.32	0.32	0.30	0.46	0.22	0.32	0.29
Manhattan	0.11	0.12	0.09	0.08	0.14	0.11	0.07
Queens	0.29	0.24	0.42	0.20	0.26	0.28	0.54
Staten Island	0.06	0.03	0.13	0.03	0.03	0.24	0.04
Panel B: characteristics of top3 high school choices							
Commuting time (minutes)	40	41	39	45	38	38	39
Regents math VA (percentile)	65	61	75	60	62	73	77
SAT math VA (percentile)	70	65	83	64	65	82	85
Peer quality (percentile)	74	68	86	69	68	85	87
White+Asian %	0.35	0.25	0.56	0.24	0.26	0.58	0.54
Graduation rate	0.78	0.75	0.82	0.76	0.75	0.82	0.83
Panel C: student outcomes							
Regents Math σ	0.04	-0.13	0.54	-0.18	-0.10	0.47	0.62
Regents Ela σ	0.33	0.16	0.69	0.14	0.17	0.71	0.67
SAT Math σ	0.13	-0.25	0.72	-0.30	-0.21	0.54	0.86
SAT Ela σ	0.16	-0.14	0.63	-0.15	-0.13	0.69	0.60
Graduates in 4 years	0.78	0.73	0.90	0.73	0.73	0.90	0.91
Enrolls in college	0.67	0.58	0.82	0.58	0.59	0.79	0.85

Notes: This table provides descriptive statistics for the sample of 9th grade applicants applying to enroll in high school between 2006 and 2016. Panel A describes applicants' demographic composition, baseline test scores and residential boroughs. Panel B summarizes the characteristics of their first three high school choices. Panel C restricts the applicant samples to students who enrolled in the district and for whom I observe achievement outcomes. Column (1) reports averages across all students, while columns (2)-(7) consider student subgroups by race or ethnicity. Minority refers to Black and Hispanic students, while non-minority to white and Asian students.

Table 1.2: School - Year Descriptives by Letter Grade

	All schools		A	B	C/D/F	N/A	A	B	C/D/F
	Mean	Sd	Mean by letter grade				Mean by "correct" grade		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Black %	0.39	0.26	0.32	0.39	0.44	0.41	0.31	0.38	0.49
Hispanic %	0.44	0.24	0.45	0.42	0.46	0.44	0.40	0.48	0.44
White %	0.08	0.13	0.12	0.09	0.04	0.07	0.13	0.07	0.03
Asian %	0.09	0.12	0.11	0.09	0.06	0.07	0.15	0.06	0.04
FRPL %	0.80	0.17	0.78	0.77	0.83	0.83	0.75	0.81	0.82
Ell %	0.12	0.19	0.14	0.10	0.11	0.14	0.13	0.12	0.09
Regents Math VA (percentile)	50	29	66	51	33	49	82	49	17
Regents Math VA σ	0.01	0.21	0.13	0.00	-0.11	0.00	0.23	-0.02	-0.20
SAT Math VA (percentile)	50	29	60	52	42	45	70	48	36
SAT Math VA σ	0.00	0.15	0.06	0.00	-0.05	-0.03	0.11	-0.02	-0.06
Peer quality (percentile)	50	29	60	54	41	43	71	50	38
Peer quality (avg. 7th grade Math σ)	-0.15	0.41	0.01	-0.11	-0.30	-0.24	0.13	-0.21	-0.34
Graduation rate	0.72	0.16	0.82	0.71	0.60	0.73	0.80	0.68	0.62
Average Regents Math σ	-0.18	0.45	0.07	-0.14	-0.39	-0.28	0.22	-0.26	-0.41
Average SAT Math σ	-0.42	0.42	-0.29	-0.37	-0.53	-0.49	-0.03	-0.47	-0.72
Screened	0.24	0.43	0.38	0.20	0.16	0.19	0.40	0.15	0.12
Size	679	858	641	980	928	248	949	885	691
N (school-year)	2,716		733	736	507	740	733	736	507

Notes: This table provides school descriptive statistics for the 2006-2007 to the 2012-2013 schools years. The progress reports were based on data covering these school years. An observation in this sample is a school-year. Column (1) and (2) report means and standard deviations across school-year observations. Columns (3)-(6) report means by letter grades and columns (7)-(9) by the letter grade schools would have received if grades had been actually based on causal estimates of Regents VA.

Table 1.3: Race Gap in Choice of School Quality

	N	Race gap (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: school value added percentile ranking									
Mean Regents VA in top 3 choices	620,975	-14.48***	-12.74***	-7.21***	-11.13***	-10.07***	-8.88***	-4.51***	-7.52***
		(0.04)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)
Mean SAT VA in top 3 choices	620,975	-18.35***	-15.78***	-9.23***	-12.70***	-12.71***	-10.83***	-5.78***	-8.37***
		(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)
Regents VA in school of enrollment	520,275	-17.97***	-16.58***	-8.64***	-12.88***	-13.32***	-12.41***	-5.81***	-9.20***
		(0.07)	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)	(0.09)	(0.09)
SAT VA in school of enrollment	520,275	-20.89***	-18.72***	-9.75***	-13.70***	-15.80***	-14.19***	-6.71***	-9.97***
		(0.07)	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)	(0.09)	(0.08)
Panel B: school value added									
Mean Regents VA in top 3 choices	620,975	-0.09***	-0.09***	-0.06***	-0.08***	-0.06***	-0.06***	-0.03***	-0.05***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mean SAT VA in top 3 choices	620,975	-0.11***	-0.10***	-0.07***	-0.09***	-0.07***	-0.07***	-0.05***	-0.06***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Regents VA in school of enrollment	520,275	-0.12***	-0.11***	-0.06***	-0.09***	-0.08***	-0.08***	-0.04***	-0.06***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
SAT VA in school of enrollment	520,275	-0.11***	-0.10***	-0.06***	-0.08***	-0.08***	-0.08***	-0.04***	-0.06***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Borough FE			X				X		
Zipcode FE				X				X	
Test score controls						X	X	X	X
Mean and max in choice-set					X				X

Notes: This table reports race differences in the quality of the top 3 school choices and of the school of enrollment, as estimated by the coefficient β in equation (1.2). The regressions in the first column correspond to raw race gaps, while columns (2)-(8) progressively add controls for residential location, test scores and school quality available in the students' feasible set. Each row uses a different left-hand side outcome, that is a measure of school quality in applicant's top 3 choices or in the school where she enrolls.

Table 1.4: Correlation Between Elicited Beliefs, School Quality and Mean Achievement Levels

	Elicited belief					
	(1)	(2)	(3)	(4)	(5)	(6)
Value-Added (SD)	0.124***	0.096**			-0.133***	-0.037
	(0.033)	(0.043)			(0.049)	(0.068)
Value-Added (SD) · Non-minority		0.063				-0.194**
		(0.061)				(0.091)
Achievement level (SD)			0.230***	0.160***	0.351***	0.200***
			(0.035)	(0.049)	(0.053)	(0.076)
Achievement level (SD) · Non-minority				0.130**		0.278***
				(0.063)		(0.093)
Non-minority Respondent		-0.010		-0.081		-0.137
		(0.074)		(0.086)		(0.087)
N	849	849	849	849	849	849
Mean response	2.55	2.55	2.55	2.55	2.55	2.55

Notes: This table reports regression estimates of the relationship between elicited beliefs about school quality and school characteristics. Elicited school quality ranges from 1 (bottom quartile of school quality) to 4 (top quartile of school quality). The school attributes considered in the right hand side of the regressions are the average achievement in Regents exams of students enrolled in the school, and the school Regents VA, both expressed in standard deviations (SD) of the cross-school distribution. Even columns allow the relationship between left hand side variables and school attributes to vary across respondent's race, as measured by the interaction between school attributes and a dummy indicating white and Asian respondents.

Table 1.5: Demand Responses to School Letter Grades

	School share		School log share	
	(1)	(2)	(3)	(4)
A	0.15**	0.14***	0.22**	0.31***
	(0.03)	(0.03)	(0.07)	(0.06)
B	0.08*	0.07**	0.06	0.17***
	(0.04)	(0.02)	(0.05)	(0.03)
C	-0.02		-0.19**	
	(0.03)		(0.05)	
D	-0.09	-0.05	-0.35***	-0.12
	(0.05)	(0.03)	(0.05)	(0.06)
F	-0.21**	-0.15*	-0.36**	-0.21*
	(0.07)	(0.07)	(0.13)	(0.08)
Graduation % (SD)	-0.00	-0.07**	0.07	0.01
	(0.04)	(0.02)	(0.08)	(0.04)
College % (SD)	0.02	-0.01	0.01	-0.02
	(0.02)	(0.02)	(0.02)	(0.03)
Graduation % (SD) · Visible	0.02**	0.17***	0.05***	0.27***
	(0.00)	(0.02)	(0.01)	(0.04)
College % (SD) · Visible	0.00	0.03	0.03	0.07**
	(0.02)	(0.01)	(0.03)	(0.02)
Only graded schools		X		X
N	32,190	22,815	15,908	12,470
N schools	458	338	429	334
Average school share	0.606	0.766	0.606	0.766

Notes: This table presents regression estimates of letter grade effects on demand for schools. The dependent variable is the share (or log share in columns (3)-(4)) of students in a demographic cell c and application cohort t ranking the school in their top 3 choices. Demographic cells are defined by the interaction of a student residential borough and baseline test score tercile. The first 5 rows report estimates of the coefficients β_g in equation (1.4) for each letter grade. The other rows the coefficients of a school graduation or college rates in the year prior to when cohort t applies to schools and of their interaction with an indicator (*Visible*) for years in which these statistics were printed on the school directories. Other controls include school-cell fixed effects, year-cell fixed effects, a school average Regents performance and the share of white and Asian students enrolled at the school in the year before cohort t applies to school. Standard errors are clustered at the school-year level. All columns use school-years observations between 2010 and 2014 included, the years in which letters were printed on directories. Columns (2) and (4) restrict the observations to school-year with a grade, so that the omitted category is receiving a grade of C.

Table 1.6: Heterogeneity in Responses to Letter Grades by School Achievement Level

	School share		School log share	
School achievement level:	*Above median*	*Below median*	*Above median*	*Below median*
	(1)	(2)	(3)	(4)
A	0.10*	0.19***	0.19*	0.43***
	(0.04)	(0.03)	(0.08)	(0.09)
B	0.02	0.13***	0.07	0.24**
	(0.04)	(0.02)	(0.06)	(0.06)
C	-0.14**	0.07**	-0.14*	0.01
	(0.04)	(0.02)	(0.05)	(0.06)
D	-0.35	0.03	-0.49*	-0.09
	(0.19)	(0.02)	(0.19)	(0.06)
F	-0.62	-0.08	-0.58**	-0.16
	(0.36)	(0.06)	(0.18)	(0.12)
Graduation % (SD)	0.02	0.01	0.06	0.15
	(0.06)	(0.03)	(0.07)	(0.10)
College % (SD)	0.02	-0.03	0.02	-0.09**
	(0.03)	(0.02)	(0.03)	(0.03)
Graduation % (SD) · Visible	0.04***	-0.01***	0.06***	-0.02
	(0.00)	(0.00)	(0.01)	(0.01)
College % (SD) · Visible	0.03	0.04	0.04	0.12**
	(0.02)	(0.02)	(0.02)	(0.04)
N	14775	14445	8793	6159
N schools	199	204	197	187
Average school share	1.030	0.257	1.030	0.257

Notes: This table presents regression estimates of letter grade effects on demand for schools defined by the coefficient β_g in equation (1.4), distinguishing schools by the mean achievement levels of their students. Columns (1) and (3) restrict the sample to schools with above median student achievement levels and columns (2) and (4) to schools with below median achievement levels. The dependent variable is a school share among students choices (or log share in columns (3)-(4)), defined as the share of students in demographic cell c and application cohort t ranking the school among their first three choices. Demographic cells are defined by the interaction of a student residential borough and baseline test score tercile. Controls are those used in table 1.5 and always include school-cell fixed effects and year-cell fixed effects. The sample includes applicant cohorts from 2010 to 2014 included. Standard errors are clustered at the school-year level.

Table 1.7: Heterogeneity in Responses Across School Attributes

	X_j = "% White"			X_j = "Peer Quality"			X_j = "Achiev. Level"		
	B+H	W+A	All	B+H	W+A	All	B+H	W+A	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
S_{jt}	0.07***	-0.00	-0.02	0.07**	0.00	-0.02	0.06**	-0.00	-0.02
	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$S_{jt} \times X_j$	0.10	0.08**	0.09**	0.10	0.06*	0.08**	0.21*	0.08**	0.09**
	(0.07)	(0.02)	(0.02)	(0.05)	(0.02)	(0.02)	(0.10)	(0.02)	(0.03)
$S_{jt} \times M_c$			0.10**			0.10**			0.08**
			(0.02)			(0.03)			(0.02)
$S_{jt} \times X_j \times M_c$			0.02			0.01			0.11
			(0.06)			(0.03)			(0.07)
N	22,815	22,815	45,630	22,815	22,815	45,630	22,815	22,815	45,630
N schools	338	338	338	338	338	338	338	338	338
Average school share	0.788	0.753	0.771	0.788	0.753	0.771	0.788	0.753	0.771

Notes: This table presents regression estimates of equation (1.7) or a variant that does not include interactions between school attributes and student race. The dependent variable is a school share in students' top three choices, within student demographic cells defined by the interaction of student race, residential borough and baseline test score tercile. Right hand side variables include S_{jt}, a discrete letter grade rank varying from 1 to 5, a dummy X_j indicating whether school j is in the top third of schools in terms of white and Asian enrollment (columns (1)-(3)), peer quality (columns (4)-(6)) or mean achievement levels (columns (7)-(9)), a dummy M_c indicating Black and Hispanic students demographic cells, and their interactions. Controls are those used in table 1.5 and include school-cell fixed effects and year-cell fixed effects. The sample includes applicant cohorts from 2010 to 2014 included. Standard errors are clustered at the school-year level.

Table 1.8: Consequences of Letter Grade Introduction on Ranked and Offered School Attributes

	Grade A	Low grade	Regents VA σ	Regents VA pct	Peer quality pct	White and Asian %	Screened	P(matched) or P(enrolls)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: first choices								
$Post2010 \cdot M_i$	0.049***	-0.049***	0.035***	4.667***	1.487***	0.001	-0.000	
	(0.003)	(0.001)	(0.001)	(0.126)	(0.104)	(0.001)	(0.003)	
$Post2010$	0.045***	-0.024***	0.023***	1.200***	3.188***	0.021***	0.044***	
	(0.002)	(0.001)	(0.091)	(0.001)	(0.068)	(0.001)	(0.002)	
N	502,923	502,923	502,923	502,923	502,923	502,923	502,923	
Black+Hispanic mean	0.225	0.142	0.0495	58.08	67.68	0.249	0.318	
White+Asian mean	0.397	0.0400	0.172	76.32	86.37	0.572	0.568	
Panel B: first 3 choices								
$Post2010 \cdot M_i$	0.074***	-0.112***	0.031***	4.488***	1.556***	0.001	0.005*	
	(0.003)	(0.002)	(0.001)	(0.090)	(0.082)	(0.001)	(0.003)	
$Post2010$	0.051***	-0.062***	0.028***	1.588***	3.555***	0.022***	0.031***	
	(0.002)	(0.001)	(0.001)	(0.067)	(0.057)	(0.001)	(0.002)	
N	502,923	502,923	502,923	502,923	502,923	502,923	502,923	
Black+Hispanic mean	0.484	0.340	0.0388	56.43	65.18	0.241	0.558	
White+Asian mean	0.645	0.107	0.155	74.43	84.16	0.550	0.729	
Panel C: offers								
$Post2010 \cdot M_i$	0.026***	-0.052***	0.029***	4.492***	1.582***	-0.006***	-0.005*	-0.000
	(0.003)	(0.002)	(0.001)	(0.142)	(0.118)	(0.001)	(0.003)	(0.002)
$Post2010$	0.017***	-0.042***	0.025***	1.967***	4.108***	0.021***	0.036***	-0.031***
	(0.002)	(0.001)	(0.001)	(0.108)	(0.084)	(0.001)	(0.002)	(0.001)
N	459,617	459,617	459,617	459,617	459,617	459,617	459,617	502,923
Black+Hispanic mean	0.144	0.211	-0.0264	46.75	53.93	0.167	0.224	0.929
White+Asian mean	0.276	0.0857	0.123	69.78	79.49	0.503	0.443	0.919
Panel D: enrollment								
$Post2010 \cdot M_i$	0.022***	-0.040***	0.030***	4.549***	1.488***	-0.004***	-0.002	-0.038***
	(0.003)	(0.002)	(0.001)	(0.157)	(0.130)	(0.001)	(0.003)	(0.002)
$Post2010$	0.012***	-0.041***	0.022***	1.530***	4.106***	0.017***	0.033***	0.008***
	(0.003)	(0.002)	(0.001)	(0.126)	(0.099)	(0.001)	(0.002)	(0.002)
N	422,654	422,654	422,654	422,654	422,654	422,654	422,654	502,923
Black+Hispanic mean	0.142	0.213	-0.0237	47.13	53.50	0.176	0.157	0.885
White+Asian mean	0.262	0.105	0.116	68.79	77.61	0.494	0.214	0.728

Notes: This table presents pooled differences in differences estimates of the differential changes in the attributes of school choices (panels A and B), school offers (panel C) and enrollment schools (panel D) by student race after the introduction of letter grades. The sample includes students applying to enroll in 9th grade between 2006 and 2014. Controls include gender, ell status, subsidized lunch status and fixed effects for combinations of student borough and baseline test score terciles.

Table 1.9: Model Estimates - Summary Statistics

	By race			By 7th grade Math tercile		
	Black	Hispanic	White	Low	Median	High
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: first step						
δ_{cjt} SD	24.8	21.5	25.6	22.1	22.0	27.5
δ_{cjt} range	160.5	130.7	133.9	143.5	133.0	144.0
Corr(δ_{cjt}, VA)	0.33	0.37	0.60	0.24	0.47	0.61
Corr(δ_{cjt}, Peer quality)	0.48	0.49	0.74	0.34	0.63	0.77
Corr(δ_{cjt}, % white)	0.33	0.42	0.70	0.32	0.51	0.65
Panel B: second step - preferences						
γ_c	5.2	4.2	5.4	3.2	4.8	6.7
β_c^{white}	2.0	1.5	1.1	1.2	1.3	2.0
$\beta_c^{peerquality}$	4.4	4.4	3.7	3.7	4.1	4.7
$\tilde{\xi}_{cj}$ SD	17.1	14.2	20.3	14.3	16.4	20.8
$\tilde{\xi}_{cj}$ range	104.9	86.9	107.2	92.4	97.7	106.8
$\tilde{\xi}_{cj}$ skewness	0.06	0.23	0.71	0.15	0.31	0.58
Corr($\tilde{\xi}_{cj}$, VA)	0.16	0.21	0.54	0.07	0.34	0.52
Corr($\tilde{\xi}_{cj}$, Peer quality)	0.30	0.30	0.67	0.15	0.48	0.66
Corr($\tilde{\xi}_{jc}$, % white)	0.14	0.26	0.68	0.18	0.38	0.56
Panel C: second step - beliefs						
μ_{cL}	-0.15	-0.11	-0.08	-0.10	-0.13	-0.09
μ_{cH}	-0.01	0.04	0.18	-0.09	0.05	0.26
σ_{cL}^{-1}	2.12	2.01	2.39	1.98	2.09	2.45
σ_{cH}^{-1}	2.26	2.44	3.17	2.11	2.74	3.09
Absolute bias	0.48	0.49	0.46	0.53	0.47	0.42

Notes: This table summarizes the model estimates. Panel A reports summary statistics for the estimates of the mean school utility δ_{cjt} obtained in the first step. Panel B reports the second step estimates of the preference parameters $\gamma_c, \beta_c, \xi_{cj}$ and panel C of the prior moments $\boldsymbol{\mu_c}, \boldsymbol{\sigma_c^{-1}}$ taking a weighted average of cell-specific estimates across cells sharing the same covariate (race or baseline test score), using weights proportional to cell size.

Chapter 2

Overcoming Racial Gaps in School Preferences: the Effect of Peer Diversity on School Choice

Written jointly with Clemence Idoux

2.1 Introduction[1]

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

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

2011).

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

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

We first document the determinants of racial gaps in choices by surveying 3,000 parents and guardians of high-school applicants during the application cycle 2022-2023.[2] This survey delved into various factors that could influence high school choice: sources of information, significance of different school attributes, awareness of school offerings, perceptions of academic performance and admission probabilities to competitive schools, aspirations for higher education, and perceptions about discrimination. The survey also included a vignette study to sep-

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

arately estimate preferences for various school characteristics, such as high-achieving peers, school safety, and racial composition of the student body.

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

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

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

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

The instrumental variable (IV) estimates show that attending a more diverse middle school significantly affects high school choice. Based on our IV estimates, Black and Hispanic students who attend middle schools with one SD (26 p.p.) more white and Asian peers choose high schools which enroll 3 p.p. more white and Asian students and have 0.06 SD higher value-added on average. This corresponds to a reduction of the racial gaps in value-added and racial make-up of preferred school choices of approximately 30%. High school choices of white and Asian students are less affected by middle school peer diversity: enrollment in a middle school with one SD (29 p.p.) more Black and Hispanic peers increases the Black and Hispanic peer share of the average high school choice by 2 p.p. The differences in high school choice characteristics induced by middle school peer diversity translate into differences of similar magnitude in the characteristics of the high school offered in the centralized match.

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

This paper builds on research that considers how interactions with peers of different backgrounds may impact social attitudes and beliefs (Corno et al., 2019; Boisjoly et al., 2006; Carrell et al., 2019; Rao, 2019), showing how school integration affects preferences for contact with other races in a real, high-stakes, setting. It also contributes to the literature on frictions in school choice (Kapor et al., 2020; Arteaga et al., 2021; Ainsworth et al., 2023) and their unequal impact by socio-economic status (Hastings and Weinstein, 2008; Hoxby and Turner, 2013, 2015; Allende et al., 2019; Pathak and Sönmez, 2008). Our study provides novel insights by documenting directly the existence of information frictions and their role in shaping inequality in access to high quality schools. Thanks to a unique combination of survey data and administrative records, we also show how peer networks reduce these frictions. Moreover, our findings suggest that biases in beliefs and sophistication may play a lesser role in explaining racial disparities in school applications.

Finally, our findings highlight the importance of path-dependence in school choice. As such, we speak to the literature on the impact of school integration reforms (Idoux, 2021; Laverde, 2020; Bjerre-Nielsen and Gandil, 2020), uncovering a potential dynamic effect mostly overlooked so far.[3] If exposure to more diverse peers is important in shaping student preferences, reducing school segregation in earlier school grades could lower school segregation in later grades through changes in demand for schools.

2.2 Institutional Setting

2.2.1 NYC School Assignment System and School Segregation

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

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

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

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

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

Each academic program ranks applicants using a set of eligibility and admission criteria based on residential location and on academic achievement. Geographic eligibility and admission criteria are more stringent at lower grade levels. At the elementary level, 85% of schools only admit students in their school zone and the remaining 15% non-zoned schools still give priority to students in their zone. NYC middle schools are intended to serve students residing in their local district, with 83% of middle school programs having zone or district eligibility requirements across the 32 districts.[5] On the contrary, high schools are open to all students in the city, with only approximately 39% of schools giving priority to students residing in their borough or zone. Finally, high school and middle school programs rely on academic admission criteria to the same extent: approximately a third of these programs rank individual students based on prior grades, standardized test scores, talent test scores, and behavioral measures, in addition to the eight highly selective specialized high schools.[6]

In line with the higher importance of geographic priorities in earlier school grades, racial segregation across schools is also higher in elementary and middle schools. Appendix Figure B.1.1 compares overexposure to Black and Hispanic peers for students of different races and grades. Across all grade levels, Black and Hispanic students attend schools which enroll disproportionately more Black and Hispanic students than their representation in the city's student population. For example, on average Black and Hispanic students attend high schools where the proportion of Black and Hispanic students is 11 p.p. higher than the city's average of 68%. In contrast, white and Asian students typically go to high schools with 23 p.p. fewer Black and Hispanic peers than the city average. Moreover, the over-exposure of Black and Hispanic students to Black and Hispanic peers is more marked in elementary school than in middle school and high school. Black and Hispanic students attend respectively elementary schools with 28 p.p. more Black and Hispanic than the elementary school population, middle schools with 25 p.p. more Black and Hispanic than the middle school population and high schools with 23 p.p. more Black and Hispanic than the high school population.

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

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

2.2.2 Race Differences in School Choice

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

In Figure 2-1, we document that preferences for school attribute vary by race even for students with similar baseline achievement. The figure compares mean characteristics of the high school listed first by applicants of different race as a function of their 7$^{\text{th}}$ grade test scores.[7] Across all racial groups, higher test score applicants prefer schools that enroll more white and Asian students and students with higher baseline achievement. These schools are also more likely to screen applicants on academic achievement and have higher math value-added on average.[8] Nonetheless, for any given level of baseline achievement, white and Asian applicants favor more these characteristics than Black and Hispanic applicants. Panel A of the figure shows that white and Asian students first school choices have on average 20 p.p. more white and Asian students and enroll students with 0.25 standard deviations higher 8$^{\text{th}}$ grade test scores than Black and Hispanic students' first choices. Similarly, Panel B shows that white and Asian students are 10 p.p more likely to choose as first choice a school that screens on academic performance than Black and Hispanic students and favor schools with 0.30 higher math value-added standard deviations on average.

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

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

[8]Schools' math value added are estimated using an OLS regression of student SAT scores on school fixed effects controlling for demographics, baseline test scores and student assignment risk to the school as in Angrist et al. (2021).

Asian applicant whose test scores is worse by 0.47 standard deviations. This suggests that race plays a different role than test scores and geographic residence in school choice.

2.3 Conceptual Framework

To shed light on the potential underlying causes of racial differences in school choice, we introduce a simple school choice framework where applicants choose which high schools to apply solving a portfolio choice problem as in Chade and Smith (2006). Specifically, each applicant, indexed by i, chooses a ranked-ordered list (ROL) of schools $R \in \mathscr{R}_i$, where $\mathscr{R}_i$ comprises the sets resulting from all the k-permutations of A_i, the set of schools the applicant is aware of, which includes her outside option outside the traditional public sector denoted by school 0.[9] Each ROL can be mapped to a lottery over high schools whose weights depend both on the ordering of schools in the list and on applicant beliefs about admission probabilities.[10] Hence, applicants choose their ROL to maximize their expected utility, which depends on their expected utility for enrolling in any given school, the lottery over schools induced by the ROL and their beliefs about admission probabilities, and the cost of submitting the ROL:

$$\max_{R \in \mathscr{R}_i} \sum_{s \in A_i} p_{is}(R, \widehat{q}_i) E[u_{is}(\theta_i, X_{is}) | \mathscr{I}_i] - c_i(R) \tag{2.1}$$

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

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

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

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

dent knowledge of school admission rules, of demand for the program among other students, and of her relative ranking in terms of priority and test scores.

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

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

1. *Differences in preferences* - $u_{is}(\theta_i, X_{is})$. Applicants may put different weight on different school features, even when these attributes are perfectly observed.[11] For instance, experimental evidence from Hailey (2022) reveals that parents tend to prefer schools enrolling students of similar races or ethnicity.

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

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

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

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

3. *Differences in perceived probabilities of admissions* - $\widehat{q}_i$. Applicants may differ in their probability of admission at each school, as well as in their subjective belief about this probability. Admission probabilities differ across students due to differences in priority and test scores. Nonetheless, holding fixed these attributes, subjective belief about admission probability may still differ if applicants do not hold rational expectations but

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

hold biased beliefs instead.[12] Applicants from different socio-economic and racial background might hold different subjective beliefs about admission chances because they differ in their degree of optimism and confidence in their relative ability.[13]

2.4 Data

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

2.4.1 Administrative Data

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

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

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

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

Column (1) of Table 2.1 includes summary statistics for the sample of middle school applicants who were also observed applying to 9th grade seats within NYC public schools. The sample is racially diverse and includes many low SES students (72% are eligible for subsidized lunch). The average student attends middle schools where 60% of peers are Black or Hispanic.

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

- High school quality: we measure it using value-added models (VAM) which capture the contribution of schools to student achievement. Our main measure of achievement is SAT math scores. In particular, we adopt a recent methodological improvement in the school VAM literature introduced by Angrist et al. (2021) and referred to as *Risk-controlled value-added (RC VA)*. [14]

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

- Measures of high school selectivity and popularity: we construct a dummy indicating high schools that have at least one program screening students on the basis of test scores, audition or other ability assessments. We also construct a measure of popularity using the ratio of rejected to accepted applicants at a school. [15]

On average students rank 8 high school programs on their list. Top choices tend to be of higher quality, more selective and popular than the average school in the city. On average, applicants' top 3 program choices have a SAT math VA which is 0.8 standard deviation higher than the average school in the city. In addition, 70% of applicants rank at least one screened program among their top three choices which reject on average 4.15 applicant for each admitted applicant, compared to a rejection rate of 1.36 for the average school in the city.

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

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

2.4.2 Survey Data

We conducted a post-application survey of guardians of 9th grade applicants, in partnership with the NYC Department of Education. The survey was conducted from February 17th to March 6th, 2023, after applicants had submitted their high school application but before they had received their match offer. The survey was sent electronically to the email addresses provided during the high school application. Respondents could answer in English, Spanish or Chinese. Upon completion of the survey, participants that had answered at least one survey question were sent a 10-dollar amazon gift card. Only parents/guardians were permitted to respond to the survey.[16]

We selected 21,401 potential participants who were general education high school applicants enrolled in a NYC traditional public middle school and who had baseline test scores. Of the participants, 18% completed some questions of the survey, and 14% – referred to as respondents in Table 2.1 – answered over half of the questions. As shown in column (5) of Table 2.1, survey respondents are more likely to be white or Asian and less likely to be low-income, compared to the general NYC high school applicant population (column (1)). Furthermore, respondents' students scored higher on tests than the average NYC student.

The survey examined various dimensions of the choice process for families, including sources of information, essential school characteristics, knowledge of school options and their features, perceptions of admission probabilities and their influence on choice, perceptions of discrimination and its impact on decision-making, and educational aspirations. Additionally, the survey conducted a vignette experiment, described in more details in section 2.5.1, which aimed at disentangling families' relative preference for different school characteristics and uncover potential statistical discrimination. This comprehensive approach enables us to systematically document differences in the three main potential drivers of choice differences outlined in our conceptual framework: differences in preferences, information about schools and beliefs about admissions probabilities among different racial and socio-economic groups. The complete survey, as presented to participants, together with detailed information on the construction and randomization of the questions, is available in Online Appendix B.2.

[16]75% of the survey respondents reported that parents/guardians played an essential role in their student's high school selection.

2.5 Family Survey Results

In this section, we use survey answers to comprehensively investigate the reasons behind observed racial gaps in school choice. Motivated by the framework presented in section 2.3, we consider three main channels: differences in preference for school characteristics in section 2.5.1, differences in information about school existence and attributes in section 2.5.2, and differences in beliefs about likelihood of admissions in section 2.5.3.

2.5.1 Racial Gaps in Preferences for School Characteristics and Peers

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

Stated Preferences

As suggested by Figure 2-3, respondents from all racial groups prioritize the same school features when selecting schools. Most respondents report the same six school features as most important when selecting a school: safety, academic progress of students at the school, college and graduation rates, commuting time, the number of AP classes offered, and whether their students would feel they belong. Each of these features was mentioned by at least 20% of respondents as one of the three most important attributes of a school. The race differences in the proportion of respondents mentioning each school feature as most important are not statistically significant for the top three school features (safety, academic progress, and college and graduation rates). This is supported by Appendix figures B.1.3, wherein a majority of families from various racial backgrounds express that their child would thrive in an academically rigorous school environment, yet may not fit as well in school with significant disparity in peer achievement levels.

The emphasis on academic performance of schools by respondents across all racial groups may reflect their high aspirations for their children's future education. As depicted in Figure 2-5 panel (a), more than half of respondents agree that attending college is crucial for achieving success in life, and 85% express a desire for their children to obtain at least a four-year college degree. Educational aspirations among respondents from different racial groups are similar. When controlling for district of residence and baseline test scores, Black and Hispanic respondents are equally likely as white and Asian respondents to desire that their children attend college for at least four years. However, they are 10 percentage points less likely to consider college important for achieving success in life, potentially due to differences in personal experiences or trajectories. This evidence suggests that, overall, families across demographic groups share the same school selection criteria and have comparable aspirations for their children's education.

Revealed Preferences: Vignette Experiment

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

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

The academic performance information provided to participants varied based on the treatment arm they were assigned to: 60% of participants received precise information about the schools' academic performance, while 40% received imprecise information. The precise academic information consisted of the 4-year graduation rate and the college and career program enrollment rate. The imprecise information consisted of the share of students that earned enough credits in 9$^{\text{th}}$ grade to be on track for graduation. Participants who received precise information were presented with either a high-performing or low-performing school, whereas those who received imprecise information were always presented with a school with median academic performance. Table B.2.14 outlines the precise information presented to participants in both the precise and imprecise academic information treatment arms.[17]

The experiment employed a factorial design to randomly combine these characteristics, resulting in 16 unique combinations for the precise-information treatment arm and 8 unique combinations for the imprecise-information treatment arm. The two schools for the first part

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

of the experiment were randomly selected without replacement from these unique combinations. For the second part of the experiment, two sets of three distinct schools were randomly chosen without replacement.[18] Figure 2-4 shows an example of the vignettes as seen by survey participants.

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

$$u_{is} = \alpha Z_i + \beta X_s \times Z_i + e_{is}$$

.

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

We combine the absolute preference for schools and relative rankings of schools provided by respondents to estimate their respective weights for different school characteristics.[19] The scale and location of the utility is thus normalized by respondents' likelihood of listing the schools. For a respondent indicating a likelihood of a to list a school, it implies that $u_{is} \in [a-0.5, a+0.5]$. The full parameter vector $\theta = (\beta, \sigma)$ is estimated using a Gibbs sampler to maximize the likelihood of observing the responses to both questions.

Table 2.3 presents the estimates in likelihood units with respect to a racially-balanced, low-safety and low-performing school. Column (1) shows that school academic and safety ratings are the primary factors that influence families' school choices for all respondents. Holding all else constant, a high academic rating increases utility by 1.4 points, while a high safety rating increases it by 0.7 points. The magnitude of these effects is consistent across racial groups. Nonetheless, contrary to Black and Hispanic respondents, white and Asian respondents also hold some preference over schools' demographic make-up.[20] White and Asian respondents are 0.28 points more likely to list a majority white and Asian school, and respectively 0.28 and 0.44

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

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

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

less likely to list a majority Hispanic or Black school, compared to a racially-balanced school. Conversely, there is no evidence to suggest that Black and Hispanic respondents take into account the demographics of schools, as the coefficients on the racial make-up of schools are not statistically significant.

This preference over the school's racial composition is partly due to statistical discrimination, whereby respondents infer the academic performance of a school based on its demographic makeup. The comparison between columns (1) and (2) of Table 2.3 suggests that respondents who receive less precise information about schools' academic performance are more influenced by the schools' racial composition in their rankings.

This effect is observed across all racial groups. In the case of Black and Hispanic respondents, the coefficients pertaining to the school's demographic makeup are only significant when they receive imprecise information about the schools' academic performance. In such instances, they are approximately 0.2 points more likely to list racially balanced schools than other types of schools. Similarly, white and Asian respondents exhibit stronger preferences for the racial composition of their peers when they receive imprecise information about the schools' academic performance. They demonstrate a heightened preference for majority white and Asian schools, as indicated by the corresponding coefficient increasing from 0.28 to 0.36. Additionally, they show a stronger aversion towards majority Hispanic or Black schools, with the corresponding coefficients decreasing from -0.22 to -0.59 and from -0.44 to -0.70, respectively.

Overall, the vignette experiment suggests that the racial gaps in choice is unlikely to be due to differences in preferences for peer achievement or safety, in line with the evidence shown in the previous subsection, but may be due to homophily. White and Asian respondents show consistent preference for schools that enroll more white and Asian students, while Black and Hispanic respondents are less likely to choose majority white and Asian students when they have more imprecise information about schools' academics performance. As suggested by Appendix Figure B.1.4, this racial preference for peers may stem from concerns regarding racial discrimination by students' teachers and peers, which are more common among Asian, Black, and Hispanic families compared to white families. Indeed, among the 23% of respondents who reported that their students might face discrimination from their teachers or peers, 70% mentioned that these concerns influenced their high school choices.

2.5.2 Racial Gaps in Information About Schools

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

Awareness Sets

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

Panel (a) of figure 2-6 presents the share of schools respondents were aware of among the schools presented to them, categorized by different types of schools. Panel (b) shows racial differences controlling for district of residence and baseline test scores. On average, respondents from all racial backgrounds appeared to be familiar with approximately one-third of the schools presented to them. While there were no notable differences in the total number of schools respondents were aware of, Figure 2-6 reveals significant racial disparities in the types of schools respondents are familiar with. Compared to their white and Asian counterparts, Black and Hispanic families appear to be less aware of majority white and Asian schools, high quality and high performing schools, as indicated by having high-VAM or high college enrollment and graduation rates. Black and Hispanic respondents are aware of 4.3 percentage points fewer majority white and Asian schools, 4.1 percentage points fewer high-VAM schools and of 3.0 percentage points fewer schools with high college enrollment and graduation rates. In contrast, Black and Hispanic respondents are aware of 4.3 percentage points more high Black and Hispanic schools, 4.8 percentage points more low-VAM schools.

A potential reason why Black and Hispanic families are less likely to know of high quality school options is that they rely on different sources to gather information about schools. In Figure B.1.6 we document the difference in the probability that a given information source is listed among the most used by respondents' race. Black and Hispanic families are 20 percentage points less likely to rely on networks of family and friends to collect information about schools compared to white respondents, and 5.6 percentage points less likely to rely on parent networks within their student middle school. They also appear less likely to use information sources that are costlier and more time consuming, such as attending individual high school information sessions and browsing on websites different from the official NYC DOE website. On the contrary, they are more likely to rely on institutional resources provided by their student middle school, such as guidance counselors and other school staff and middle school information sessions, and equally likely to use DOE online resources.

Accuracy of Information

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

Figure 2-7 suggests that families have somewhat accurate information about certain school features that are important to them when selecting a school but not about others. The left chart in panel (a) shows that respondents are more likely, on average, to correctly rank the two schools they are presented based on school safety, college and graduation rates, commuting distance, and peer preparedness than if they had guessed randomly. Respondents' most precise information is about commuting time, with respondents 18.4 percentage points more likely to accurately guess the school with the shortest commuting time compared to guessing randomly. The left chart in panel (b) provides additional evidence that families possess knowledge about these features: respondents' quartile rankings of schools are positively correlated with actual quartile rankings of schools.

Families, instead, appear to be quite misinformed about which schools have higher value-added.[21] In panel (a), respondents are 9.1 percentage points more likely to correctly identify, out of two schools, the school that best contributes to student learning than if they had guessed randomly. However, they do not accurately rank the school value-added compared to other schools in their borough, as suggested by the non-significant correlation between respondents' ranking and actual ranking in panel (b). Similarly, applicants are not well-informed about which schools offer more AP classes, as suggested by non-significant estimates in both panels. In summary, this evidence is consistent with the view that families are better informed about aspects that are more easily observed, such as where the school is located or the type of students it enrolls, but are less well informed about value-added which is arguably harder to observe.

The schools selected for these questions are well-known in the applicant's neighborhood,

[21] To elicit families' beliefs about school value-added we ask them to compare schools in terms of which one is best for promoting student learning.

even though they may not necessarily belong to the respondent's awareness set.[22] Even when restricting the sample to the schools that are for sure in the respondent's awareness set, information accuracy improves only for easily observed attributes, such as commuting time and peer achievement levels, but not for school value-added. We show this in appendix Figure B.1.5. Panel (a) reports the raw average probability of answering correctly in the pairwise comparison when using all questions (first bar), when restricting to questions in which one school is for sure known (second bar) and when we know that both schools are known (third bar).[23] Panel (b) reports the rank-rank correlation coefficient, restricting the sample to schools we are sure the applicant knows in the second bar.

While families appear substantially misinformed about school characteristics such as value-added, differences in information accuracy across race are unlikely to drive inequalities in application behavior. The right chart in panel (a) shows that the race difference in the probability of guessing correctly in pairwise school comparisons is never significantly different from zero. The right chart in panel (b) instead shows that Black and Hispanic applicants' beliefs about peer achievement and college rates are less correlated with the actual school rankings, and this is driven by their lower propensity to select extreme answers. However, there is no significant difference in belief accuracy about school value-added across respondent race. In appendix Table B.1.3 we pool the answers to the two sets of questions by regressing an indicator for answers that are approximately correct on respondent race, baseline test score and district of residence, finding no significant difference in information accuracy by race.[24]

2.5.3 Racial Gaps in Admission Beliefs

Finally, differences in belief about admission probabilities at competitive schools might contribute to racial gaps in application. When deciding where to apply, applicants may exclude

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

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

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

programs from their lists if they perceive their chances of admission as being too low.[25] The survey provides direct evidence supporting this possibility. According to column (1) of Table 2.4, 16% of survey respondents indicated that they did not apply to their "dream program", i.e. their preferred program if guaranteed admission.[26] Nevertheless, we do not find evidence that racial differences in pessimism about admission chances to most preferred programs is a channel underlying racial gaps in applications.

Panel A of Table 2.4 shows that there is no significant difference in the likelihood of Black and Hispanic applicants applying to their favorite programs compared to white and Asian applicants. Controlling for the respondent's "dream program", the difference in application rates to the dream school, as reported in column 1, is less than 1.5 percentage points and lacks statistical significance. Additionally, there appears to be no disparity in pessimism about admission chances as applicants from all racial groups are equally inclined to apply to their favorite program when faced with the same admission probability. This finding suggests that applicants' actual admission probabilities are equally associated with their beliefs about admission chances. This association is further supported by column 2, where a 1 percentage point increase in admission probability corresponds to a 0.16 percentage point increase in reported admission belief, without any significant differences across racial groups.

As suggested in Panel B of Table 2.4, the similarity in beliefs about admission chances across racial groups partly arises from similar beliefs about their students' relative performance compared to students attending sought-after schools. While at the bottom of the performance distribution, Black and Hispanic respondents are more optimistic about their kids relative academic performance, their higher optimism fades at higher levels of student achievement. For each increase in actual performance by one tercile, Black and Hispanic beliefs about their students' relative performance increase by 0.121 less than those of white and Asian respondents. Moreover, Column 2 shows no racial differences in beliefs about relative academic performance conditional on student achievement when respondents are asked how their students would compare to other students enrolling in high-demand schools, who are typically higher-achieving. These results suggest that Black and Hispanic families are not less likely to apply to popular programs enrolling high-achieving students due to under confidence about their admission chances or their student academic ability.

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

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

2.6 Middle School Effects on the Racial Choice Gap

Our results so far indicate that families prefer schools enrolling students with similar demographics and that Black and Hispanic families choose lower value-added schools despite caring equally about school quality. Most of the Black and Hispanic shortfall in preferences over quality is explained by racial gaps in awareness of higher quality schools, with the remainder possibly explained by differences in preferences over the demographic composition of schools.

Middle schools may contribute to leveling the playing field for families across income and racial groups by offering more equitable access to information and providing a setting where families from diverse backgrounds can interact. First, middle schools serve as a place for parents to share information: 75% of respondents reported discussing high school applications with other parents at their student's middle school at least once, 26% engaged in such discussions more than five times and over a quarter indicated other parents as one of their most important information sources. Second, middle schools are an institutional source of information about high school applications: they organize information sessions and school staff may provide guidance to families during the application process.[27] Finally, diversity within earlier grades may attenuate preferences for more homogeneous peers in high schools, which drive part of the differences in application behavior across racial and socio-economic groups.

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

2.6.1 Correlating Peer Exposure and School Choices

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

$$Y_i = \alpha C_i + X_i'\Gamma + u_i \tag{2.2}$$

[27] 16% of families cite middle school sessions as one of the most important sources of information, while 26% of respondents overall, and over 30% of low-income, Black, and Hispanic families rely on middle school staff as one of the main sources of information about high schools

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

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

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

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

Appendix Table B.1.6 shows the correlation between middle school peer diversity and additional attributes of high school top choices. It reports OLS estimates of α in equation (2.2), both when C_i indicates having a majority of other-race peers (top rows), and when it indicates the share of other-race peers in middle school (bottom rows). Here we consider the correlation with the average attributes of applicants top 3 choices, rather than first choices as in Figure 2-2. In addition to the patterns highlighted in the figures, the table also shows that, for minority students, having more white and Asian middle school peers is associated to choosing more popular schools. On the contrary, white and Asian students with more minority peers choose schools with lower achieving peers and are marginally less likely to choose less popular schools.

These patterns may reflect selection bias due to students with different high school preferences sorting in middle schools with different racial composition. In the next section, we present our instrumental variable approach to identify the causal effects of attending middle schools with a more diverse set of peers.

2.6.2 Instrumental Variables Framework

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

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

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

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

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

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

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

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

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

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

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

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

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

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

.

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

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

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

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

. Intuitively, μ_i controls for any variations in offered peer race that is due to applicant type θ_i.

Hence, once controlled for μ_i, any remaining variation in offered peer race is due solely to the tie-breaking randomness in the match.

The research design deployed here is thus a two-stage least squares (2SLS) procedure that uses Z_i to instrument for C_i, controlling for the expected other-race share μ_i. We also control for local-linear functions of non-lottery-school tie-breakers; these functions employ the bandwidth used to define $\tau_i(\delta_N)$.[28]

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

$$Y_i = \beta C_i + \kappa_2 \mu_i + \sum_s g_s(R_{is}) + X_i' \Gamma_2 + \epsilon_i \tag{2.3}$$

$$C_i = \gamma Z_i + \kappa_1 \mu_i + \sum_s h_s(R_{is}) + X_i' \Gamma_1 + \nu_i \tag{2.4}$$

. Because β might differ by race, we estimate this system of equations separately by race. First and second stage models control for linear control functions $g_s(\cdot)$ and $h_s(\cdot)$ are linear control functions of the running variables R_{is} at non-lottery programs.[29] Both stage models also include a set of baseline covariates, denoted X_i. [30]

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

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

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

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

.

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

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

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

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

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

For all the 2sls estimations, our sample consists of middle school applicants with non-degenerate variation or risk for the continuous instrument. That is, the analysis is restricted to applicants who have risk of being assigned to more than one other-race peer share value. Appendix Table B.1.1 describes the restrictions applied to construct this experimental sample with greater detail. Columns (2)-(4) in Table 2.1 compare demographics, other-race shares in middle schools and high school choices of students in the experimental samples to those of the universe of students observed applying to both middle school and high school in NYC in the study period. While Black and Hispanic students are slightly over represented in the experimental sample, the sample appears to be quite similar to the population of applicants in column (1).

Appendix tables B.1.4 and B.1.5 report a set of results meant to validate our research design. The first panel of both tables checks whether differential attrition may lead to selection bias. Virtually all the middle school applicants in our analysis sample are observed enrolling in 6th grade within the public school system, while only 89% of them subsequently apply to enroll in a public high school in NYC. Both tables show that the likelihood of observing these outcomes is unrelated to the majority other-race offer (Table B.1.4) and offered other-race share instruments (Table B.1.5).

A second set of diagnostics evaluates covariate balance. In both tables, Panel B reports coefficients on offer instruments from regressions of covariates on the instruments, with appropriate controls for estimated μ_i and for functions of non-lottery program tiebreakers. For the discrete instrument in Appendix Table B.1.4, the estimates show no statistically significant relationships between the majority other-race offer instrument and baseline covariates. For the continuous instrument in Appendix Table B.1.5, the estimates show small differences in baseline math test scores and subsidized lunch status. Nonetheless, the magnitudes of these differences seem unlikely to lead to substantial omitted variables bias. In any case, all 2SLS estimates are from models that include the baseline covariates listed in the table as controls.

2.6.3 2SLS Estimates on School Choices

Black and Hispanic students with higher shares of white and Asian middle school peers apply to high schools that enroll fewer Black students and more white and Asian students. This is documented in Table 2.5 which reports the 2SLS estimates of attending a majority-white and Asian middle school (top rows in each panel) and of attending a middle school with a 10 p.p. higher share of white and Asian peers (bottom rows) on Black and Hispanic high school choices. The table shows estimates separately for the top three choices in Panel A and for all the choices in an applicant's list in Panel B.[31] To account for a change in the number of choices, the table also reports the effect on the length of rank order lists.

The estimated effects indicate that attending a middle school with a higher proportion of white and Asian peers significantly influences the overall application profile of Black and Hispanic students. As a result of attending a majority-white and Asian middle school, the first three high school choices of Black and Hispanic applicants have on average 3.5 p.p. fewer Black students and 5.9 p.p. more white and Asian students. This corresponds to an increase in the chosen share of white and Asian students of more than 20%. Similarly, attending a middle school with 10 p.p. more white and Asian students decreases the share of Black students in top school choices by 0.8 p.p. and increases the share of white and Asian students by 1.1 p.p. The table also shows that attending white and Asian middle schools induces Black and Hispanic students to rank schools enrolling higher achieving peers, plausibly because white and Asian students tend to have higher test scores. All these effects are significant at the 5% level. The magnitude and significance of these effects are similar when considering all the high school choices, although results are larger when looking at the most preferred schools.

Middle school peer diversity also impacts other dimensions of choice. In particular, attending majority-white middle schools (middle schools with 10 p.p. more white students) increases average value added in top 3 choices by 0.15 SD (0.02 SD). Nonetheless, other-race peers seem to have little to zero effects on the popularity or ranked high schools and the probability of applying to a program that uses screened admission methods. Finally, the list length is unaffected suggesting that students are changing most of their choices within a fixed-length list rather than adding extra schools. The pattern and magnitude of the effects of exposure to white and Asian peers are similar when estimated independently for Black and Hispanic students in tables B.1.7 and B.1.8.

The estimated peer effects on high school choice attributes for white and Asian students are reported in Table 2.6. White and Asian students' high school choices are affected by the racial make-up of their middle schools by a lesser degree. First, we observe much smaller effects on the racial composition of high school choices: attending majority-minority middle schools increases the share of Hispanic students in choices by 1.9 p.p. but only when focusing on all ranked schools, while it has no significant effect on the average share of Black students. Similarly, the magnitude of the decrease in the chosen same-race share is about half of what we estimated for Black and Hispanic students. Effects are larger when considering all choices rather than the top 3 choices, contrary to what we observed for Black and Hispanic students. Since list

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

length is, if anything, shortened, these estimates suggest that white and Asian students respond to a change in middle school peer diversity by only modifying their bottom high school choices, while contact with white and Asian students seem to affect Black and Hispanic students top choices.

Attending middle schools with larger shares of Black and Hispanic students induces white and Asian students to rank schools with lower achieving peers. This negative effect, while smaller in magnitude, is comparable to the increase in chosen peer achievement by Black and Hispanic students observed in Table 2.5. The effects on choices popularity and screening method are non-significant, while the effect on school value-added is negative and significant and comparable in magnitude to the positive peer effect found for Black and Hispanic students.

Overall, attending a majority-white and Asian middle school closes respectively about half of and two-thirds of the racial gaps in the school value-added and in the racial composition of school choices, conditional on baseline achievement. A comparison of the 2SLS estimates with the OLS estimates in Table B.1.6 reveals that OLS estimates of peer effects are not extremely biased, especially for Black and Hispanic students. 2SLS estimates of white and Asian peer effects are less than 40% smaller than OLS estimates for the racial composition of school choices and not statistically distinguishable for peer achievement and school value added.

Table 2.7 shows that differences in high school choices induced by middle school peer diversity translate into differences in high school offers for all students. Estimated effects on the offered high school's racial make-up and peer achievement are generally larger for Black and Hispanic students than for white and Asian students. Black and Hispanic students who attend a majority-white and Asian middle school receive as a result high school offers with 5.6% more white and Asian students on average. Nonetheless, this increase in high school peer diversity partly comes at the cost of a statistically significant increase in the probability of being unmatched by 6 percentage points. Middle school diversity also has a negative effect on the offered high school's value-added for white and Asian students and a positive, although non-significant effect on the offered high school's value-added of Black and Hispanic students.

In appendix Table B.1.9 we estimate models that allow for a more granular definition of race and distinguish between the effects of Black and Hispanic peers. These models have two endogenous regressors, one for each race share different from own in the middle school of enrollment (e.g. Black peer share and Hispanic peer share if the student is white or Asian). The two instruments are the two corresponding potential race shares in the offered middle school. Similarly to what observed for the binary other-race exposure treatment, attending middle schools with students from a particular ethnicity induces students to rank schools with a higher share of students from that ethnicity and a lower share of students from their own-race.

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

2.6.4 Peer Effects Mechanisms

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

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

We measure the effect of middle school diversity on students' achievement using the same 2SLS strategy used to study peer effects on high school choices. Table 2.8 reports middle school peer race effects on 6th, 7th and 8th grade Math and Ela standardized state test scores. We find no effect of other-race peers on the test scores of any students. These results are in line with several studies finding small to zero peer effects in achievement when using well-identified empirical designs (Angrist, 2013).

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

To investigate the effect of middle school diversity on racial preferences for high schools, we re-estimate respondents' preferences in the vignette experiment as a function of whether their student attended a majority white and Asian middle school. These estimates are reported in Table 2.9. Respondents whose students attended a majority-white and Asian middle schools on average prefer hypothetical high schools that are majority white and Asian over hypothetical schools that are racially neutral. This is in contrast with other Black and Hispanic respondents who tend to prefer racially-neutral schools and other white and Asian respondents who are indifferent between racially-neutral schools and majority white and Asian schools.

One way of interpreting these results is that interaction with other-race families in earlier grades reduces taste-based discrimination. An alternative hypothesis is that it reduces statistical discrimination, i.e. the extent to which household rely on race to make inference about the school academics. The second explanation is likely to play a larger role in this setting, given the smaller importance of pure taste-based discrimination found in the context of this experiment, as discussed in section 2.5.1.

An additional mechanism might be reduction in perceived discrimination, which we investigate in Table B.1.10. We only report the differential effect of having more middle school peers of different races on perceived discrimination by respondent race or ethnicity. We find mostly null effects, except for Asian students, the group with the highest levels of stated perceived discrimination. Attending majority Black and Hispanic middle schools makes Asian respondents more likely to agree with the statement "*My student would fit well in a school where the majority of peers are from a different race*" and less likely to report that their high school application choices were influenced by fear of discrimination.

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

Information In section 2.5.2 we showed that Black and Hispanic students were significantly less likely to have heard of high-value-added schools and schools enrolling higher achievers and a high share of white and Asian students. Here we study whether attending schools with white and Asian peers might close some of these gaps. We measure the effect of attending majority-white schools on the probability of knowing different types of schools by estimating the following regression:

$$Know_{is} = \alpha + Minority_i + \beta W_i + \gamma Minority_i \cdot W_i + X_i^{\prime}\delta + \epsilon_{is} \tag{2.5}$$

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

The OLS and IV estimates of β in Table 2.10 are non-significantly different from zero while estimates of γ are positive and significant, and somewhat larger when using IV. They indicate that while middle school peer demographics matter little for white and Asian families, attending majority white schools significantly expands the set of schools known by Black and Hispanic students. Black and Hispanic students are less likely to know schools with high value-added or higher achieving students, but attending schools with more informed peers entirely closes these gaps. Moreover, the effect of attending a white and Asian middle school on the total number of schools known for Black and Hispanic students appears larger than the race gap, suggesting that white and Asian peers appear to expand Black and Hispanic students awareness sets, rather than simply changing their composition.

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

2.7 Conclusions

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

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

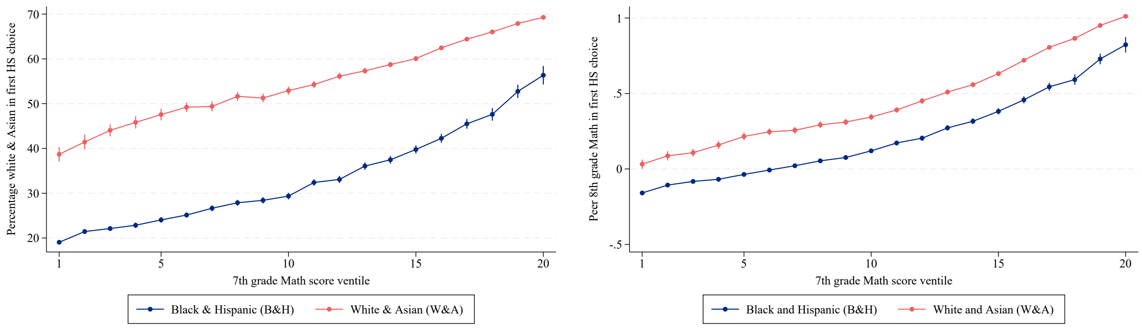
We also find large information frictions in the form of inaccurate beliefs about school attributes and about admission chances to high demand programs, but these are not differential by race. These results highlight that the interventions trying to correct biased beliefs, which has often been the focus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

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

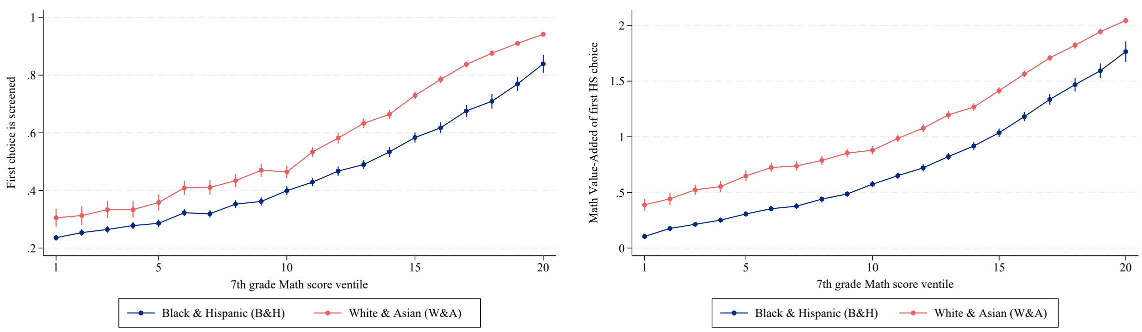
2.8 Figures

Figure 2-1: Differences in High School Choices by Race and Middle School Test Scores

(a) Differences in peer composition



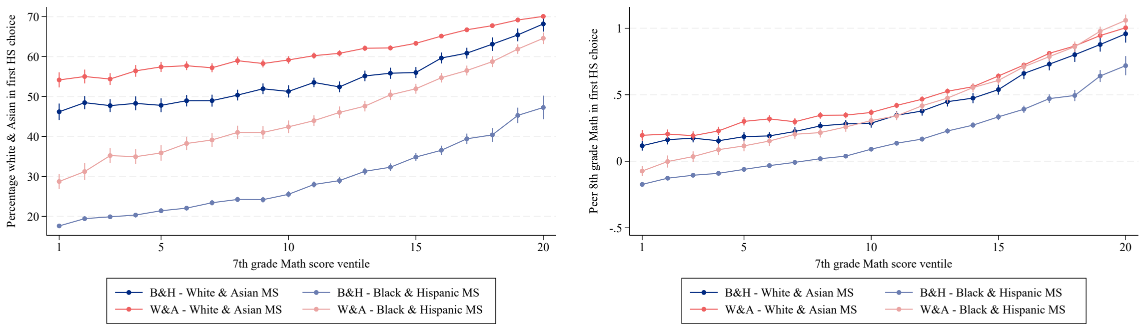
(b) Differences in school selectivity and quality



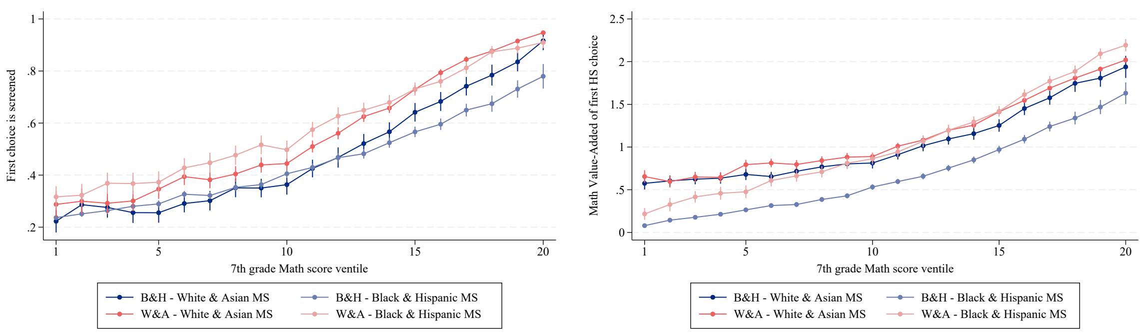
Notes: This figures plots mean characteristics of the school ranked first on the high school application by applicants' race and middle school test score ventiles. Panel (a) considers the percentage of white and Asian students and the mean 8th grade math scores of students at the school. Panel (b) considers the probability the school is screened and the math value-added (VA) of the school. School value-added is measured with school fixed effects in a standard OLS model that uses SAT Math as a dependent variable and controls for lagged student achievement.

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

(a) Differences in peer composition



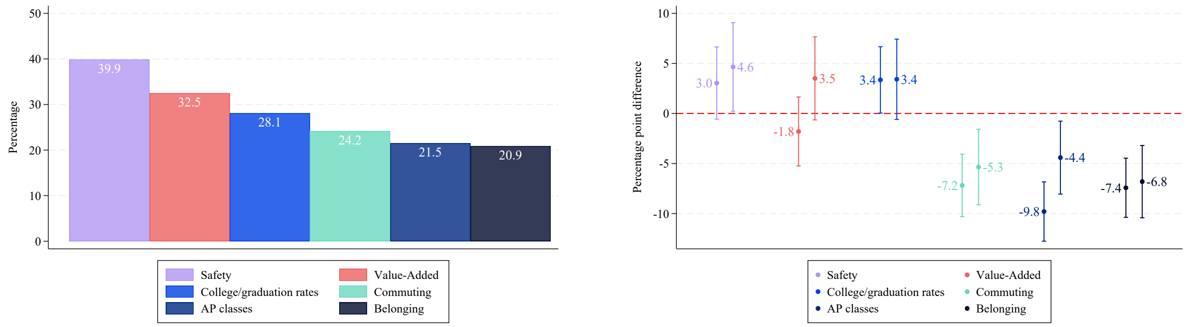
(b) Differences in school selectivity and quality



Notes: This figures plots mean characteristics of the school ranked first on the high school application by applicants' race and middle school test score ventiles and by the the racial composition of the middle school attended by the applicant. Characteristics of high school choices are depicted using a lighter shade for students enrolled in a majority-white and Asian middle school ($\geq 50\%$ of white and Asian enrollment) and in a darker shade for students enrolled in a majority-Black and Hispanic Middle School. The characteristics considered are the same as in Figure 2-1.

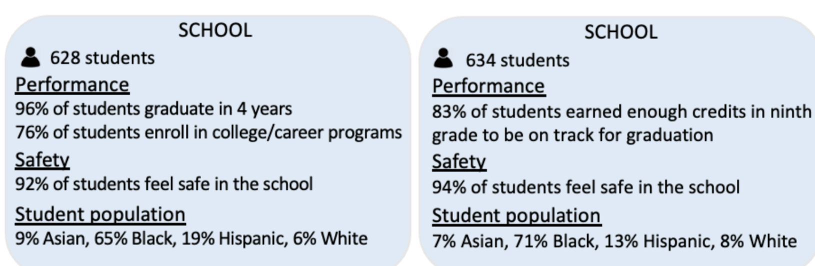
Figure 2-3: Differences in Most Important School Features

(a) Means (b) Racial gaps



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

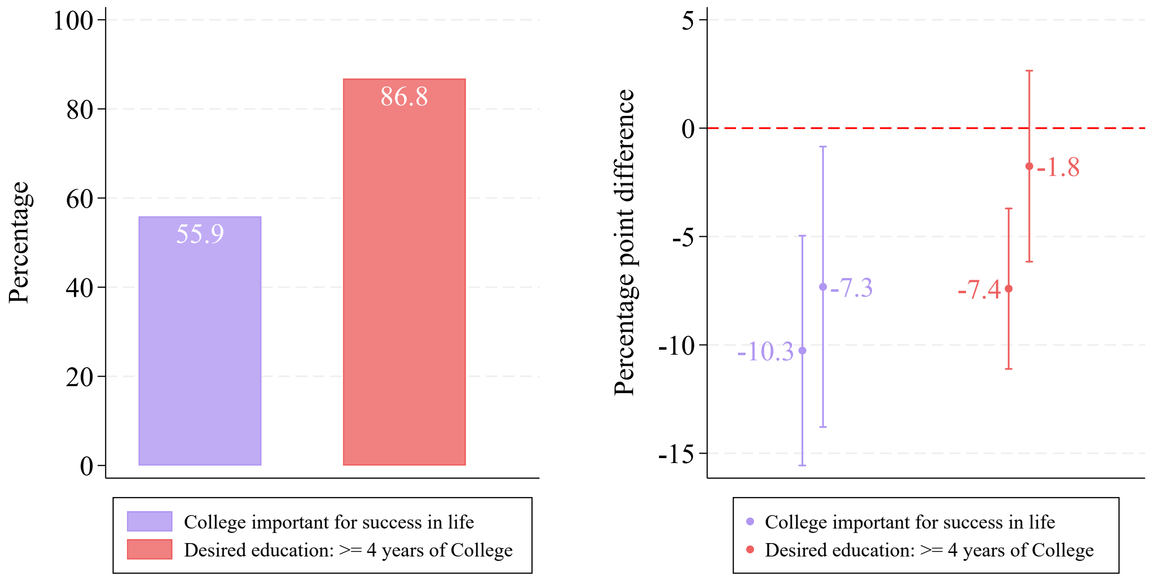
Figure 2-4: School Cards for Vignette Experiment



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

Figure 2-5: Differences in Aspirations

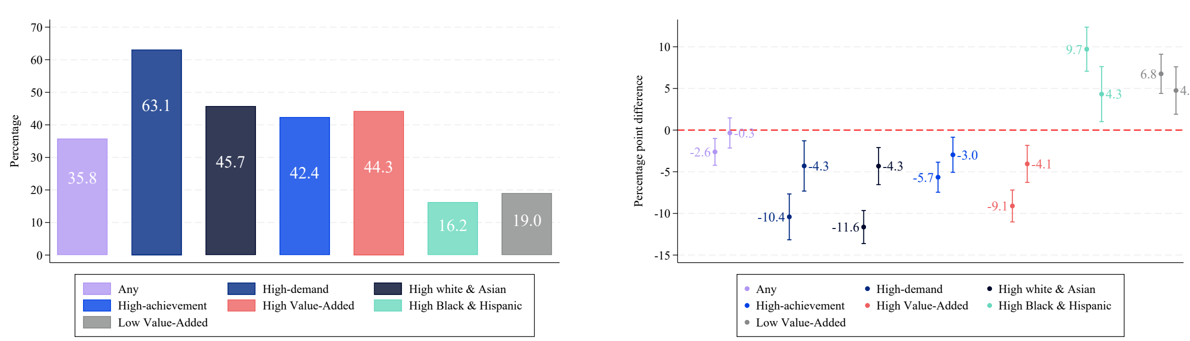
(a) Means (b) Racial gaps



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

Figure 2-6: Differences in Awareness Sets

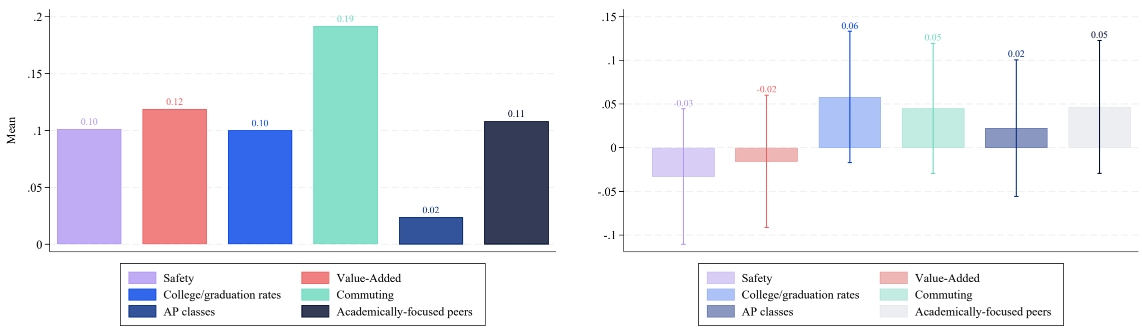
(a) Means (b) Racial gaps



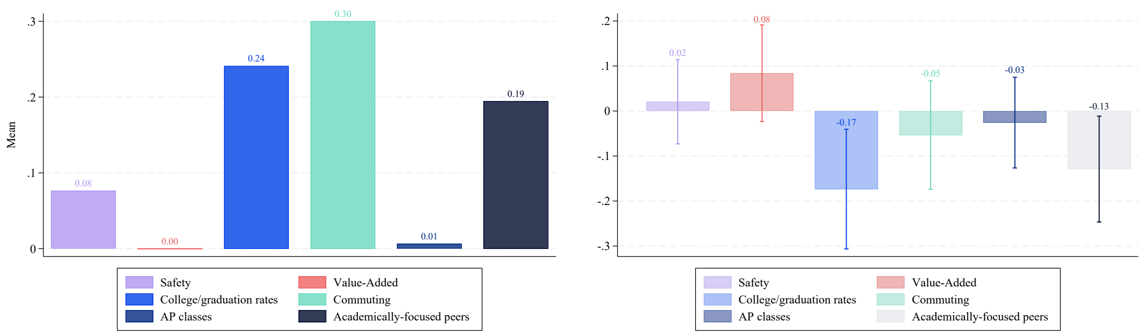
Notes: This figure reports differences in respondents' awareness sets. Panel (a) reports the mean share of schools respondents were aware of by school type. Panel (b) reports the differences in the share of schools Black and Hispanic respondents were aware of compared to white and Asian respondents.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

Figure 2-7: Differences in Information About School Characteristics

(a) Excess p(correct) - pairwise comparisons



(b) Correlation between real quartile and answer



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

2.9 Tables

Table 2.1: Summary Statistics

	Administrative data sample				Survey respondents		
	MS applicants	Experimental sample					
	applying to HS	All	Black+ Hispanic	White+ Asian	All	Black+ Hispanic	White+ Asian
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Demographics and baseline scores							
Black	0.20	0.22	0.36	0.00	0.14	0.29	0.00
Hispanic	0.39	0.40	0.64	0.00	0.33	0.69	0.00
White	0.17	0.14	0.00	0.38	0.21	0.00	0.40
Asian	0.21	0.22	0.00	0.62	0.29	0.00	0.56
FRPL	0.72	0.75	0.84	0.62	0.66	0.80	0.53
Female	0.52	0.51	0.53	0.50	0.52	0.52	0.51
Ell	0.06	0.06	0.07	0.04	0.06	0.08	0.04
4th gr. Math	0.24	0.20	-0.12	0.76	0.51	0.11	0.88
4th gr. Ela	0.24	0.21	-0.04	0.61	0.46	0.15	0.75
Panel B: Middle school characteristics							
% Black+Hispanic peer in MS	61	63	79	37	54	72	37
Panel C: High school choices							
Number of HS choices	8.2	8.0	8.2	7.6	8.9	8.9	9.0
% Black+Hispanic in top3 choices	57	59	71	40	49	62	38
Mean baseline peer math in top3 choices	0.24	0.20	0.04	0.48	0.40	0.19	0.59
Mean popularity in top3 choices	4.15	4.13	3.55	5.10	5.71	4.16	7.16
Lists a screened program among top3 choices	0.71	0.69	0.60	0.85	0.78	0.67	0.88
Mean RC SAT math VA in top3 choices	0.80	0.74	0.45	1.23	1.04	0.66	1.40
N	211,146	69,336	43,043	25,040	3,628	1,749	1,879

Notes: The administrative data sample in columns 1 to 4 includes students who applied to middle school for enrollment in 2015-2016 to 2019-2020 and then successively applied to high school for enrollment in 2018-2019 to 2022-2023. Columns 1 reports descriptive statistics for the sample of applicants who have demographic information. Columns 2 to 4 restrict the sample to the experimental sample which includes offered Middle school applicants who have (i) non-degenerate risk of school assignment, (ii) non-missing baseline test scores, and (iii) non-missing geographic information. The survey respondents in column 5 to 7 include any survey participants who answered at least one survey question. The baseline scores are 4th grade scores from the NY state standardized assessments. High school popularity corresponds to the number of applicants rejected by the program divided by the number of accepted applicants (city-mean is 1.37). Screened programs are programs that admit students based on their Middle school grades and/or auditions and essays. The risk-controlled value-added computation (RC SAT math VA) follows that in Angrist et al. (2021).

Table 2.2: Different School Characteristics for the Vignette Experiment

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

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

Table 2.3: Vignette Experiment Preference Estimates

Respondent race	Characteristic	Precise Info (1)	Imprecise Info (2)
White+Asian	Constant	2.17***	3.03***
		(0.08)	(0.10)
	High-academics	1.44***	
		(0.06)	
	High-safety	0.74***	1.09***
		(0.06)	(0.08)
	Majority Black	-0.44***	-0.70***
		(0.08)	(0.11)
	Majority Hispanic	-0.28***	-0.59**
		(0.08)	(0.11)
	Majority white+Asian	0.28***	0.36***
		(0.08)	(0.12)
Black+Hispanic	Constant	2.72***	3.41***
		(0.09)	(0.10)
	High-academics	1.28***	
		(0.07)	
	High-safety	0.66***	1.16***
		(0.07)	(0.09)
	Majority Black	-0.11	-0.29**
		(0.09)	(0.12)
	Majority Hispanic	0.01	-0.16
		(0.09)	(0.12)
	Majority white+Asian	-0.08	-0.22*
		(0.09)	(0.12)
N respondents		1,212	957

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

Table 2.4: Beliefs About Admission Probability

Panel A: Beliefs about admission probabilities

	Applied to "dream school"	Admission beliefs
	(1)	(2)
Actual admission probability	0.116***	0.164***
	(0.040)	(0.023)
(Actual admission probability) × (White+Asian)	0.031	0.013
	(0.044)	(0.026)
White+Asian	-0.012	-0.035
	(0.045)	(0.027)
Mean	0.169	0.612
N	2,363	3,297

Panel B: Beliefs about performance tercile

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

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

Table 2.5: 2SLS Estimates of Peer Share Effects on Black & Hispanic Students' HS Choices

	% Black	% Hispanic	% White	Peer Ela	Peer Math	Popularity	Screened	SAT math VA	Length of rol
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
					Top 3 choices				
Majority white+Asian MS	-3.523***	-2.100**	5.886***	0.086***	0.147***	0.289	-0.005	0.149**	0.007
	(1.075)	(1.115)	(1.507)	(0.029)	(0.033)	(0.246)	(0.038)	(0.053)	(0.398)
Share white+Asian (10pp)	-0.835***	-0.283	1.131***	0.017***	0.023***	0.104**	0.009	0.021**	-0.021
	(0.216)	(0.209)	(0.273)	(0.005)	(0.006)	(0.054)	(0.007)	(0.010)	(0.070)
mean	24.70	45.85	27.70	-0.02	0.04	3.55	0.60	0.37	8.23
					All choices				
Majority white+Asian MS	-2.580***	-1.799**	4.675***	0.069***	0.120***	0.208	0.004	0.116***	
	(0.935)	(0.877)	(1.168)	(0.024)	(0.029)	(0.214)	(0.030)	(0.041)	
Share white+Asian (10pp)	-0.558***	-0.286*	0.861***	0.014***	0.018***	0.069*	0.006	0.020**	
	(0.186)	(0.165)	(0.213)	(0.004)	(0.005)	(0.047)	(0.006)	(0.009)	
mean	25.51	46.75	26.03	-0.08	-0.02	3.18	0.78	0.28	
N	43,042	43,042	43,042	43,042	43,042	42,731	43,042	43,034	

Notes: This table reports 2SLS estimates of middle school demographic composition effects on Black and Hispanic high school choices. Panel A focuses on each applicant's top 3 choices, panel B includes all the choices. The sample includes students with non-degenerate risk of middle school assignment, who applied to Middle schools for enrollment in 2015-2016 to 2019-2020 and then successively applied to high school for enrollment in 2018-2019 to 2022-2023. All models control for application year, student demographic characteristics (ELL status, gender, poverty status, district of residence), and 4th grade math and ELA test scores. High school popularity, screened status and RC VA are defined in the notes of Table 2.1. Standard errors clustered at the Middle school × year level in parenthesis.

Table 2.6: 2SLS Estimates of Peer Share Effects on White & Asian Students' HS Choices

	% Black	% Hispanic	% White	Peer Ela	Peer Math	Popularity	Screened	SAT math VA	Length of rol
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
					Top 3 choices				
Majority Black+Hispanic MS	0.500	1.466	-2.438**	-0.085**	-0.114***	0.350	-0.003	-0.171**	-1.281***
	(0.931)	(1.290)	(1.486)	(0.036)	(0.045)	(0.247)	(0.044)	(0.079)	(0.472)
Share Black+Hispanic (10pp)	0.093	0.534**	-0.686**	-0.015**	-0.021**	0.049	0.002	-0.029**	-0.237***
	(0.186)	(0.225)	(0.289)	(0.007)	(0.008)	(0.053)	(0.008)	(0.014)	(0.090)
mean	13.44	26.48	58.10	0.38	0.48	5.10	0.85	1.01	7.58
					All choices				
Majority Black+Hispanic MS	0.711	1.949**	-3.019**	-0.069**	-0.094***	0.196	-0.046*	-0.155***	
	(0.852)	(1.086)	(1.295)	(0.030)	(0.037)	(0.215)	(0.038)	(0.063)	
Share Black+Hispanic (10pp)	0.210	0.522***	-0.764***	-0.010**	-0.016**	0.017	-0.004	-0.027**	
	(0.165)	(0.194)	(0.253)	(0.006)	(0.007)	(0.043)	(0.006)	(0.011)	
mean	14.20	29.52	54.37	0.24	0.35	4.50	0.91	0.81	
N	25,040	25,040	25,040	25,040	25,040	25,014	25,040	25,022	

Notes: This table reports 2SLS estimates of middle school demographic composition effects on white and Asian high school choices. Panel A focuses on each applicant's top 3 choices, panel B includes all the choices. The sample, controls and endogeneous variables are as defined in the notes of Table 2.5. Standard errors clustered at the Middle school × year level in parenthesis.

Table 2.7: 2SLS Estimates of Peer Share Effects on Characteristics of Offered High School

	Matched in 1st round	Offered rank	% Black	% Hispanic	% White	Peer Ela	Peer Math	Popularity	Screened	SAT math VA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
					Panel A: Black & Hispanic students					
Majority white+Asian MS	-0.060**	0.112	-3.783**	-1.407	5.558**	0.080**	0.125**	-0.138	-0.025	0.072
	(0.024)	(0.207)	(1.540)	(1.342)	(1.756)	(0.035)	(0.038)	(0.239)	(0.042)	(0.060)
Share white+Asian (10pp)	-0.012**	-0.022	-0.693**	-0.326	1.022***	0.012**	0.014**	0.086**	0.005	0.017
	(0.004)	(0.043)	(0.289)	(0.263)	(0.278)	(0.006)	(0.007)	(0.042)	(0.007)	(0.012)
mean	0.96	2.41	29.20	49.10	20.07	-0.21	-0.17	1.35	0.22	0.05
N	43,037	34,530	41,117	41,117	41,117	41,093	41,093	41,128	41,587	41,060
					Panel B: White & Asian students					
Majority Black+Hispanic MS	0.040	-0.757**	0.087	1.199	-1.876	-0.019	-0.085	0.871**	0.084	-0.132
	(0.026)	(0.268)	(1.495)	(1.890)	(2.139)	(0.051)	(0.058)	(0.353)	(0.064)	(0.098)
Share Black+Hispanic (10pp)	0.003	-0.144**	0.287	0.670*	-1.000**	-0.000	-0.012	0.160**	0.028**	-0.019
	(0.005)	(0.059)	(0.301)	(0.363)	(0.446)	(0.010)	(0.011)	(0.072)	(0.012)	(0.018)
mean	0.93	2.87	16.15	31.12	50.85	0.17	0.27	3.00	0.47	0.66
N	25,033	19,743	23,310	23,310	23,310	23,309	23,309	23,316	23,614	23,272

Notes: This table reports 2SLS estimates of middle school demographic composition effects on high school offers. Panel A focuses on Black and Hispanic applicants, while panel B focuses on white and Asian applicants. The sample, controls and endogeneous variables are as defined in the notes of Table 2.5. Standard errors clustered at the Middle school × year level in parenthesis.

Table 2.8: 2SLS Estimates of Peer Share Effects on Test Scores

	Has 6th grade test	Has 7th grade test	Has 8th grade test	6th grade Ela	6th grade Math	7th grade Ela	7th grade Math	8th grade Ela	8th grade Math
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				Panel A: Black & Hispanic students					
Majority white+Asian MS	-0.009	0.003	-0.012	0.001	0.046	0.052	0.077	-0.029	0.073
	(0.008)	(0.014)	(0.036)	(0.041)	(0.047)	(0.051)	(0.053)	(0.092)	(0.071)
Share white+Asian (10pp)	-0.003*	-0.001	-0.004	-0.000	-0.002	-0.005	0.013	-0.021	-0.006
	(0.002)	(0.003)	(0.007)	(0.009)	(0.010)	(0.011)	(0.012)	(0.015)	(0.019)
mean	0.99	0.97	0.92	-0.02	-0.12	-0.00	-0.10	0.03	-0.00
N	36,328	29,227	20,251	35,891	35,743	28,175	27,847	18,542	14,635
				Panel B: White & Asian students					
Majority Black+Hispanic MS	-0.006	0.002	-0.009	0.038	-0.036	-0.056	0.010	0.089	-0.067
	(0.013)	(0.019)	(0.038)	(0.066)	(0.062)	(0.072)	(0.060)	(0.111)	(0.149)
Share Black+Hispanic (10pp)	0.002	-0.000	-0.004	0.022*	-0.007	-0.016	0.001	0.020	0.027
	(0.003)	(0.004)	(0.008)	(0.013)	(0.014)	(0.015)	(0.015)	(0.026)	(0.034)
mean	0.99	0.99	0.95	0.71	0.81	0.73	0.83	0.72	0.83
N	21,112	16,256	10,951	20,896	20,813	15,988	15,854	10,336	6,852

Notes: This table reports 2SLS estimates of middle school demographic composition effects on 6th, 7th and 8th grade State standardized test scores. Panel A focuses on Black and Hispanic applicants, while panel B focuses on white and Asian applicants. The sample and controls variables are as defined in the notes of Table 2.5. Standard errors clustered at the Middle school × year level in parenthesis.

Table 2.9: Vignette Experiment Preference Estimates by MS

	Black & Hispanic	White & Asian
	(1)	(2)
Constant	2.76***	2.23***
	(0.09)	(0.13)
Majority white+Asian MS	-0.66***	-0.07
	(0.20)	(0.15)
Imprecise info	0.81***	0.99***
	(0.09)	(0.12)
(Imprecise info) × (Majority white+Asian MS)	0.21	-0.13
	(0.19)	(0.15)
High-academics	1.26***	1.58***
	(0.08)	(0.11)
(High-academics) × (Majority white+Asian MS)	0.36**	-0.24*
	(0.18)	(0.13)
(High-safety)	0.85***	0.94***
	(0.06)	(0.09)
(High-safety) × (Majority white+Asian MS)	0.11	-0.15
	(0.14)	(0.10)
Majority Black	-0.22**	-0.54***
	(0.09)	(0.13)
(Majority Black) × (Majority white+Asian MS)	0.22	0.03
	(0.20)	(0.15)
Majority Hispanic	-0.13	-0.46***
	(0.09)	(0.12)
(Majority Hispanic) × (Majority white+Asian MS)	0.37*	0.11
	(0.20)	(0.14)
Majority white+Asian	-0.24***	0.05
	(0.09)	(0.12)
(Majority white+Asian) × (Majority white+Asian MS)	0.50***	0.33**
	(0.19)	(0.15)
N respondents	914	1,086

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

Table 2.10: Peer Effects on Consideration Sets

	Any school		Popular		High white+Asian %		High bl. Math		High VA	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(Black+Hispanic) × (High white+Asian MS)	0.08***	0.20**	0.08**	0.07	0.10***	0.22**	0.11***	0.24**	0.11***	0.23**
	(0.02)	(0.09)	(0.03)	(0.15)	(0.02)	(0.11)	(0.02)	(0.10)	(0.02)	(0.11)
High white+Asian MS	-0.03**	-0.10	0.02	-0.00	-0.03*	-0.12	-0.02	-0.13*	-0.02	-0.12
	(0.01)	(0.06)	(0.02)	(0.11)	(0.02)	(0.08)	(0.01)	(0.07)	(0.02)	(0.08)
Black+Hispanic	-0.04***	-0.04***	-0.06***	-0.06**	-0.08***	-0.09***	-0.07***	-0.07***	-0.08***	-0.08***
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
N	25,690	25,690	5,138	5,138	13,801	13,801	15,683	15,683	12,994	12,994
mean white+Asian	0.380	0.380	0.680	0.680	0.500	0.500	0.450	0.450	0.470	0.470

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

Chapter 3

Collective Bargaining for Women: How Unions Can Create Female-Friendly Jobs

Written jointly with Lorenzo Lagos and Garima Sharma

3.1 Introduction[1]

Despite significant labor market progress over the past decades, women continue to disproportionately suffer large earnings losses because they are in-charge at home (Kleven et al., 2019). Across 142 countries, over 30% of working women cite having to balance family and work as their main challenge (ILO and Gallup Inc., 2017). While governments and scholars alike have argued that making workplaces more female-friendly is key to lowering gender disparities—for example, Goldin (2014) argues that changing the structure of jobs may cause all remaining gender earnings gaps to vanish—little is known about if and how labor market institutions can be redesigned from within to ameliorate the stark trade-offs faced by working women.

Per one view, making workplaces female-friendly—providing maternity leave, childcare, and flexible work schedules—is not worth the expense to employers since the marginal worker does not value them enough. This paper tests an alternate view: that, instead of the marginal

[1]We thank Daron Acemoglu, David Atkin, David Autor, Abhijit Banerjee, Esther Duflo, Brandon Enriquez, Henry Farber, and Simon Jäger, as well as seminar participants at MIT, Brown, Princeton, NBER Labor Studies Summer Institute, SOLE, LERA, PUC-Rio, FGV-EESP, Insper, Central European University, and Norwegian Business School for helpful advice and comments. We are grateful to João Fernandes, Roberto Hsu Rocha, Stanley Gacek, Juvandia Moreira, Juliana da Penha Thomaz, and Beatriz Santos for their incredible help in better understanding the context, and Iacopo Morchio and Christian Moser for their help with the PageRank codes. Access to Brazil's RAIS database is governed by the Data Use Agreement between MIT and Brazil's former Ministry of Labor. We thank David Atkin and Mayara Felix for procuring MIT's access to the database, and Mayara Felix for de-identifying, harmonizing, and translating the RAIS datasets pursuant to MIT COUHES guidelines.

worker's preference, the priorities of those designing compensation determine workplace amenities. Because a few individuals typically decide workplace policies, their priorities take precedence and may not always feature women's needs on top. When these priorities change, so too do workplaces. Unions provide a natural setting in which to test this hypothesis since, for nearly 20% of the world's workers, a few union representatives negotiate pay and benefits (Visser, 2019). Since few union leaders are women, they may not represent women's interests in collective bargaining.[2]

This paper investigates how changing leaders' priorities in women's favor changes the workplace. The ideal experiment to study this question requires a top-down change in priorities that is uncorrelated with changes to a firm's labor demand or workers' preferences. We exploit such a natural experiment in Brazil, that spurred leaders of its largest trade union federation (or "union central"), the *Central Única dos Trabalhadores* (CUT), to prioritize women's needs in collective bargaining.[3] Starting in 2015, the CUT reserved half its leadership positions for women and emphasized the provision of female-focused policies, such as 6 months of paid maternity leave, flexible work schedules, and childcare. Because unions seldom change affiliation to a union central, and neither workers nor establishments choose their union, the CUT reform represents a top-down pro-women directive to union leaders that is unrelated to an establishment's labor demand or supply. This motivates using a difference-in-differences design to compare amenities and costs (wages, employment) at establishments negotiating with CUT unions (treated) to non-CUT affiliates (comparison). The two sets of establishments closely resembled each other at baseline; together they comprise 19% of formal employment in Brazil, and employ 11.5 million workers across 80,000 establishments.

Unique to the Brazilian setting, our analysis relies on linking three rich sources of data: (i) establishment-level amenities from the text of all collective bargaining agreements (CBAs), (ii) worker outcomes from linked employer-employee records covering all formal employment (RAIS), and (iii) union affiliation and leadership covering all unions. CBAs offer uniquely high quality information on 137 different amenities offered by establishments, including maternity leave, workplace safety, absences, and work hours. The administrative data track workers over time and report their gender, wages, and instances of maternity leave.

We begin by using a revealed preference approach to identify which amenities are highly valued by women and which by men, relying on the idea that workers flock to employers with

[2]For example, nearly half of all workers but only 12% of union leaders in Brazil are women. In continental Europe, where collective bargaining covers a majority of workers, including Germany, Austria, and the Netherlands, less than 30% of union members are women (Skorge and Rasmussen, 2022).

[3]Union centrals are umbrella organizations that coordinate priorities among local unions. Over half of all formal workers in Brazil are covered by collective bargaining and 20% of unions affiliate with CUT.

better work conditions. Employer-to-employer moves thus reveal valuable firms (Sorkin, 2018; Morchio and Moser, 2020), and correlating these values with CBA clauses reveals valuable amenities. We find that women value amenities enabling work-life balance, including maternity protections, childcare payments, absences, and workday reductions ("female-centric" amenities). In contrast, men value higher pay and safety, such as clauses governing profit sharing, hazard pay, life-insurance, and safety equipment ("male-centric" amenities).[4] In an out-of-sample sense check, we find that female amenities increase—and male amenities decrease—with the share of women in an establishment's workforce, providing the first clue that representation could influence amenities.

The second part of our analysis studies the causal effect of shifting union leaders' priorities on female and male-centric amenities, and its downstream effect on workers and establishments, on their wages, retention, and employment.

Our first main takeaway is that female-centric amenities increase on paper and in practice. On paper, we find a 19% increase in female-centric amenities. This is a large improvement, equivalent to moving from the average baseline amenity count at a minority female establishment to one where over 80% of workers were women. Provisions governing leaves and childcare account for much of the gain, suggesting that the reform spurs benefits especially for childbearing women. The largest gains occur at establishments where women had limited voice at baseline, either by being in the minority among workers or among union leaders.

Amenity improvements on-paper translate into practice. Following the reform, women at treated establishments take longer maternity leaves, enjoy job protection following these leaves, and constitute a larger share of managers.

Our second main takeaway is that women value changes to the work environment induced by the CUT reform, ruling out a pure compensating differences story for the amenity gains. Specifically, we find women separating less from and queuing for jobs at treated establishments, both of which are revealed preference measures of firm value (Krueger and Summers, 1988; Holzer et al., 1991). Retention among women increases by 10% and reflects a decline in voluntary separations. While we do not directly observe job queues, we proxy for them using probationary contracts that are commonly used by employers to screen applicants. Women's share among probationary workers rises by 10%. In sum, higher female-centric amenities cause women to flock to CUT-affiliated establishments.

[4] We mitigate simultaneity bias, i.e., that employers increase female-centric amenities when wanting to hire more women, by using amenities from sectoral agreements negotiated with multiple employers in an industry instead of firm-level agreements negotiated with a single employer. Unlike the latter, sectoral CBAs are unlikely to be influenced by employer-level demand shocks.

Our third main takeaway is that these improvements for women manifest without observed tradeoffs in wages or employment. Compensating differences would suggest that women's wages fall to finance amenity improvements (Rosen, 1986). However, we find no effect on the earnings of either new or incumbent workers, ruling out even very small declines with a high degree of confidence. Given no wage change, establishments may reduce women's employment because they are now more expensive to employ. We find no evidence of this; employment remains unchanged. Instead, women comprise a larger share of the workforce (by 0.2pp relative to 36% at baseline).

If women are not losing, perhaps men are. However, there is little evidence of this. We find no decline in the earnings or employment of incumbent male workers. Male amenities do not decline. If anything, there is a small positive treatment effect on retention among incumbent male workers, suggesting that men *value* the changes to the work environment spurred by the CUT reform. Overall, our findings are consistent with a model of the labor market wherein firms post utility offers for each gender (e.g. Card et al. (2018); Berger et al. (2022)). The reform causes this posted utility for women to rise without a corresponding decline in men's utility.

If workers do not finance amenity improvements, perhaps firms finance them through lower profits. Both the empirical evidence and theoretical reasons point against this explanation. Empirically, we find no treatment effect on establishment exit—which is a non-trivial margin of adjustment in Brazil, with 8.7% of control establishments having exited within two years of the reform. For the subsample of establishments that report to Orbis, we find no evidence of a decline in measured profits. Theoretically, the CUT reform shifted union priorities rather than raising unions' bargaining power. As such, CUT unions were not positioned to capture a larger share of surplus and thereby reduce profits. Indeed, while increasing union bargaining power generally predicts changes in employment, we find a precisely estimated zero effect.

How, then, are these amenity gains paid for? One explanation is that the union shifts rents from men to women (albeit not observed in the data as wage or amenity declines), but men are not marginal to these rents as they would not obtain them elsewhere. A second possibility is that bargaining was inefficient ex-ante and changing union priorities led to a pareto improvement for workers and firms. At least two general models could explain these results. In one, frictions in the bargaining process or in aggregating workers' interests to the union level (e.g., information or contracting frictions) yield the possibility for win-win situations once union attention is refocused on previously ignored issues. In another model, behavioral firms and unions did not conceive of providing female-centric amenities until changing union priorities put these issues front-and-center.

The final part of our analysis develops a revealed preference method to quantify the welfare effect of changing the work environment by drawing an analogy with consumer theory (Feenstra, 1994; Redding and Weinstein, 2016). Just as gains to consumer welfare from improving product varieties are quantifiable via an increase in the expenditure share on these improving varieties, gains to worker welfare from improving workplace amenities are quantifiable via an increase in the wage bill at these amenity-improving employers. In other words, workers vote with their feet toward desirable employers. A few sufficient statistics then quantify gains in welfare.[5] This sufficient statistics approach allows us to remain agnostic regarding the precise functional form linking amenities to utility. Consistent with our reduced-form findings, we find that the CUT reform raises women's welfare by 6% while leaving men's welfare unchanged.

This paper contributes to four literatures. First, on unions and inequality. While firms care about the marginal worker, it is unclear who the union cares about (Farber, 1986). Unions have long struggled to organize workers with competing interests (Hill, 1996) and unionization has mixed effects for different worker groups, raising wages for low skill workers (Card, 1996; Farber et al., 2021) and black workers (Ashenfelter, 1972), but not necessarily women (DiNardo et al., 1996; Card et al., 2004, 2020; Bolotnyy and Emanuel, 2022). We provide quasi-experimental evidence that union leaders' priorities determine whose interests they represent. When unions prioritize women, they can lower within-firm gender inequality.

Second, on the importance of leaders' priorities in how institutions function. Political leaders are found to better represent their own group's preferences than the average constituent's (Chattopadhyay and Duflo, 2004; Pande and Ford, 2012). In the labor market, women negotiate less over pay than men (Dittrich et al., 2014; Leibbrandt and List, 2015; Biasi and Sarsons, 2022), suggesting that, here too, leaders could step in on their behalf. While women on company boards have been found to have limited effects on gender gaps (Bertrand et al., 2018; Flabbi et al., 2019; Matsa and Miller, 2011; Maida and Weber, 2020), we find an important role for union leaders. Just as in politics, top-down changes to union leaders' priorities alter the workplace, in this case making it better for women.

Third on whether providing female-focused amenities leads employers to lower women's

[5]For tractability, we assume that workers possess nested CES preferences over employers. Just as gains to consumer welfare from improving product varieties are quantifiable via changes to the price index, i.e., change in the cost of purchasing an additional util of utility, gains to worker welfare from improving workplace amenities are quantifiable using changes to a wage index, i.e., change in the wage for working one disutility-weighted hour. Under CES, only four sufficient statistics quantify these gains: an increase in the share of labor income at treated establishments (capturing workers flocking to amenity-improving employers), workers' elasticity of substitution across establishments (capturing how difficult these moves were), change in the dispersion of labor income at comparison establishments (capturing where workers are drawn away from), and any change in the wage at comparison establishments (capturing pro-competitive responses).

wages (Gruber, 1994) or reduce hiring when they cannot (Summers, 1989). We find no evidence of this: although the work environment improves for women, we cannot reject the null that their wages and employment do not suffer (and, indeed, rule out very small declines with high confidence). By way of benchmark, Lagos (2021) estimates that leave clauses—many of which emerge as female-focused in our revealed preference approach—are valued at 8.4% of a worker's wage. Instead, although recent work demonstrates limited gains for workers from greater voice on corporate boards (Harju et al., 2021; Blandhol et al., 2020), we find substantial gains from elevating women's voices on union boards. One exception is Boudreau (2023), who finds that elevating worker voice through Occupational Safety and Health (OSH) committees in Bangladeshi garment factories has a small, positive effect on workplace safety without detectable impacts on wages and employment.

Finally, our paper contributes to the revealed preference literature in three ways. First, we provide quasi-experimental evidence that workers move toward improving amenities, consistent with several papers that infer amenity values using such moves (Krueger and Summers, 1988; Sorkin, 2018; Taber and Vejlin, 2020; Morchio and Moser, 2020; Lagos, 2021; Lamadon et al., 2022). Second, we use worker moves and variation in amenities across establishments to identify what workers value, using a richer set of amenities and higher stakes environment than possible in experiments. Encouragingly, our results match this experimental work—in particular, women value flexibility (Mas and Pallais, 2017; Wiswall and Zafar, 2017; Maestas et al., 2018). Finally, we bring a revealed preference approach from consumer theory to the labor setting to quantify the welfare effect of changes to a firm's work environment. This sufficient statistics approach remains agnostic regarding the precise functional form linking a rich set of amenities to utility; it can be similarly employed in future work to quantify the welfare effect of a change to the work environment.

The rest of the paper proceeds as follows. Section 3.2 discusses the institutional context and CUT reform. Section 3.3 describes the data and details our approach for classifying amenities as female- or male-centric. Section 3.4 presents our empirical strategy. Section 3.5 presents our main results on the causal effect of changing union priorities on amenities, revealed preference measures of firm value, and labor market outcomes. Section 3.6 quantifies the welfare impact of improving (female-centric) amenities on men and women. Section 3.7 concludes.

3.2 Institutional Context

We begin by describing the collective bargaining structure in Brazil, emphasizing the distinction between unions (which represent workers in collective bargaining) and union centrals (which coordinate activities among affiliated unions). We then describe the 2015 pro-women reform enacted by Brazil's largest union central (the CUT), which provides the top-down shift in priorities at affiliated unions that we use for identification.

3.2.1 Collective Bargaining and Union Centrals

Types of CBAs Brazil has two types of collective bargaining agreements (CBAs): sectoral and firm-level. In sectoral CBAs, unions negotiate with employer associations representing establishments in a specific industry and geography, for example, the car manufacturers in Curitiba. In firm-level CBAs, unions negotiate with individual employers, for example, Volkswagen. Given their wider coverage, sectoral agreements typically set general floors for wage and non-wage benefits. By contrast, firm-level agreements generally build on these floors to expand benefits for workers at individual employers (Horn, 2009). Our main analysis studies the impact of the CUT reform on firm-level CBAs. However, we leverage amenities contained in sectoral CBAs to identify the clauses that are highly valued by female and male workers (Section 3.3.2).

Union determination The union that negotiates CBAs on behalf of workers at any given employer is chosen neither by the workers nor by the employer. Rather, representation depends on two factors: 1) industry (or category); and 2) geographic location (municipality).[6] Examples of unions include the bank workers' union of São Paulo and the teachers' union of Florianopolis.

Neither workers nor employers can change their union. As a legacy of Brazil's corporatist past, the first union approved to represent a given category-geography cell enjoys a lifetime monopoly.[7] As such, workers can only influence their union priorities from within, for example, by voting in union elections, running for union leadership, or voicing their concerns to union leadership. At the same time, employers cannot avoid unions by virtue of this predetermined assignment of the same union to all employers in a category-geography cell. Naturally, union assignment by these cells produces an incredibly fragmented union landscape, with over 11

[6] For a few professions, the worker's occupation rather than the industry determines representation in collective bargaining, e.g., for elevator operators, journalists, and musicians. These cross-industry, occupation-based unions comprise approximately 5% of all unions in Brazil.

[7] President Getúlio Vargas instituted this "monopoly union" framework, known as *unicidade sindical*, in the late 1930s as a means to co-opt the labor movement by enabling the federal government to control the union given the right to represent workers in collective bargaining.

thousand labor unions operating in Brazil.[8]

CBA coverage Neither workers not employers cannot opt out of CBAs negotiated by their union. Coverage is universal, which means that workers need not be union members to enjoy negotiated benefits.[9] Consequently, union membership is low (at around 20%) consisting of workers willing to pay membership dues in exchange for additional benefits that are not in CBAs, e.g., recreational facilities and private health insurance plans. Importantly, individual work contracts cannot take away benefits negotiated in CBAs, meaning that CBA provisions constitute a general floor for all represented workers. Similarly, CBAs cannot derogate provisions granted by the federal labor code. CBA clauses therefore build on top of these basic guarantees that are enjoyed by all workers.

Negotiation process Unions' priorities play a central role in determining the content of CBA negotiations. Before the expiration of an existing CBA, the union organizes a General Assembly where workers vote on the list of demands (or *pauta de reivindicações*) that they want to prioritize in the next negotiation. Union leaders typically select the topics that are discussed at these assemblies and are up for vote into the *pauta*. Negotiations officially begin when the union sends these demands to employers. They occur over several rounds. Most CBAs are signed for a duration of 12 months, giving rise to annual negotiations.[10] The union board also decides which representatives sit at the bargaining table, which is not restricted exclusively to board members.

Union centrals Unions can affiliate with union centrals (or *centrais sindicais*), which are somewhat analogous to trade union federations such as the AFL-CIO in the United States. These centrals are national level, umbrella organizations that coordinate the activities of local unions and lobby for political favor (Liukkunen, 2019). While union centrals do not directly participate in collective bargaining, they are indirectly involved in coordinating union priorities *across* worker categories. For example, union centrals regularly organize general strikes, plan annual conferences attended by union representatives, provide support to local unions, participate in public discussion forums on behalf of constituent unions, and steer union attention toward broad priorities such as gender or racial equality.

[8]It's worth noting that the assignment of representation rights (known as *enquadramento sindical*) is not always clear-cut, e.g., separate unions may claim the same set of workers and the employer may claim yet another union already holds the representation rights. All such matters are dealt by the labor courts.

[9]Despite universal coverage, CBA coverage in Brazil is around 50% partly because not every union has a CBAs covering all (sometime any) of the municipalities they represent.

[10]In some cases negotiations occur every two years—the maximum possible duration for a CBA.

There are 9 union centrals in Brazil, depicted in the right panel of Figure 3-1. The *Central Única dos Trabalhadores* (CUT) is the largest of these organizations, representing 30.4% percent of all organized workers in Brazil as of 2016.[11] CUT is Latin America's largest union central, and among the largest in the world. It has close ties with the *Partido dos Trabalhadores* (PT), or Workers' Party, which is Brazil's most prominent left-leaning political party. President Luiz Inácio Lula da Silva (founding member of PT) was the leader of a metalworkers' union within CUT before his move into politics—a common path for PT politicians (Lang and Gagnon, 2009).

CUT has vertically organized congresses and executive boards at the regional, state, and national levels. Congresses are meetings of delegates who are elected by individual unions to develop a coherent agenda for unions within CUT.[12] They meet once every three years to vote on CUT's overarching priorities for the subsequent 3 years, recorded in a book of resolutions or "fight plan". Executive boards comprise a smaller group of leaders elected by congresses to oversee CUT's day-to-day functioning. They manage CUT's finances, oversee the execution of the fight plan, organize meetings and training for local union leaders, and organize committees to tackle specific topics like gender equality within CUT.[13]

3.2.2 CUT Reform

The origin of the CUT reform we study arises from the tight link between this union central and the Workers' Party (PT). In 2011, PT instituted a 50% quota for women in its leadership, and its female presidential candidate, Dilma Rousseff, was elected as Brazil's first woman president. Together these events spawned a demand for greater gender parity within the CUT. Prominent female CUT leaders authored op-eds demanding greater say for women within CUT's leadership and a similar quota for women in the union (Godinho Delgado, 2017). They were successful. CUT's 2015 state and national congresses witnessed an unprecedented focus on women and instituted a pro-women reform that had two parts.

1) Gender quota First, CUT reserved 50% of seats in its state and national executive boards for women. This quota was voted in by the 2012 state and national congresses and came into effect

[11]The other union centrals are: *Força Sindical* (FS), *União Geral dos Trabalhadores* (UGT), *Central dos Trabalhadores e Trabalhadoras do Brasil* (CTB), *Nova Central Sindical de Trabalhadores* (NCST), *Central Geral dos Trabalhadores do Brasil* (CGTB), *Central dos Sindicatos Brasileiros* (CSB), *Intersindical—Central da Classe Trabalhadora*, and *Central Sindical e Popular Conlutas*.

[12]Elected delegates are typically local union leaders. The number of delegates that each union gets to elect to different levels depends on the number of workers it represents. Outlined in the CUT constitution here.

[13]For instance, CUT established the National Committee of Working Women in 1986 to campaign for universal childcare. In 2003, it gained a broader mandate to organize gender-related advocacy within CUT and became institutionalized as the Department of Working Women.

in 2015. Figure 3-2a shows that the quota had bite: the share of women in CUT's national board rose sharply from 35% to 50% in 2015. To accommodate having more women in its national board, the board size was almost doubled from 30 to 50 members. Importantly, there is no indication that other union centrals directly reacted to CUT's quota, maintaining a rather stable share of women on their national boards of around 21-25% (averaged across union centrals).[14]

Along with this large increase at the union central level, the quota had spillover effects on the representation of women in CUT-affiliated unions. Figure 3-2b employs a difference-in-differences strategy to compare the share of women on the local union boards of CUT and non-CUT affiliated unions. There is a small positive treatment effect, of 3% relative to baseline. This estimated effect is not mechanical as the quota only applied at the union central level and not also for its affiliates. Hence, this first part of the CUT reform should be interpreted as a leadership change favoring women mainly at the national level—where the involvement in collective bargaining is only indirect through, for example, coordinating the activities of the affiliated unions and their bargaining priorities.

2) Female-centric fight plan Second, the 2015 CUT national congress adopted a bargaining agenda more attentive to the needs of female workers. Its new fight plan featured a 14-page section on achieving gender equality in the workplace, which was the first time that such a section was authored in at least 10 years. Figure C.1.2 shows the cover of the 2015 fight plan. Some of its demands included advocating for 6 months of paid maternity leave (up from the state mandate of 120 days), reduced work hours and flexible schedules to accommodate women's household duties, and childcare as a universal right. The word *mulheres* (women) appeared 203 times in the 2015 fight plan, compared with 46 occurrences in 2012 and 74 in 2009.

CUT's 2015 fight plan also detailed a series of measures to promote gender parity within local unions. These included giving women chairmanship of important committees (like finance and communications) and involving women in the drafting of *pautas de reivindicações*, i.e., the list of union demands which are taken to employers for negotiation.[15] Therefore, independent of any change in women's representation on local union boards, these recommendations potentially translated into practices that elevated women's voices within local union boards.

[14]The small increase in the share of women on the boards of non-CUT union centrals in 2015 is driven by *Conlutas*—an even more combative left-leaning union central with a very small number of affiliated unions. CUT's main competitor union central is *Força Sindical*, which saw a small decline in the share of women on its national board in 2017 (Figure C.1.1).

[15]These strategies were developed at the 2015 meeting of CUT Women, and voted in as official CUT policy by delegates at CUT's 2015 national congress. The full text of the book of resolutions can be accessed here.

Summary In sum, starting in 2015, the CUT had more female leaders and vowed to advocate for women's priorities at the bargaining table. It made its commitment to gender equality especially evident to the local union delegates who attended its congresses. Importantly, the CUT reform did not change the bargaining power of unions relative to employers but merely refocused union priorities toward women. Hence, any improvements for women realized due to the reform are likely to reflect these new priorities, as opposed to a change in the share of surplus accruing to workers.

3.3 Data and Amenity Classification

To study how the CUT reform affects the workplace for women and at what cost, we need establishment-level information on wages, amenities, and employment, as well as each negotiating union's affiliation to a union central. This section first describes the data that satisfy these requirements. We then detail our data-driven approach to classifying amenities as male- or female-centric.

3.3.1 Data Sources

Our analysis relies on linking three rich sources of data: (i) amenities at the establishment-level from the text of all CBAs; (ii) worker outcomes from linked employer-employee data on the universe of formal sector workers; and (iii) union affiliation and leadership from the registry of unions. For information on amenities, we use CBA clauses scraped from the Ministry of Labor's *Sistema Mediador* registry, which tracks and stores every CBA signed in Brazil since 2009. To register an agreement, clauses need to be classified into 137 different clause types, e.g., overtime pay, childcare assistance, profit sharing, paid leave, etc.[16] We extract the number of clauses of each type as a measure of amenities offered to workers.

For information on worker-level outcomes we use linked employer-employee data known as *Relaçao Anual de Informacões Sociais* (RAIS). These are administrative data covering the universe of formal sector workers. Essentially, the federal government requires each employer to report key information regarding each worker employed in any given year. For each work spell, RAIS reports average monthly earnings, leaves taken, and (6-digit) occupation. It also reports worker characteristics like gender, age, and education; and establishment characteristics such as location (municipality) and industry (6-digit). We link RAIS to CBAs using an establishment identifier, known as CNPJ, that is common to both datasets.

[16] Figure C.1.3 shows an example of a maternity leave clause.

For information regarding each union's affiliation to a union central and its leadership composition over time, we use the national registry of unions, known as *Cadastro Nacional de Entidades Sindicais* (CNES). We infer the gender of leaders using the R package *genderBR*, which codes a name as female if most people with that name in the Brazilian census are women (and similarly for men).[17] Among all union leaders between 2005 and 2019, 27.7% are women, 67% are men, and 5% are unclassified. CBAs record the same union identifier as CNES, which we use to link contracts to unions, and, thus, union central affiliation and board composition.

3.3.2 Classifying Female-Centric Amenities

By matching CBAs to signing establishments in RAIS we can track workers across jobs, observing not only their wages but also a comprehensive set of amenities provided at each job. However, whether a CBA clause is differentially valued by women relative to men (what we denote as a female-centric amenity) is not directly observed in these data. We adopt two approaches to classify clauses as female-centric. Here we describe the key steps of each approach, with details in Appendix C.2.

1) Intuitive approach In the intuitive approach, we classify 20 of the 137 pre-specified clause types in *Sistema Mediador* as disproportionately valued by female workers (Table 3.2, Column 1). They fall into four broad themes, detailed in Table C.1.1: (1) Leaves, e.g., following maternity, adoption, or a miscarriage; (2) Maternity and childcare, e.g., employment protection after maternity, childcare assistance, and policies for dependents; (3) Workplace harassment and discrimination, e.g., sexual harassment and equal opportunities in promotions; and (4) Flexibility and part-time work, e.g., workday controls, uninterrupted shifts, and part-time contracts. Themes (1)-(3) include clauses that one could reasonable associate with women. The last theme reflects the fact that women disproportionately value flexibility in work hours (Goldin and Katz, 2011; Mas and Pallais, 2017; Maestas et al., 2018).

2) Data-driven approach In the data-driven approach, we aim to identify CBA clauses that correlate with women's disproportionate desire to work at an establishment relative to men. The underlying model motivating this approach is one where workers of gender $G \in \{F, M\}$ share a common ranking over establishments $j \in \mathscr{J}$. A worker's utility from working at establishment j is rising in the wage and amenities that it offers to their group G. In particular, we assume that the gender-specific value of working at an establishment (denoted V_j^G) is a linear function of

[17] Developed by Fernando Meireles and posted on GitHub.

wages, amenities, and an unobserved component:

$$V_j^G = \beta_w^G \psi_j^G + \sum_{z\in Z} \beta_z^G a(z)_j + \epsilon_j^G \tag{3.1}$$

where Z denotes the set of all amenities. Our classification problem is then to find the set of amenities for which the difference $\beta_z^F - \beta_z^M$ is positive, which we denote as "female-centric", as well as those for which this difference is negative, which we denote as "male-centric".[18]

At a minimum, we must measure the value of employment, wages, and amenities provided at each establishment. For the value of employment, we estimate gender-specific PageRank values by leveraging worker flows across establishments (Sorkin, 2018; Morchio and Moser, 2020). This is a revealed preference measure of the value of working at an establishment, which relies on the idea that good employers attract more workers, especially from other good employers.[19] For wages, we estimate the gender-specific wage premium at an establishment (ψ_j^G) using gender-specific AKM models.[20] For amenities, we use the average annual count of clauses $a(z)_j$ for each of the 137 clause types $z \in Z$ included in CBAs covering establishment j.

Hence, while we measure the gender-specific value of employment and wage premia at each establishment, we only observe a proxy for amenities without knowing which clause types are disproportionately valued by women and which by men. To identify these clauses, we take the difference between the female and the male version of Equation (3.1) and estimate the following hedonic regression:

$$V_j^F - V_j^M = \beta_w^F \psi_j^F - \beta_w^M \psi_j^M + \sum_{z\in Z} \beta_z a(z)_j + \epsilon_j \tag{3.2}$$

where $\beta_z = \beta_z^F - \beta_z^M$ captures the value of the amenity for women relative to men. We estimate this regression using lasso to select amenities that are the most predictive of utility differences between women and men, controlling for gender-specific wage premia. We deem the top 20 clauses with the highest β_z "female-centric", and the bottom 20 with the lowest β_z "male-centric". To the best of our knowledge, this is the first time that such a rich description of the work environment has been combined with administrative data on worker flows to uncover which features of the workplace are valued by different groups of workers.[21]

[18] An advantage of the data-driven approach relative to the intuitive approach is that it identifies male-centric clauses, allowing us to test for tradeoffs in male amenities following the CUT reform.

[19] Appendix C.3 describes the approach in detail and Appendix C.2 describes our implementation.

[20] AKM is the acronym for Abowd et al. (1999), which is the original paper estimating firm-specific wage premia using linked employer-employee data. Their underlying model also assumes a common job ladder among workers and identifies the firm effect using worker flows (see Appendix C.3 for details and Appendix C.2 for implementation).

[21] Several papers elicit workers' willingness-to-pay for a small set of workplace attributes such as flexibility and wage growth (e.g. Mas and Pallais (2017) for workers on an online platform, and Wiswall and Zafar (2017) for NYU

Omitted variable bias While the data-driven approach is a predictive exercise, mitigating omitted variable bias is still important. For example, establishments that want to more hire women may redouble their recruitment efforts or provide other job features that are valued by women, in addition to increasing observed clauses. Because we do not directly observe recruitment intensity or perfectly observe the work environment, we may erroneously identify a clause as valuable because it covaries with these unobserved features.[22] To mitigate this bias, we use amenities $a(z)_j$ from sectoral CBAs negotiated with several employers in an industry and geography instead of firm-level agreements negotiated with a single employer. Unlike the latter, sectoral CBAs are not influenced by demand shocks affecting individual employers.[23] Using sectoral CBAs for classification is also important because we use firm-level CBAs to study the CUT reform's causal effect. Using separate CBAs for classification and analysis prevents a mechanical relationship between clauses identified as female-centric and those that increase after the reform. Women flocking to treated establishments following the rise in female-centric amenities is then not a pre-determined result.

Estimation sample We estimate Equation (3.2) in the cross-section of establishments for which we can estimate V_j^G, ψ_j^G, and $a(z)_j$. First, because we must observe PageRank values for both genders, which can only be estimated for the largest super-connected set of employers (i.e., each establishments must hire from and lose a worker to another establishment in the set), our sample is restricted to the 2009-2016 intersection of these gender-specific super-connected sets. Second, AKM wage premia are only estimated for the largest connected set of establishments for which estimates are not noisy (average size ≥ 10 workers). The sample is thus also restricted to the 2006-2016 intersection of these largest connected sets between genders. Third, we reduce noise in $a(z)_j$ (i.e., the over-year average of clause type z), by restricting the sample to employers covered by at least four sectoral CBAs between 2009-2016.

Normalization Both PageRank values and AKM wage premiums must be normalized to make the gender difference in them interpretable. For AKM premiums, we normalize ψ_j^F and ψ_j^M relative to the restaurant sector—a fairly competitive industry where one can reasonably assume a zero wage premium for both genders. For PageRank values, V_j^F and V_j^M are unique up to unknown multiplicative factors. Our results are robust to three alternative methods for calculating

college students). They find that women value flexibility in work schedules more than do men. In the same context as ours, Lagos (2021) quantifies the wage-equivalent value of CBA clauses undistinguished by gender.

[22] Including ψ_j^G partly addresses this concern by accounting for recruitment efforts operating through wages.

[23] The results are not driven by industry-specific amenities and are similar when including industry fixed effects to leverage variation across geography; see footnote 23.

$V_j^F - V_j^M$. The first chooses the establishment with the smallest gender gap in wage premiums as the normalizing establishment, and divides the female value of all other establishments by the ratio $\frac{V_j^F}{V_j^M}$ at this establishment. The second simply assumes the multiplicative factor is the same for both genders, i.e., no normalization is needed. The third method re-scales the values V_j^F and V_j^M to a scale from 0 to 100. The base method for identifying male and female-centric amenities in the data-driven classification uses a 50% random sample of establishments and the first method for normalizing PageRank values.

Results Table 3.2, Columns 2 and 3 list amenities identified as female and male-centric using the data-driven approach. Clauses are ranked in descending order of the absolute value of $\hat{\beta}_z$. The clauses in red are those also intuitively classified as female-centric.

In line with the intuitive definition, the data-driven approach reveals that women disproportionately value clauses governing leaves (e.g., following adoption and miscarriage), childcare, and maternity (e.g., childcare assistance, maternity protections, and policies for dependents). In addition, they value 12 other provisions missing from the intuitive classification, including absences, extensions or reductions of the workday, medical exams, and health education campaigns.

On the male side, we also obtain sensible results. Men highly value additional pay, such as clauses governing on-call pay, profit sharing, hazard pay, workday compensation, life insurance and death or funeral assistance. They also disproportionately value workplace safety, such as protections for injured workers, machine and equipment maintenance, and safety equipment.[24]

The fact that "female workforce" clauses appear among those disproportionately valued by men highlights the fact that our approach does not account for variation in the text of clauses. These "female workforce" clauses vary widely in content, including items that are clearly beneficial to women (e.g., free provision of sanitary pads), as well as those clearly beneficial to men (e.g., forbidding women to cast concrete or install scaffolding). It is likely, then, that our data-driven approach captures the latter. While the availability of pre-specified clause types allows us to have a simple measure of CBA content that avoids the drawbacks that plague more complicated topic models—such as text pre-processing, choosing the number of topics, and noisy estimates—it is by no means a faultless measure.

[24]The clauses classified female-centric remain similar across various normalizations of PageRank values (Tables C.1.4 and C.1.5). Moreover, the classification is not driven by industry or geography-specific amenities, since it is invariant to including industry- and state- fixed effects. The rank correlation of the coefficient β_z on the selected clauses with and without these fixed effects is positive and statistically significant (0.56 with $p - value < 0.01$). Tables C.1.2 and C.1.3 offer specific examples of clauses identified as female and male-centric.

Sense checks Out-of-sample sense checks indicate that both the “intuitive" and “data-driven" approaches identify clauses that women (or men) value disproportionately more than the other gender. Using firm-level CBAs signed in 2014—the year prior to the CUT reform—we find that female (male)-centric clauses increase with the share of women (men) at an establishment.[25] Figure 3-3a shows that intuitively classified female-centric clauses increase almost linearly with this share. Figure 3-3b shows a similar relationship for male and female-centric clauses defined using the data-driven method. Specifically, all-male workplaces offer ≈1.5 more male than female clauses, with this gap shrinking to almost zero at all-female workplaces. Interestingly, female clauses per the data-driven classification only begin to increase once women comprise the majority in an establishment (above the 50% threshold). This suggests either that women successfully advocate for these amenities once in the majority, or that establishments provide them to attract female workers—both implying higher value among women.

3.4 Empirical Strategy

We employ a difference-in-differences strategy to study the CUT reform's effect on amenities and labor market outcomes. This section first describes the three analysis samples we use to study the reform's effect on collective bargaining agreements, establishments, and workers. We then detail our empirical approach and identifying assumptions.

3.4.1 Analysis Samples

We construct three analysis samples to study the CUT reform's effects on negotiated CBAs, establishments, and workers. Appendix C.2 provides detail.

1) Amenities sample To study the evolution of amenities, we construct a balanced panel of each pair of establishment-and-negotiating union covered by firm-level collective bargaining between 2012 and 2017. Each of these pairs can be thought of as constituting a unique worker group, because the same union represents any category (usually industry) of workers in a given geography.[26] Our analysis focuses on clauses in firm-level CBAs because most improvements in amenities and working conditions are achieved through these agreements (Horn, 2009; Liukkunen,

[25] In addition, the number of female clauses is strongly positively correlated with the difference between women and men's PageRank valuation of an establishment (Figure 3-3).

[26] Most signing establishments (93%) negotiate with a single union over the entire study period, meaning that employers rarely negotiate with more than one worker category.

2019).[27]

While not every establishment-union pair renegotiates its contract every year, we obtain a balanced panel of contracts by exploiting the fact that, during our period of study, the coverage of old CBAs is automatically extended until a new agreement is negotiated (Lagos, 2021). Given both that all CBAs were required to be registered in *Sistema Mediador* beginning in 2009, and that they span at most 2 years, our panel paints an accurate picture of active CBAs between 2012 and 2017. Our results are robust to instead using an unbalanced panel that comprises only new contracts.

2) Establishment sample To study the possible downstream effects of changing amenities on labor market outcomes as well as wage and employment tradeoffs, we construct a sample of establishments signing CBAs in our *amenities sample*, and track their outcomes in RAIS. Outcomes include employment, the share of women among workers, and mean log wages. We make two additional restrictions to this sample. First, we restrict attention to establishments that employed both men and women at baseline (2014). Second, we only consider an establishment signing a contract as covered by its contents if it lies within the contract's geographic coverage. This restriction allows us to exclude headquarters that sign contracts on behalf of their subsidiaries, and are hence outside the contract's geographic coverage.

3) Incumbent worker sample To study individual worker-level outcomes such as wages and retention, we construct a sample of incumbent workers employed at establishments in the *establishment sample* at baseline (2014). We track these workers wherever they go, i.e., not conditional on staying at their baseline employer.

Treatment definition Following the 2015 reform, CUT-affiliated unions prioritized women in their collective bargaining strategy. While the reform was enacted in 2015, the gender quota was approved in 2012 (see Section 3.2.2), suggesting that CUT's pro-women pivot may have been anticipated and spurred unions to switch affiliation to avoid or benefit from the pivot. Although unions rarely switch their union central affiliation, we define treatment using a union's 2012 CUT affiliation to avoid bias from selection into or out of CUT affiliation. Figure C.1.5 confirms that neither treated nor comparison unions systematically switched affiliation away from or

[27] In an informal conversation, the President of the bankers' union of São Paulo also confirmed that most amenity improvements are achieved through firm-level CBAs. Sector-level negotiations typically involve several tens (or even hundreds) of employers, making it difficult to reach consensus over a rich set of amenities. Unions therefore typically reserve these topics for negotiation with individual employers.

toward the CUT following its 2012 announcement of a gender quota. Thus, there is no concern from endogenous selection even had we used a later affiliation year.

Treatment is defined in the following way. In the *amenities sample*, a treated establishment-union pair is one where the negotiating union was affiliated with the CUT in 2012. In the *establishment sample*, a treated establishment is one belonging to treated pair.[28] Finally, in the *incumbent worker sample*, a worker is treated if employed at a treated establishment at baseline (2014).

Descriptive statistics Table 3.1 describes our starting sample, i.e., the *amenities sample*. Column 1 describes the full sample, and Columns 2 and 3 report information by treatment status.

Panel A reports sample sizes. Our sample comprises more that 211 thousand firm-level CBAs signed by 89,920 establishment-union pairs. These pairs cover 80,131 signing establishments and 4,409 signing unions. On average, each pair signs new contracts in 2.4 out of the 6 years spanning our study (2012-2017). Of all pairs, 21% are treated and 79% are in the comparison group.

The amenities sample covers over 19% of total formal employment in Brazil, and 2.1% of establishments. These numbers highlight two points. First, only a select set of employers negotiate firm-level CBAs. Second, these establishments are substantially larger than the average establishment in Brazil, employing 143 workers on average compared to 16 among all establishments (Table C.1.6).[29] The establishment sample, where establishments must additionally have been employing both men and women in 2014, covers 15% of the total 2014 workforce, and otherwise resembles the amenities sample in the size, sector and regional distribution of its establishments.

Panel B of Table 3.1 describes contract provisions at baseline (2014). CBA negotiations (at the pair-year observation level) feature 24.7 clauses on average, of which 3.2 are classified “female-centric" per our data-driven definition (Section 3.3.2). On average, contracts feature 1.7 more male clauses than female clauses. These numbers are statistically indistinguishable across treated and control contracts. Although the share of female-specific clauses may appear to be small, this statistic may not accurately represent the value and importance of these clauses. For example, even a single contract provision extending maternity leave by 60 days may prove very valuable to young women. Thus, in addition to considering how the CUT reform af-

[28]Over 93% of establishments negotiate with a single union and 98% with all unions with the same union central affiliation. For the remaining 2% of establishments, treatment is defined as negotiating with any treated union.

[29]Compared to the average Brazilian establishment, an establishment signing firm-level CBAs is more likely to operate in manufacturing rather than commerce (difference of 16-19pp for each); these establishments are more likely to be located in the affluent Southeast and less in the poorer Northeast region of Brazil (Table C.1.6).

fects amenities on paper, we will infer how valuable these changes are to women by studying revealed preference changes in their sorting behavior across establishments.

Panels C and D document establishment- and union-level characteristics, respectively, at baseline (2014). Our sample comprises large employers (especially in the treated group). The average establishment employs 143 workers, over a third of whom are women. A majority of establishments employ both men and women (82%). On the union side, treated unions have larger boards but with a similar share of women as comparison unions (around 23%), indicating no baseline difference between CUT and non-CUT affiliates. Only about 17% of unions have a female president.

Treated and comparison establishments exhibit substantial overlap along a number of observable dimensions, including their distribution of size, geography, industry, and share of women in the workforce (Figure C.1.6). Appendix Table C.1.7 statistically explores differences by treatment status. Treated establishments are larger than control establishments, but employ a similar share of women. They are more likely to be located in the North-East region (15% treated versus 11% control) and engage in manufacturing (32% treated versus 28% control). All analyses control for differences in industry and geography across treatment status through 2-digit-industry by year and geography by year fixed effects.

3.4.2 Differences-in-Differences Design

To measure the causal effect of the CUT reform on negotiated amenities and labor market outcomes, we compare treated units of observation (i.e., pairs, establishments, or incumbent workers) with the comparison group using a dynamic difference-in-differences specification:

$$Y_{it} = \sum_{j=2012}^{2017} \beta^{t=j}(D_i \times \delta_{t=j}) + \alpha_i + \gamma X_{it} + \varepsilon_{it} \tag{3.3}$$

where i indexes the unit of observation and t indexes a year. The treatment indicator D_i is interacted with year fixed effects δ_t. The specification also includes unit fixed effects α_i, as well as time-varying fixed effects X_{it}, i.e., industry-year and geography-year fixed effects.[30,31] Idiosyncratic errors are captured by ε_{it} and standard errors are clustered by establishment.[32]

[30]For industry we use the first two digits of Brazil's CNAE codes. There are 87 unique industries, including textile production, road transportation, and construction.

[31]For geography we use either states (27 in total) or microregions, which are neighboring municipalities grouped into 543 units that capture local labor markets.

[32]Clustering by establishment assumes that establishments negotiate with unions that, as of 2012, were affiliated at random with a union central. Results are unchanged when clustering by union.

The coefficients of interest, β^t, capture the effect of treatment in year t relative to the baseline year (β^{2014} is normalized to zero). The model allows for average differences between treated and the comparison units, absorbed by unit fixed effects α_i. The identifying variation occurs within the same unit, comparing outcomes in any year relative to 2014, and within the same time period, comparing treated and comparison units. The identifying assumption is that outcomes would have evolved in parallel at treated and comparison units absent the CUT reform, conditional on covariates. We assess the plausibility of this assumption by testing for parallel trends in the pre-period.

To summarize the average post-period impact of the CUT reform we run a "pooled" version of the above regression, which amounts to replacing the full interaction of D_i with year-specific indicators δ_t with a single interaction for the post-period, $D_i \times \delta_{t \geq 2015}$. In addition, to make treatment effects in worker-level regressions interpretable as establishment-level averages, we weight each incumbent worker by the inverse of (own-gender) employment at their baseline employer (Jäger et al., 2021). Finally, it is worth noting that outcomes that may change as a downstream consequence of changing amenities (e.g., wages and retention) are unscaled by the amenity change since we do not directly observe the value workers assign to said amenities.

3.5 Impact of the CUT Reform

This section presents our main results. We start by analyzing the CUT reform's effect on amenities, finding disproportionate gains in women's amenities on paper and in practice. We then explore whether women value these changes to CUT workplaces by studying the reform's impact on two revealed preference measures of firm value—retention, and job queues. We conclude by evaluating potential tradeoffs from the improvement in female-centric amenities—in men and women's employment, wages, and in firm profits.

3.5.1 Amenities: On Paper and In Practice

Negotiated amenities Table 3.3 reports the CUT reform's pooled DiD treatment effect on female and male-centric clauses and Figure 3-4 reports year-specific effects.[33] Pre-reform female amenities evolve in parallel, supporting our identification assumption. Immediately following the reform, female clauses in treated contracts rise sharply in number (intensive margin), incidence (extensive margin), and as a share of all clauses. On the intensive margin, the number of

[33] Figure C.1.7 plots the raw path of female-centric clauses in treated and comparison contracts.

intuitively defined female clauses increases by 0.157 (SE 0.013)—a 17% increase relative to baseline (Panel A). Data-driven female clauses rise by 0.301 (SE 0.021), a 19% increase. These effects represent a substantial improvement, equivalent to moving from the average baseline amenity count at an establishment with a minority female population to one with over 80% women. The effects do not reflect a mere increase in the number for clause types already being provided in contracts, for example, going from 1 to 5 maternity leave clauses. Rather, they represent the inclusion of *new* female-focused clauses, with the sum of unique clause types increasing by 12% over its baseline value (Panel B).

The CUT reform also increases the occurrence of any female-centric clause (extensive margin, Panel C) and these clauses as a share of all clauses in the contract (Panel D). On the extensive margin, the incidence of female-centric clauses increases by 1.7pp (SE 0.003)—a 5% gain over baseline. Using the data-driven classification, this effect is 3.4pp (SE 0.003), representing a 10% increase. As a share of all clauses, intuitive female clauses rise by 0.5pp (SE 0.001), a 10% increase over baseline, and data-driven clauses by 2.1pp (SE 0.001), denoting a 30% increase.

All four types of female clauses rise—leaves, childcare payments, anti-harassment, and flexibility (Column 2-5), with clauses governing leaves and childcare accounting for 76% of this gain. The CUT-driven improvement in amenities is thus likely to differentially impact workers of childbearing age, a fact that we will later exploit to zoom in on labor market outcomes among these workers.

There is some evidence that unions trade off men's interests in favor of women's, but only negligibly. Both the extensive margin and share of male amenities decline by small amounts: by 0.1pp (SE 0.003) relative to 46% at baseline for the former, and by 0.3pp point (SE 0.002) relative to 14% for the latter (Column 7). While the number of male-centric clauses increases, this gain is more than overshadowed by the gain in female-centric clauses. Moreover, while the treatment effect on female amenities occurs sharply in 2015, for male amenities it occurs in 2017, suggesting that the male effect is unlikely to be driven by the CUT reform (Figure 3-4).[34] Overall, the ratio of female-to-male clauses rises by 21% over its baseline value in treated versus comparison contracts (Column 8). The CUT reform therefore increases the female-orientation of contracts, driven by an increase in female-centric clauses.[35]

Turning to the question of *where* union priorities exert the greatest influence on female amenities, we find the largest gains at establishments where women could not already advo-

[34]The increase in male amenities is not robust to clustering by the union, whereas the gain in female amenities is (Table C.1.8).

[35]These results are robust to reasonable amendments to the data-driven definition of male- and female-centric amenities, the inclusion of more granular industry-geography-year fixed effects, and conditioning on establishment-union pairs with coverage in 2014 (Tables C.1.9, C.1.10, C.1.11).

cate for themselves either as workers or as union leaders (Table 3.4, Figure C.1.9). Specifically, at establishments with a small baseline share of women in the workforce (below median, Column 2), in union leadership (Column 3), or without a female union president or vice-president (Column 4).

In terms of mechanisms, our findings could either reflect a change in the composition of union leadership through more female leaders, or a shift in broader union priorities without a direct increase in female leadership. Figure C.1.10 shows a small positive treatment effect on the share of women among union leaders (0.7pp or 3%). However, while these newly elected female leaders may have been instrumental in implementing the CUT's new priorities, they do not account for the reform's full effect, as we also find large improvements in female amenities in contracts negotiated with unions without any new female leaders. We interpret this as evidence that the amenity increase stems from a broader shift in union priorities toward women, rather than simply changes enacted by the women themselves.

On a final note, it is worth highlighting that CBA clauses represent equilibrium outcomes resulting from negotiations between unions and employers. As such, our results show employers' willingness to sign off on female-friendly amenities. Upcoming analyses explore whether this willingness reflects changes on paper not translating into practice, employers adjusting compensation along other dimensions (such as wages), a reallocation of surplus toward workers, or the proposed changes leading to pareto improvements for workers and employers.

Actual amenities To assess whether the change in amenities on paper translates into practice, we draw on the text of female-centric clauses to identify three outcomes that they can directly affect: (i) whether women are managers—corresponding with equal opportunity clauses; (ii) whether women take longer maternity leaves—corresponding with clauses that extend maternity leave; and (iii) if women enjoy job protection post maternity leave—corresponding with job protection clauses.

The reform positively affects outcomes along all three dimensions (Figure 3-5). The share of women among managers at treated establishments increases by 2% relative to baseline. Women also take longer maternity leaves, with a 14% treatment effect on the share of mothers taking leaves longer than the state mandate of 120 days. Despite these longer leaves, mothers are no less likely to return to their employer following maternity, implying that they enjoy longer periods of job security. Thus, the new union priorities enacted by the CUT reform lead to actual improvements in the workplace for women.

We similarly draw on the text of male-centric clauses to study whether workplaces deteriorate for men. Per the data-driven approach, men value safety. We find no treatment effect on workplace safety as captured by the share of workers taking work-related injury leave. If anything, there is a -3% treatment effect on this outcome. Thus, at least on this dimension, the workplace does not deteriorate for men.

3.5.2 Revealed Preference Changes in Firm Value

Our analysis of improvements in actual amenities is limited to observables in the RAIS data. To more comprehensively understand whether workers actually value these changes to CUT workplaces, we study the reform's impact on two revealed preference measures of job quality: employee retention and job queues.

Retention Retention serves as a revealed preference measure of an employer's attractiveness relative to others (Krueger and Summers, 1988). We find a 1.8pp (SE 0.004) increase in retention among incumbent women, a 2.5% improvement over baseline.[36] The gender difference in this treatment effect is 0.08pp (SE 0.003), suggesting that incumbent women disproportionately value the reform over its value for incumbent men (Figure 3-6a). Since we find the largest improvement in amenities related to maternity and childcare, we also zoom in on retention among workers of childbearing age (20-35 years), finding a similar treatment effect (Figure 3-6b).

However, higher retention need not imply that women value these jobs more if it reflects fewer firings instead of fewer quits. To assess this possibility, we decompose the treatment effect on retention into a component explained by employer-to-employer transitions, likely reflecting quits, versus transitions into unemployment, more likely after a firing. Consistent with a revealed preference story, we find that the treatment effect on retention is explained by fewer voluntary employer-to-employer transitions as opposed to fewer firings into unemployment (Table C.1.12).[37]

Since the share of male-centric clauses negligibly falls, men may value CUT employers less. However, we find a 1.0pp *increase* in retention among incumbent men (Table C.1.12), representing a 1.5% increase over baseline. That men quit less provides strong evidence against the hypothesis that men are worse off due to the CUT reform. Thus, although the reform disproportionately improves working conditions for women, it does so without driving men to other jobs.

[36] The two-year baseline retention rate among women is 71%.

[37] Voluntary transitions among incumbent women (men) decline by 1.1pp (0.8pp).

Job queues Job queues are a second revealed preference measure of value (Holzer et al., 1991). Because we do not directly observe applications, we use workers in the probationary period, i.e., the first 3 months of tenure, as a proxy measure. Since Brazilian labor law permits employers to terminate probationary workers without severance pay, such contracts are commonly used to screen workers.[38] We find a 0.6pp increase (SE 0.003) in women's share among probationary workers (Figure 3-6c), a 1.7% improvement over baseline. This suggests that women are more likely to queue for jobs at treated establishments.

Although precise, the magnitude of this estimate is small. Three factors likely dampen the estimate of women's queuing response at CUT establishments. The first (as previously discussed) is our inability to directly observe changes in amenity values using which to scale treatment effects.[39] The second is information frictions that may prevent workers from learning of newly instituted amenities at CUT establishments.[40] Finally, employers may potentially screen women out at the hiring stage, such that any change in composition among probationary workers is already muted.

In sum, we find that women flock to CUT establishments following the reform. Together their lower separation from, and higher likelihood of queuing for jobs at, CUT establishments translate into a 0.2pp increase in women's share among employees. Section 3.6 uses these revealed preference changes in firm value to quantify the CUT reform's effect on worker welfare.

3.5.3 Tradeoffs

How are the improvements in female-focused amenities paid for? Table 3.5 explores three potential explanations.[41] First, employers may finance amenity improvements by reducing women's wages, as predicted by compensating differences (Rosen, 1986). Alternatively, if unable to pass the cost of amenities onto workers' wages, employers may reduce employment or employ relatively more men or inexpensive workers/older women (Summers, 1989). Finally, firms may finance improvements through lower profits.

[38] For example, 25% of all separations occur between 3 months and 3 months and 1 day.

[39] Since PageRank values can only be estimated for the super-connected set of firms, it is infeasible to separately estimate pre and post-period values covering a reasonably large sample of firms given only 3 years of data per period.

[40] As an anecdotal example, an economics professor believed that she was eligible for extended maternity leave because a co-worker at the same institution had obtained such an extension. However, this professor's location was not covered by the same CBA as her colleague, meaning that she was ineligible for the maternity leave extension.

[41] Using the establishment sample.

Wages If amenity improvements operate in a compensating differences world, women's wages should disproportionately decline. Table 3.5, Panel A reports the treatment effect on wages and Figure C.1.11 shows parallel pre-trends. Because Brazilian employers cannot cut nominal wages for existing workers without the union's approval, wage adjustments may only realize for new workers. We therefore separately study the reform's effect on the mean log wage of established workers, with tenure over 12 months, and new workers, with tenure under 12 months, separately by gender.

There is no treatment effect on the mean log wage of any worker group—established, new, men, or women. All point estimates are negative but very small and precise—the largest decline occurs for new male workers, whose wages fall by 0.6pp (SE 0.003).[42] We rule out negative effects greater than 1.2-1.3pp for new workers, and 0.7-0.8pp for established workers, at a 95% confidence level.[43] Given the similar point estimates for wage changes among men and women, there is no change in the gender wage gap. Overall, there is little evidence in favor of employers lowering wages to pay for higher female-centric amenities.

There are two important caveats to this finding. First, the average worker may not adequately represent workers whose wages are actually influenced by unions. For a more direct measure of union-negotiated wage changes, we extract the percentage wage adjustments negotiated in collective bargaining agreements. The treatment effect on these wage adjustments is 0.032pp (SE 0.021), allowing us to rule out a more than 0.009pp fall in wages with a high degree of confidence (95%). Second, employers may respond by changing the composition of their workforce, such that zero wage effects mask effective wage changes for new workers. However, we find an incredibly precise null treatment effect on the wages of incumbent workers, whose composition is unchanged (Table C.1.12).

Employment If employers cannot pass the cost of amenity improvements onto workers' wages, they may lower employment. Table 3.5, Panel B reports the treatment effect on employment and Figure C.1.11 shows parallel pre-trends. Column 1 reports the effect on overall employment and Column 4 on hiring. We find no statistically significant impact on employment or new hiring among treated employers, and can rule out negative effects larger than 0.15pp with a high degree of confidence (95%). We also find no decline in the employment or hiring of female workers; if anything, as previously discussed, women's share among all workers increases

[42]This result is not robust to including industry-geography-year fixed effects.

[43]By way of reference, Lagos (2021) finds that workers value leave clauses, many of which are classified as female-centric, at 8.4% of their wage. That paper pools men and women together.

by 0.2pp and among probationary workers by 0.6pp.[44]

A second dimension of adjustment is worker composition—employers may hire more skilled or older workers. Table C.1.13 provides evidence against this hypothesis. There is no change in the proportion of female workers poached from other employers (a measure of positive selection). Moreover, there is no treatment effect on the mean age, tenure, contracted hours, or schooling of female workers.

In sum, we find no evidence that employers hire fewer women, fewer workers, or different or more productive workers as a result of the CUT reform. Of course, we cannot rule out productivity gains among female workers *as a result of* the change in workplace environment. Indeed, this is a candidate explanation for our finding of no wage or employment tradeoffs due to the CUT reform.

Profits If workers do not finance the amenity improvement through lower wages or employment, perhaps firms finance it through lower profits. We provide empirical evidence and theoretical reasons against this explanation.

Table 3.5, Panel C shows no treatment effect on firm profits, measured in two different ways. First, we find no treatment effect on establishment exit. Exit is a non-trivial margin of adjustment in Brazil, with 8.7% of control group establishments exiting between 2014 and 2017. Second, we estimate a statistically insignificant 0.70pp (SE 1.17) treatment effect on profits among the sample of establishments that is observed in Orbis data during our study period. For this restricted sample, we rule out a higher than 1.59pp decline in profits with a high degree of confidence (95%).

Theoretically, profits could only fall if CUT unions were able to bargain away a larger share of surplus from employers. However, there is little reason to think that the CUT reform increased unions' bargaining power; rather, it merely shifted union priorities to favor women. If anything, the position of CUT-affiliated unions grew increasingly precarious around this time, following the 2015 impeachment of President Dilma Rousseff of the left-wing Workers' Party with which the CUT has close ties. Moreover, while increasing union bargaining power generally predicts a change in employment—either moving right along a firm's upward-sloping labor supply curve, or left along its labor demand curve—we find a precisely estimated zero.

In sum, we find no evidence that profits decline to pay for the female-focused improvement in amenities.

[44] There is also a small positive effect on the share of women among separators due to more women being hired and working at the firm. However, on net, the share of female workers increases.

3.5.4 Robustness

Brazil experienced a recession between 2014 and 2016. Our findings may be driven by the recession as opposed to a shift in union priorities if CUT unions either represent systematically different industries that are differently impacted by the recession, or if these unions differently respond to the recession. Several findings point against the differential impact of, or response to the recession as driving our findings. First, our results reflect an increase in female amenities in CUT contracts as opposed to a potentially-recession-induced-decline in amenities in non-CUT contracts (Figure C.1.8). Second, there is little reason to expect the recession to have increased the CUT's demand for female-centric amenities such as maternity leave or childcare payments (as opposed to clauses that shield workers' wages, which may arguably constitute a more natural demand during a recession). Third, we find heterogeneous treatment effects, with the largest amenity gains occurring at establishments with a small baseline share of women; this heterogeneity counters the idea that the CUT in general responded differently to the recession. Finally, all specifications include 2-digit-industry and location-specific time varying shocks through industry by year and microregion by year fixed effects.

3.5.5 Discussion

The CUT reform that pushed union leaders to prioritize women's needs in collective bargaining improved the work environment for women relative to men, both on paper and in practice. Women valued these changes, becoming less likely to separate from and more likely to queue for jobs at CUT establishments. Perhaps surprisingly, we find no evidence that these gains in female-focused amenities come at the expense of women or men's wages and employment, or of firm profits. While amenities for men may have fallen (in some unobserved way), men do not exit more.

Together our findings demonstrate that shifting union priorities can reduce the gender compensation gap. Just as in politics, where leaders' priorities determine policy design (Chattopadhyay and Duflo (2004); Pande and Ford (2012)), we show that unions' priorities determine workplace design. We consider a broader definition of compensation than wages alone, including also amenities such as family allowances, leaves, and flexibility, and show that these are key levers through which unions influence inequality.

There are at least three models that could explain our results. In one model, men lose rents due to the reform—which are not observed in amenity or wage changes—but they are not marginal to this loss since these rents are not provided elsewhere. A second model is one in which the union was inefficiently aggregating workers' preferences. Shifting union priorities caused it to focus on previously ignored female amenities that could be provided at net zero cost to employers. Finally, providing amenities may have increased firm profits and the total size of rents split between unions and employers. Behavioral unions and firms may have been leaving these gains on the table until the reform spurred a shift in focus. The last two explanations represent pareto improvements. Since male retention slightly improves, and male wages and amenities do not decline, our findings are most in line with one among the last two explanations.

3.6 Quantifying the Welfare Effect of the CUT Reform

The CUT reform increased female-centric amenities and made CUT establishments more valuable to women. By how much did women's welfare change? What about the reform's impact on men's welfare? We briefly describe our approach here with details in Appendix C.4.

Approach and Intuition We quantify the CUT reform's effect on worker welfare through a revealed preference approach that (i) relies on a few sufficient statistics that are easily computable in the data; and, thus, (ii) takes no stance on the precise functional form linking amenities to worker utility. In particular, we adapt a framework used to evaluate changes in consumer welfare from introducing new or improved product varieties (Feenstra, 1994; Redding and Weinstein, 2016) to our labor market setting.

For tractability, we assume that workers possess CES preferences over employers, as is common in the consumer setting (Feenstra, 1994; Atkin et al., 2015). As shown in Anderson et al. (1992), a key advantage of CES is that it generates the same labor supply to firms as obtained by aggregating workers' discrete choices over where to work based on where they obtain the highest utility. This is a common way of modeling the labor market (in Card et al. (2018); Sorkin (2018); Berger et al. (2022); Lamadon et al. (2022)). In Appendix C.4 we microfound CES demand using such discrete choices.

Then, just as gains to consumer welfare from improving product varieties can be measured though changes to the price index—i.e., the change in cost of purchasing one util worth of utility—the gains to worker welfare from improving workplace amenities can be measured through changes to the wage index—i.e., how much more (or less) the representative worker earns to work one disutility-weighted hour.

Under CES preferences, only four sufficient statistics quantify the change in worker welfare, i.e., measure the change in the wage index. First, welfare increases with the share of total labor income found at treated establishments, which captures workers flocking to these employers after they improve amenities. Second, the same change in labor income at treated establishments corresponds with a higher increase in welfare if workers are less elastic to begin with, since it takes a larger improvement in amenities to draw them away. Third, welfare is higher if workers are drawn away from non-CUT firms with initially low value, capturing a bigger upgrade in employer quality across regimes. Finally, welfare increases with wages at non-CUT establishments, potentially capturing the pro-competitive spillover effects of the reform.

Model In each period, a representative household with CES preferences over employers is willing to work a fixed number of (dis)utility-weighted hours. It chooses labor supply to each firm to maximize total income, subject to this hours constraint:

$$\max_{\{n_{jt}\}} \sum_{j \in \mathscr{J}_t} w_{jt} n_{jt} \quad s.t. \quad \left[\sum_j (b_{jt} n_{jt})^{\frac{1+\eta}{\eta}} \right]^{\frac{\eta}{\eta+1}} = N, \tag{3.4}$$

where $\mathscr{J}_t$ denotes the set of firms operating at time t, n_j is the number of hours supplied to firm j, w_j is the wage at j, η is the elasticity of substitution across firms, and b_j represents the "taste-shifter" for firm j. b_j captures all non-wage attributes that commonly affect each worker's utility at j. Worse amenities increase this disutility b_j. We assume a utility-posting world without job rationing, where a firm accepts any worker who wishes to work there. For simplicity, since worker welfare only depends on firms' final wage and amenity offers, regardless of how firms arrive at them, we do not model the firm side.

The wage index measures how much the representative worker is paid to work a disutility-weighted hour, and serves as a measure of welfare:

$$\tilde{W} = \left[\sum_{j \in \mathscr{J}} \left(\frac{w_j}{b_j} \right)^{1+\eta} \right]^{\frac{1}{1+\eta}}$$

Any change in the wage index across two periods captures changes to worker welfare, measured by the ratio:

$$\phi_{t-1,t} = \frac{\tilde{W}_t}{\tilde{W}_{t-1}}$$

The CUT reform changes amenities, or taste shifters b_{jt}, at treated establishments. The key challenge in estimating welfare changes is that these $\{b_{jt}\}_{j \in \mathscr{J}_t}$ are unobserved. However, assuming CES preferences allows us to overcome this challenge. Under CES, any welfare change depends only on the *observed* pre- and post-reform wages and employment at CUT and non-CUT employers.[45] Formally:

$$\ln\phi_{t-1,t} = -\frac{1}{1+\eta}\ln\left(\frac{\lambda_t}{\lambda_{t-1}}\right) - \frac{1}{1+\eta}\ln\left(\frac{\overline{S^*}_t}{\overline{S^*}_{t-1}}\right) + \ln\left(\frac{\overline{w^*}_t}{\overline{w^*}_{t-1}}\right) \tag{3.5}$$

where λ_t is the share of total labor income in t at non-CUT firms, $\overline{S^*}_t$ is a geometric average of the share of labor income at each non-CUT firm in t, and $\overline{w^*}_t$ is a geometric average of period t wages at non-CUT firms. The asterisk * denotes that operations are taken over non-CUT firms.

Changes in welfare depend on three terms, as per Equation (3.5). The first, "variety-adjustment" term $\left(\frac{\lambda_t}{\lambda_{t-1}}\right)^{-\frac{1}{1+\eta}}$ is the ratio of the share of total labor income at non-CUT firms after relative to before the reform. This ratio captures welfare changes through a revealed preference logic: workers substitute toward CUT firms once their amenities improve, lowering the share of the labor income at non-CUT firms and increasing welfare. The magnitude of this change depends on the elasticity of substitution across firms. If workers are inelastic (η is low), the same move toward amenity-improving CUT-firms implies a larger welfare increase because it takes a bigger improvement in amenities to draw workers away.

The term $\left(\frac{\overline{S^*}_t}{\overline{S^*}_{t-1}}\right)^{-\frac{1}{1+\eta}}$ captures the heterogeneity in labor income at non-CUT firms: welfare increases by more if CUT firms draw workers away from less valued non-CUT firms, thereby increasing dispersion in and lowering the geometric mean of their wage bill share. As in the "variety-adjustment" term, the implied effects are larger as workers become more inelastic. The final term $\left(\frac{\overline{w^*}_t}{\overline{w^*}_{t-1}}\right)$ represents a change in wages at non-CUT firms, possibly as a pro-competitive response to the reform. As these "outside" wages increase, so too does welfare.

Estimation We separately estimate Equation (3.5) for men and women. Our estimates employ the establishment sample from Section 3.5.3. Years 2012-2014 comprise the pre-reform period $(t-1)$ and 2015-2017 the post-reform period (t). We calibrate an estimate of the cross-firm elasticity of substitution (η) from Felix (2022), but assess robustness to other reasonable values.

We estimate the log change in $\bar{w}^*$ and in $\bar{S}^*$ using average changes across non-CUT establishments between $t-1$ and t, estimated via the following regression:

$$y_{jt} = \alpha + \beta Post_t + \mu_j + \epsilon_{jt} \tag{3.6}$$

where y_{jt} is either the average log earning at establishment j ($\log w_{jt}$) or the log of the share

[45] Under CES, the relative (dis)utility of working at an employer is captured by its expenditure share, which depends exclusively on prices and quantities.

of labor income among non-CUT establishments at that establishment ($\log s_{jt}$). The specification includes establishment fixed effects μ_j. The coefficient of interest, β, captures the average within-establishment change in the dependent variable between between $t-1$ and t. Bootstrapped standard errors are clustered by establishment.

To estimate the change in λ we take a first order approximation of λ_t around λ_{t-1}. This allows us to map the market-level change in the share of labor income at CUT establishments (the desired object) to changes in quantities that are estimable through establishment-level regressions as in Equation (3.6). We refer the reader to Appendix C.4 for details.

Results Table 3.6 reports results. Women's welfare increases by 0.059 log points (or 6.1%), consistent with our reduced form results that women are more likely to remain at, and comprise a larger share of new workers among, CUT establishments.[46] Worker moves following the reform account for over half of the increase in welfare. Women become more likely to work at CUT establishments, accounting for 15% of the welfare gain (a 1.8% rise in the share of CUT wage bill). In addition, the dispersion in the labor income across non-CUT firms rises (i.e., S^* falls), accounting for 48% of the increase in welfare.

The remaining 37% of the welfare gain is accounted for by higher wages among non-CUT employers. To the extent that these wage increases reflect pro-competitive responses to the CUT reform, any change in welfare from them can also be attributed to the reform. We recognize, however, that the increase in real wages at non-CUT employers following 2015 could be driven by a host of factors that are unrelated to the CUT reform. We therefore only view the change in welfare due to worker moves across firms—amounting to a 3.8% increase—as the credible estimate of the reform's welfare impact for women workers.[47]

For men, welfare is slightly higher (1.3%), but remains essentially unchanged if one only considers the component due to worker moves across firms (0.2%). Thus, the CUT reform improves women's welfare without reducing men's welfare.

[46]As predicted by the model, workers' elasticity of substitution across employers amplifies (or dampens) the welfare effect due to the shifts in employment across firms induced by the reform. For other reasonable values of η in the literature, ranging from 0.1 (Staiger et al., 2010) to 10.9 (Berger et al., 2022), women's welfare increases by between 2.8% and 9.5%.

[47]Table C.1.14 computes the change in welfare separately for workers of child-bearing ages (i.e., between 20 and 35 years old), and finds qualitatively very similar results to those unrestricted by age.

3.7 Conclusion

This study finds that one reason that workplaces do not provide job features valued by women is that decision-makers do not prioritize women's preferences. Studying a top-down change in Brazil that led Latin America's largest trade union federation, the *Central Única dos Trabalhadores*, to adopt a bargaining plan more attentive toward women's needs, we find a sharp increase in female-centric amenities, without corresponding tradeoffs in women or men's wages and employment, or in firm profits. The reform increases female-centric amenities on paper, such as those governing maternity leaves, job protection, childcare, and flexibility. These changes on paper translate into practice, with women taking longer maternity leaves and becoming more represented among managers. We find that women value these changes; they are less likely to separate from and more likely to queue for jobs at CUT establishments. Although the reform may have reduced male amenities, men do not exit more. Finally, we find no evidence that firm profits fall.

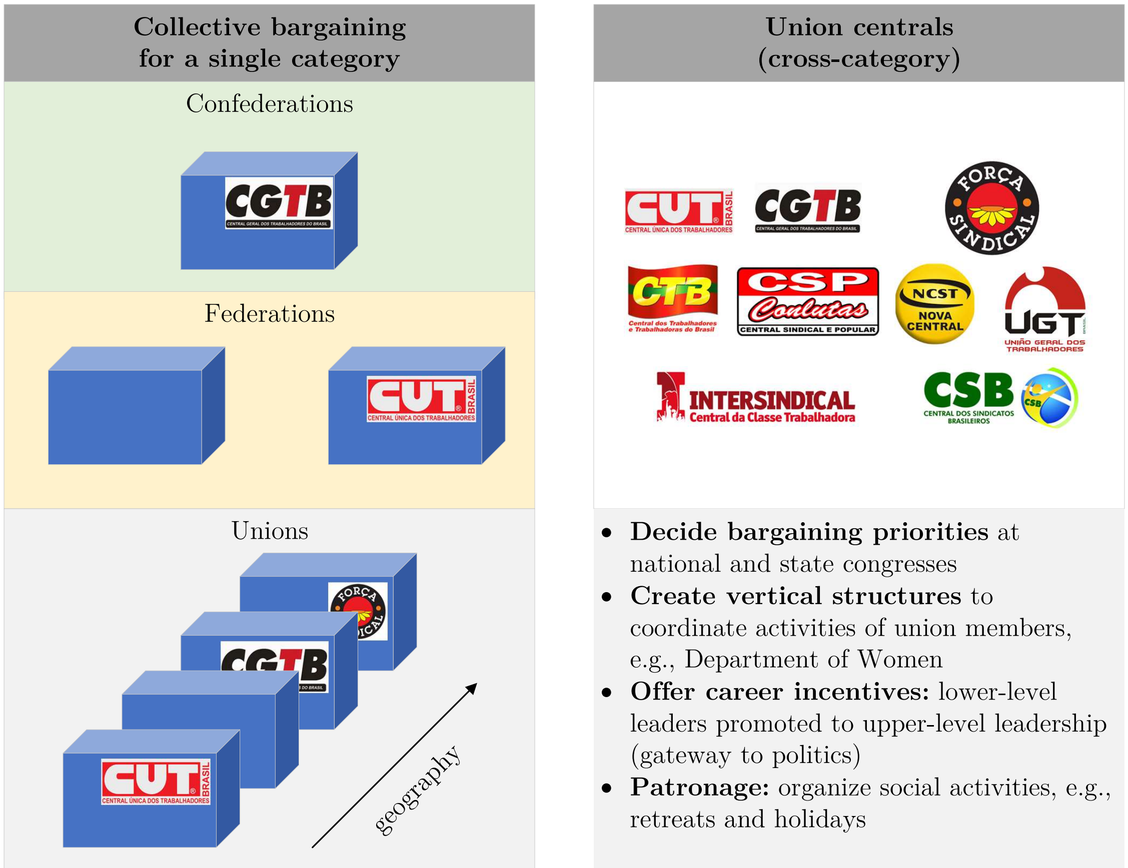
In sum, we provide causal evidence that union priorities importantly shape compensation, and, consequently, within-firm inequality. While gender gaps in virtually all labor market outcomes have narrowed at a fast pace in the last century, more recently reducing inequality has proven harder (Goldin, 2014; Blau and Kahn, 2017, 2006). Policies increasing women's representation in the workplace, for example via quotas on firm boards, have had null effects (Bertrand et al., 2018; Maida and Weber, 2020). By contrast, we find an important role for representing women's interests in collective bargaining.

In our setting, prioritizing women appears to usher in more efficient compensation for workers. One possible explanation for these findings is that the union was originally striking an inefficient bargain for workers. An alternative possibility is that the reform increased firm profits and the total size of rents split between unions and employers. Turnover is typically costly to the firm, and we find lower separations among women. Happier workers may also be more productive. Finally, the reform may have simply shifted worker rents from men to women, with no increase in male quits because men could not obtain similar rents elsewhere. This last explanation, while possible, is not supported by the evidence since we find a small increase in male retention and no observed changes in men's wages or amenities.

Our findings raise several new questions. First, given that leaders' priorities can influence compensation and inequality, how do these priorities emerge? A historical literature emphasizes the inherently political nature of labor unions, and argues that their objectives are ultimately shaped by their internal organization (Farber, 1986; Ross, 1950). In light of our findings, this hypothesis is especially fruitful to revisit empirically. Second, if leaders influence workplace conditions, might they also influence investments that affect worker productivity? Studying how productivity endogenously evolves as a consequence of leadership decisions is an exciting area for future research.

3.8 Figures

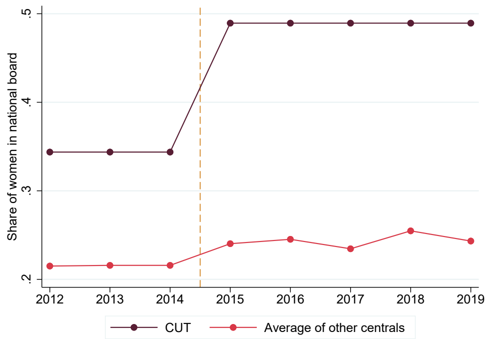
Figure 3-1: Workers' Bargaining Structure



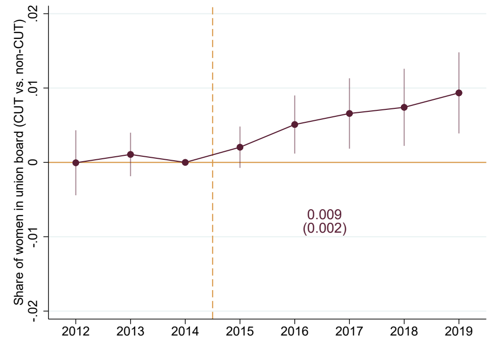
Notes: Figure depicts the organizations representing workers in collective bargaining (as blue blocks on the left panel) and the union centrals they can affiliate with (as logos on the right panel). All workers in a category-geography cell (e.g., bank workers in São Paulo) are represented by a single union. Unions can integrate geographically within the same category, forming a federation (at the state level) or a confederation (at the national level). Local unions, federations and confederations can affiliate with union centrals (*centrais sindicais*), which are depicted in the figure as union central logos "stamped" on the blue blocks. Union centrals are associations of unions, representing cross-category interests and operating on a nationwide level, with political objectives and coordination functions. Union centrals cannot directly participate in collective bargaining.

Figure 3-2: The 2015 CUT Reform

(a) Gender parity in national leadership



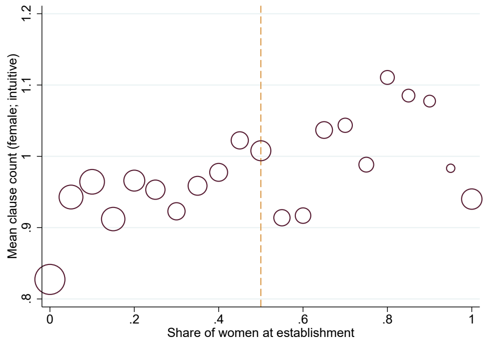
(b) Impact on local union boards



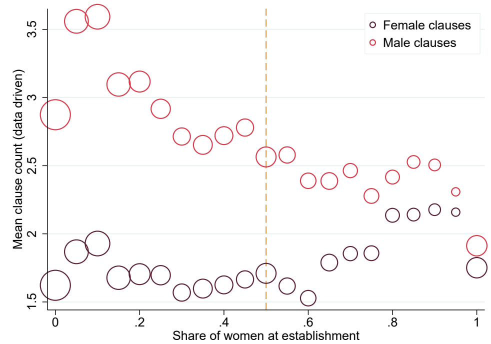
Notes: The 2015 CUT reform consisted of two parts. The first is a 50% quota for women in CUT's state and national executive bodies. The second is the adoption of a bargaining agenda more attentive to the needs of female workers. Figure 3-2a plots the annual share of women on CUT's national executive committee and the average share in the other 7 union centrals (*Intersindical* is dropped due to missing information on its board). Refer to Figure C.1.1 for the plots corresponding to each individual union central. Figure 3-2b shows how the reform had downstream effects on the gender composition of local union boards (for CUT affiliates relative to non-CUT affiliates as of 2012). The figure depicts the estimated coefficients for the interactions between a CUT affiliate dummy and year fixed effects, where the regression's dependent variable is the share of women in the board for a given union-year observation. The event-study specification omits the baseline year 2014 and includes both union fixed effects and year fixed effects. Note that the average share of women across CUT affiliates unions in 2014 is around 33%. Confidence intervals at a 95% level are reported. Standard errors are clustered by union.

Figure 3-3: Sense Checks for Female- and Male-Centric Amenities

(a) Intuitive female clauses and share of women



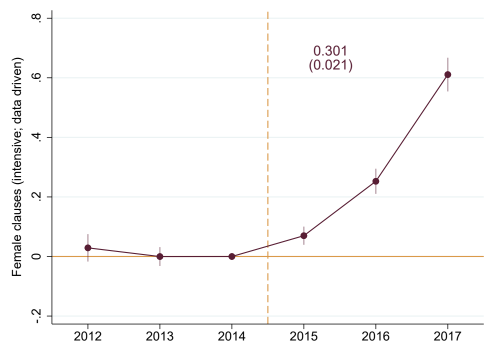
(b) Data-driven clauses and share of women



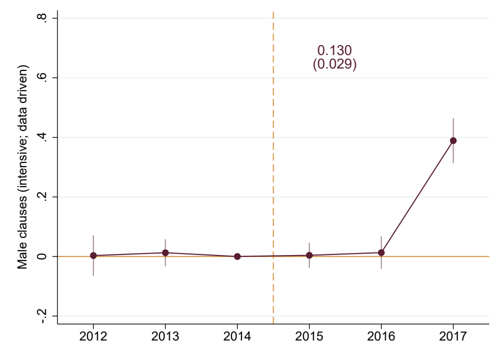
Notes: Figures depict binned scatterplots of the number of female-centric (and male-centric) clauses contained in firm-level CBAs signed at baseline (2014) by the share of women in the workforce of the establishment. The bins in the bottom figures are set to rounded values (in 0.05 increments) of the share of women at the establishment, with the size of the markers scaled to represent the number of pairs observed in a given bin. Figure 3-3a uses the intuitive definition of female-centric amenities, while Figure 3-3b uses the data-driven approach for both female- and male-centric amenities. The vertical line indicates 50% of women in the workforce. The sample consists of the establishments in our new contracts panel at baseline (2014). Regressing the y-axis variables in the bottom figures on the share of women at establishments reveals a positive (negative) and statistical significant relation between female (male) centric clauses and the share of women at the establishment. For the intuitive definition of female-centric clauses, the slope is 0.137 (SE 0.019). For the data-driven definition of female-centric clauses, the slope is 0.172 (SE 0.034). For the data-driven definition of male-centric clauses, the slope is -1.219 (SE 0.042).

Figure 3-4: Effect of the CUT Reform on Female- and Male-Centric Amenities

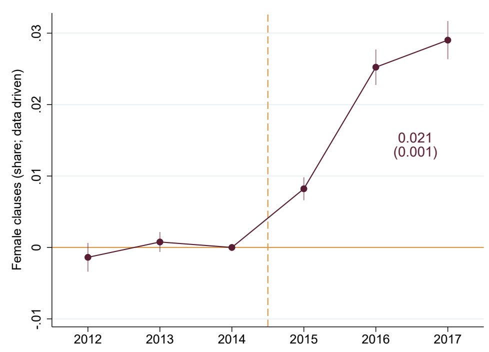
(a) Female clauses: intensive margin



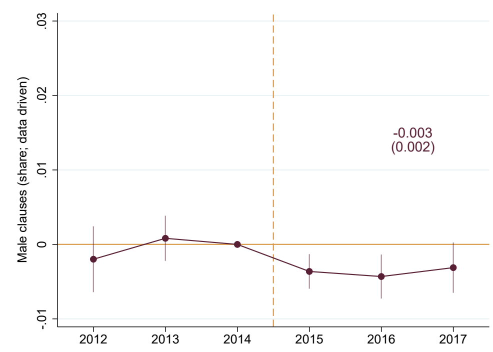
(b) Male clauses: intensive margin



(c) Female clauses: as a share of clauses

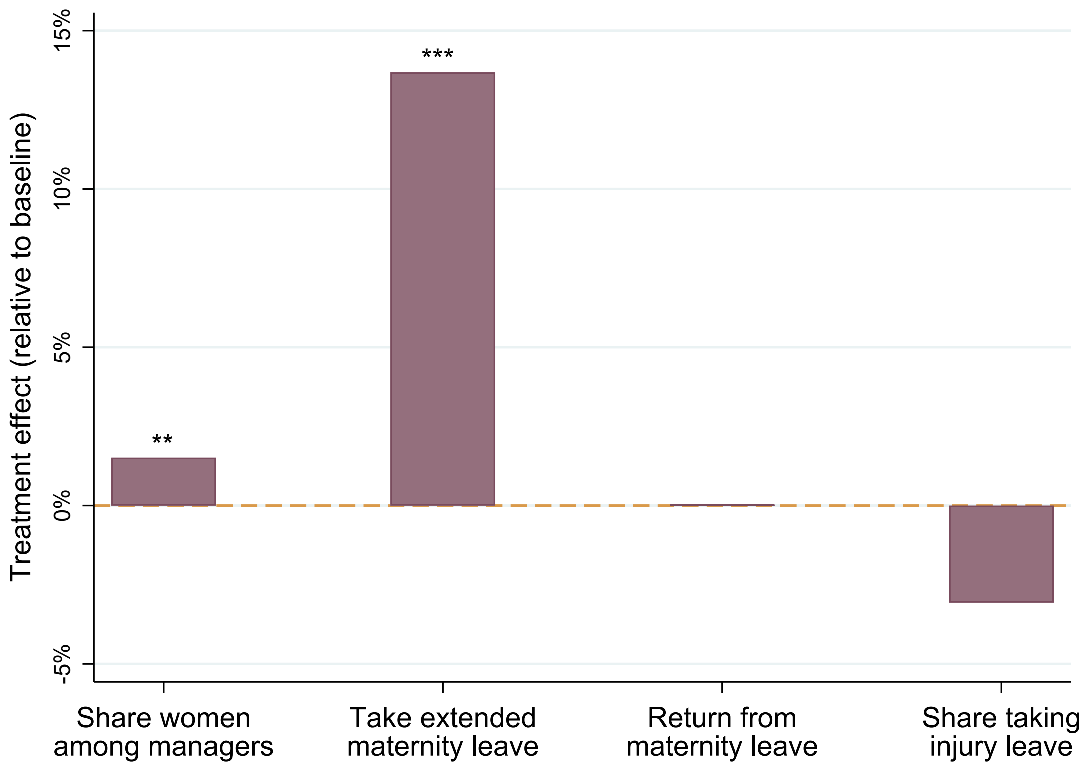


(d) Male clauses: as a share of clauses



Notes: Figures show estimates of the δ_t coefficients for $t \in [2012, 2017]$ (with 2014 omitted) from the DID specification in Equation (3.3) on the intensive margin (top figures) and shares (bottom figures) of female-centric (left side) and male-centric (right side) clauses, defined using the data-driven method. All figures use the filled panel. Confidence intervals at a 95% level are reported. Standard errors are clustered at the establishment level.

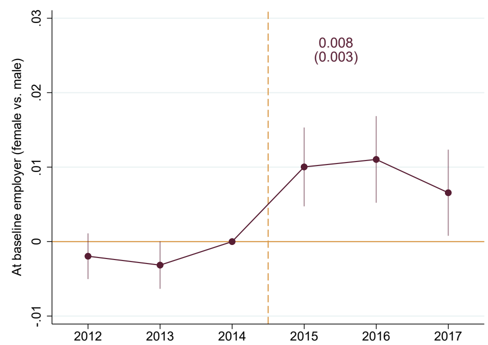
Figure 3-5: Changes in Firm Environment



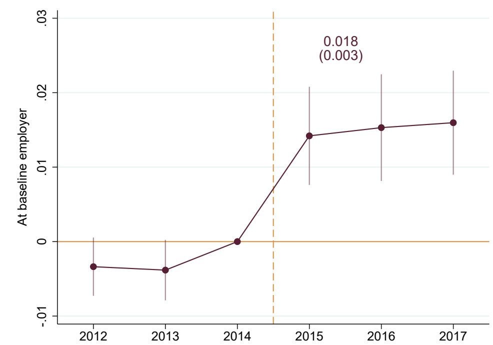
Notes: Figure reports results from four separate establishment-level DID regressions in Equation (3.3), with treatment effects reported relative to the mean among the treated at baseline (in percentage terms). The outcome variables are: 1) the share of women among managers; 2) the share of women on maternity leave who remain on leave longer than than the state-mandated 120 days (i.e., extended maternity leave); 3) the share of women taking maternity leave who remain employed at the employer where they took maternity leave (i.e., return from maternity leave); and 4) the share of workers taking leave due to a workplace injury. Each regression includes establishment fixed effects, industry-year fixed effects, and microregion-year fixed-effects. Two stars indicate significance at the 5% confidence level, while three starts significance at the 1% level. Standard errors are clustered by establishment.

Figure 3-6: Revealed Preference Measures of Firm Value

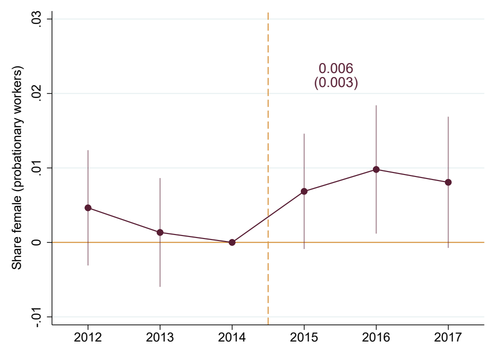
(a) Incumbent retention: women-men differential



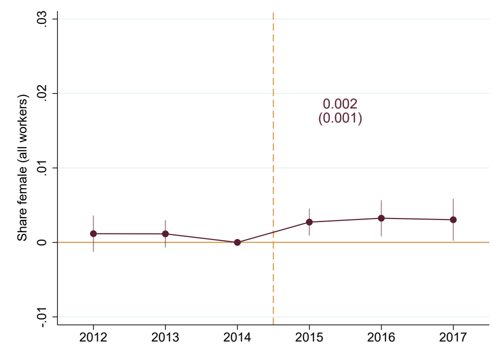
(b) Incumbent women's retention: age 20-35



(c) Share of women among probationary workers



(d) Share of women in workforce



Notes: Figures test for revealed preference measures of whether women value the changes induced by the CUT reform in treated establishments. Top figures look at retention among incumbent workers, i.e., an indicator for whether the worker is observed at their baseline (2014) employer in year t. To make treatment effects in worker-level regressions interpretable as establishment-level averages, we weight each incumbent worker by the inverse of (own-gender) employment at their baseline employer. Figure 3-6a reports the differential in retention for women relative to men using a triple DID regression, which includes worker fixed effects, industry-year-gender fixed effects, microregion-year-gender fixed effects, and tenure-year-gender fixed effects. Figure 3-6b shows effects from the baseline DID specification in Equation (3.3) among women ages 20-35, which includes worker fixed effects, industry-year fixed effects, microregion-year fixed effects, and tenure-year fixed effects. Bottom figures look at the gender composition of spells observed at the establishment level using the DID specification in Equation (3.3). The outcome in Figure 3-6c is the share of women among probationary workers, i.e., those whose tenure at the establishment does not exceed 3 months. The outcome in Figure 3-6d is the share of women among all spells observed. Regressions include establishment fixed effects, industry-year fixed effects, and microregion-year fixed effects. Confidence intervals at a 95% level are reported. Standard errors are clustered at the establishment level.

3.9 Tables

Table 3.1: Sample Descriptives

	All	Treated	Control
	(1)	(2)	(3)
Panel A: Sample characteristics			
Collective bargaining agreements	211,619	42,523	169,096
Establishment-union pairs	89,920	19,040	70,880
Signing establishments	80,131	18,103	62,028
Signing unions	4,409	886	3,523
Avg. years of CBA negotiation (per pair)	2.35	2.23	2.39
Panel B: CBA negotiation characteristics			
Avg. clause count	24.7	23.1	25.1
Avg. female clause count (intuitive)	1.66	1.81	1.63
Avg. female clause count (data-driven)	3.16	3.15	3.16
Avg. male clause count (data-driven)	4.87	4.59	4.94
Panel C: Establishment-level characteristics (2014, baseline)			
Avg. employment	143	198	127
Avg. share of women in workforce	0.38	0.36	0.38
Share employing both men and women	0.82	0.83	0.82
Share of single establishment firms	0.64	0.63	0.64
Panel D: Union-level characteristics (2014, baseline)			
Avg. size of union board	18.8	24.3	17.3
Avg. share of women in board	0.23	0.23	0.22
Share with female president or vice president	0.17	0.18	0.17

Notes: Table shows descriptive statistics for the sample of establishment-union pairs negotiating firm-level CBAs registered in *Sistema Mediador* between 2012 and 2017. All CBAs are valid, non-amendment, firm-level agreements that have a union counterpart with information on 2012 union central affiliation. We additionally drop contracts signed by more than one union if these unions have different CUT affiliation in 2012 (fewer than 0.33% of CBAs). On the signing establishment's side, we restrict to CBAs where the employer appears in RAIS and has active employees in 2014. Treated units are those where the union counterpart was affiliated to CUT in 2012. See Appendix C.2 for more details. The starting sample described in Panel A has observations at the pair-year level for years when CBA negotiations occurred, i.e., the new contracts panel. Statistics in Panel B are averages across these pair-year observations. Panels C and D use unique establishment and union observations in the baseline year (2014), respectively.

Table 3.2: Female- and Male-Centric Amenities

Intuitive definition	**Data-driven definition**		
Female clauses	Top 20 female clauses	Top 20 male clauses	Rank
Abortion leave	Childcare assistance	On-call pay	1
Abortion protections	Absences	Life insurance	2
Adoption leave	Adoption leave	Strike procedures	3
Childcare assistance	Other: holidays and leaves	Other: protections for injured workers	4
Equal opportunities	Seniority pay	Profit sharing	5
Female workforce	Maternity protections	Salary deductions	6
Maternity assistance	Abortion protections	Female workforce	7
Maternity leave	Paid leave	Transfers	8
Maternity protections	Night pay	Machine and equipment maintenance	9
On-call	Nonwork-related injury protections	Duration and schedule	10
Other: holidays and leaves	Abortion leave	Working environment conditions	11
Paid leave	Policy for dependents	Salary payment - means and timeframes	12
Part-time contracts	Extension/reduction of workday	Hazard pay (danger risk)	13
Paternity protections	Guarantees to union officers	Safety equipment	14
Policy for dependents	Renewal/termination of the CBA	CIPA: accident prevention committee	15
Sexual harassment	Medical exams	Other assistances	16
Special shifts	Unionization campaigns	Death/funeral assistance	17
Uninterrupted shifts	Health education campaigns	Workday compensation	18
Unpaid leave	Waiving union fees	Collective vacations	19
Workday controls	Salary adjustments/corrections	Tools and equipment	20

Notes: Table lists the clause types that were selected as "female-centric" based on intuition (column 1) and with our data-driven approach (column 2), which also allows us to define "male-centric" clauses (column 3)—refer to Section 3.3.2 for details on the data-driven approach. The clauses in column 1 are listed in alphabetical order while those selected with the data-driven approach are ranked on the basis of the coefficients β_z coming from the estimation of Equation (3.2). That is, the first female clause listed is the one with the highest estimate of β_z, the second is the one with the second highest value of β_z, etc. Similarly, the male clauses are ranked from the one with the lowest estimate of β_z to the one with the 20th lowest estimate. In columns 2 and 3, we highlight in red the clauses that also belong to the intuitive definition of female-centric clauses.

Table 3.3: Effect of CUT Reform on Negotiated Amenities

	Intuitive definition (female clauses)					Data-driven		
	All	Leave	Maternity	Harassment	Flexibility	Female	Male	F/(F+M+1)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Intensive margin (number)								
$D_i \times \delta_{year \geq 2015}$	0.157***	0.078***	0.042***	0.009***	0.028***	0.301***	0.130***	0.030***
	(0.013)	(0.006)	(0.004)	(0.001)	(0.008)	(0.021)	(0.029)	(0.002)
Mean outcome	0.95	0.25	0.24	0.02	0.44	1.58	2.55	0.20
Panel B: Intensive margin (sum of unique clause types)								
$D_i \times \delta_{year \geq 2015}$	0.123***	0.047***	0.042***	0.008***	0.027***	0.154***	0.067***	
	(0.010)	(0.004)	(0.004)	(0.001)	(0.004)	(0.014)	(0.017)	
Mean outcome	0.70	0.18	0.21	0.02	0.30	1.26	1.58	
Panel C: Extensive margin								
$D_i \times \delta_{year \geq 2015}$	0.017***	0.012***	0.020***	0.008***	0.022***	0.034***	-0.001	
	(0.003)	(0.002)	(0.002)	(0.001)	(0.003)	(0.003)	(0.003)	
Mean outcome	0.31	0.12	0.15	0.02	0.23	0.36	0.46	
Panel D: As a share of all clauses								
$D_i \times \delta_{year \geq 2015}$	0.005***	0.001***	0.001***	0.000***	0.003***	0.021***	-0.003**	
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)	
Mean outcome	0.05	0.01	0.01	0.00	0.03	0.07	0.14	
Observations	600,960	600,960	600,960	600,960	600,960	600,960	600,960	600,960

Notes: Table reports the coefficients for DID regressions—see Equation (3.3)—estimating the effect of the CUT reform on the female-centric and male-centric amenities included in CBAs. The unit of observation is a union-employer pair. Panel A reports effects on the total number of clauses, an intensive margin measure of amenities. Panel B reports effects on the sum of unique clause types in the corresponding categories exist in a contract, capturing changes to the *space* of female (male) clauses (as opposed to their number). For example, two anti-harassment clauses will raise the outcome value by two in Panel A of Column 6 but by one in Panel B. Panel C reports effects on a cumulative indicator for whether any clause of the corresponding type exists in a contract as an extensive margin measure of amenities. Panel D uses the share of clauses among all clauses in a contract. Under each panel we report the mean of the dependent variable among the treated at baseline (2014). The sample is the filled panel of establishment-union pairs by year. All columns control for pair fixed effects, as well as time-varying state and industry fixed effects. Standard errors are clustered at the establishment level.

Table 3.4: Heterogeneity by Baseline Female Representation

		Full interaction: $D_i \times \delta_{year \geq 2015} \times H_i$		
	Baseline	H_i = low % women in estab.	H_i = low % women in union	H_i = no woman Pres/VP
	(1)	(2)	(3)	(4)
Panel A: Intensive margin				
$D_i \times \delta_{year \geq 2015}$	0.301***	0.139***	0.002	-0.058
	(0.021)	(0.028)	(0.038)	(0.044)
$D_i \times \delta_{year \geq 2015} \times H_i$		0.307***	0.362***	0.396***
		(0.040)	(0.041)	(0.049)
Sum of coefficients		0.446	0.364	0.338
p-value		[0.000]	[0.000]	[0.000]
Mean outcome	1.58	1.58	1.58	1.58
Panel B: As a share of all clauses				
$D_i \times \delta_{year \geq 2015}$	0.021***	0.009***	0.005***	-0.004**
	(0.001)	(0.001)	(0.002)	(0.002)
$D_i \times \delta_{year \geq 2015} \times H_i$		0.022***	0.020***	0.030***
		(0.002)	(0.002)	(0.002)
Sum of coefficients		0.031	0.025	0.025
p-value		[0.000]	[0.000]	[0.000]
Mean outcome	0.07	0.07	0.07	0.07
Observations	600,960	600,960	592,344	592,344

Notes: Table tests for heterogeneity in the effect of the CUT reform on female-centric clauses (data-driven approach) according to the baseline representation of women among workers (column 2) and within union boards (columns 3-4). The dummy to test for heterogeneity in the effects (H_i) is fully interacted with the treatment dummy (D_i) and the post-period dummy ($\delta_{year \geq 2015}$). The table only reports the coefficients on the effects that determine the treatment effect for the baseline group ($H_i = 0$) and the differential effect relative to the baseline group—with the sum of both coefficients representing the treatment effect for the group of interest ($H_i = 1$). In column (2), H_i is an indicator for whether the share of women workers is below the median across our sample in 2014 (around 1/3). In column (3), H_i is an indicator for whether the share of women in union boards is below this 1/3 threshold in 2014. In column (4), H_i is an indicator for whether there is no women president of vice-president in the local union board as of 2014. All regressions use the filled panel sample and includes establishment-union pair fixed effects as well as time-varying state and industry fixed effects. Standard errors are clustered at the establishment level.

Table 3.5: Impact of CUT Reform on Establishment-Level Outcomes

Panel A: Wages	Mean log(w) [women; $t > 12$] (1b)	Mean log(w) [men; $t > 12$] (2b)	Mean log(w) [women; $t \leq 12$] (3b)	Mean log(w) [men; $t \leq 12$] (4b)	Mean gender wage gap (5b)	CBA wage adjustments (6b)
$D_i \times \delta_{year \geq 2015}$	-0.004	-0.003	-0.005	-0.006*	-0.001	0.032
	(0.002)	(0.002)	(0.004)	(0.003)	(0.002)	(0.021)
Mean outcome	7.460	7.627	7.174	7.311	-0.150	0.781
Observations	323,271	329,960	260,956	289,334	334,562	123,432

Panel B: Employment	Log employment (1a)	Share women [workforce] (2a)	Share women [probation] (3a)	Log hires (4a)	Share women [hires] (5a)	Share women [separations] (6a)
$D_i \times \delta_{year \geq 2015}$	-0.002	0.002**	0.006**	-0.009	0.004*	0.004**
	(0.007)	(0.001)	(0.003)	(0.009)	(0.002)	(0.002)
Mean outcome	4.044	0.369	0.357	3.034	0.366	0.360
Observations	353,626	353,626	275,879	325,823	325,823	332,506

Panel C: Profits	Log wage bill (1c)	Establishment exit (2c)	Profit margin (3c)
$D_i \times \delta_{year \geq 2015}$	-0.010	-0.003	0.702
	(0.008)	(0.003)	(1.167)
Mean outcome	11.431	0.087	7.759
Observations	351,593	61,716	2,874

Notes: Table reports the coefficients for the establishment-level DID regression from Equation (3.3), comparing treated to comparison establishments on wage, employment, and profit outcomes. An establishment is treated if the union with which it negotiates is affiliated to CUT in 2012. Each regression includes establishment fixed effects, industry-year fixed effects, and microregion-year fixed effects. Panel A uses workers' main spell in a given year. The terms in brackets indicate the subsample among which the mean of log wages is calculates, i.e., tenure > 12 months and tenure ≤ 12 months for either women or men. Panel B uses all spells observed at an establishment in a given year. The terms in brackets indicate the subsample among which the share of women is calculated, i.e., among all workers, among workers in probation, among hires, and among separated workers. Panel C studies three imperfect measures of firm profits. Standard errors are clustered by establishment and reported in parentheses.

Table 3.6: Welfare Estimation

		Women (1)	Men (2)
$ln\phi_{t-1,t}$		0.059	0.013
		(0.007)	(0.005)
Contribution by component:			
Wage bill	$ln(\lambda_{t,t-1}) - ln(\lambda_{t-1,t})$	15%	22%
Dispersion	$ln(\bar{S}^*_t) - ln(\bar{S}^*_{t-1})$	48%	-4%
Wages	$ln(\bar{w}^*_t) - ln(\bar{w}^*_{t-1})$	37%	82%
η (calibrated)	1.015		
N establishments		60,651	60,651
N establishments in $\Omega_{t,t-1}$		47,195	47,195

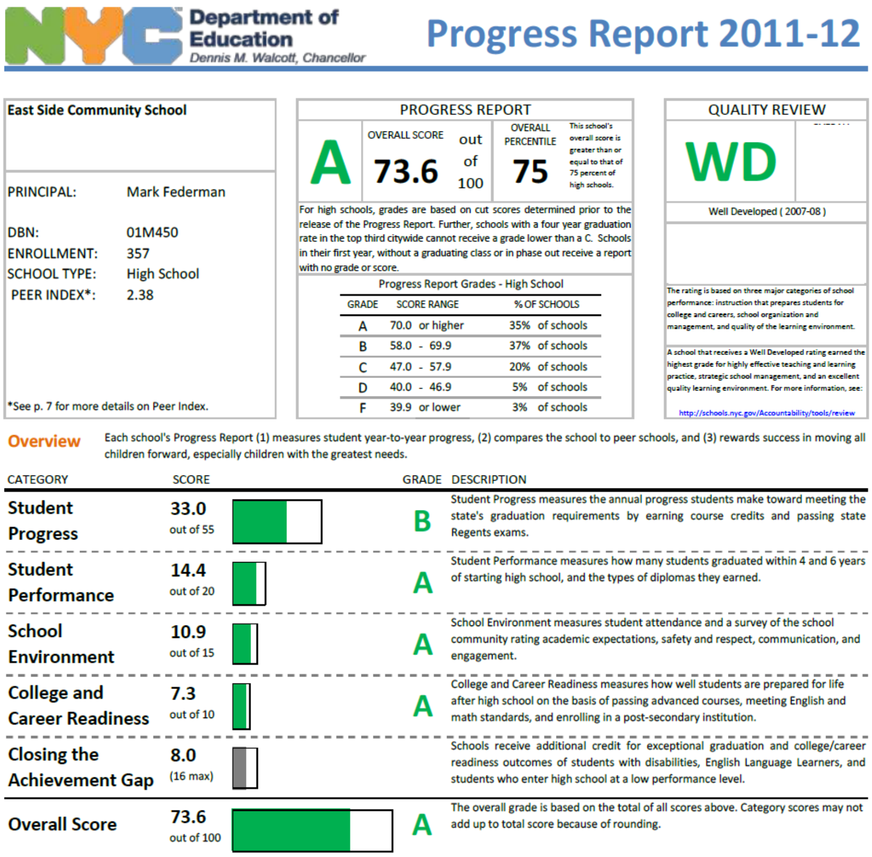
Notes: Table reports the estimated welfare change for men and women. It also reports the contribution to the overall effect by each of the three components that make the welfare index, namely the Feenstra “new varieties" term $ln(\lambda_{t,t-1}) - ln(\lambda_{t-1,t})$, the change in the geometric average of the labor income shares of non-CUT firms $ln(\bar{S}^*_t) - ln(\bar{S}^*_{t-1})$, and the change in the geometric average of the wages of non-CUT firms $ln(\bar{w}^*_t) - ln(\bar{w}^*_{t-1})$. Standard errors in parenthesis come from the bootstrap procedure described in Appendix C.4.

Appendix A

Appendix to Chapter 1

A.1 Appendix Tables and Figures

Figure A.1.1: Example of An Online School Quality Report With Letter Grades



Department of Education
Dennis M. Walcott, Chancellor

Progress Report 2011-12

East Side Community School

PRINCIPAL: Mark Federman

DBN: 01M450
ENROLLMENT: 357
SCHOOL TYPE: High School
PEER INDEX*: 2.38

*See p. 7 for more details on Peer Index.

PROGRESS REPORT

A | OVERALL SCORE 73.6 out of 100 | OVERALL PERCENTILE 75 | This school's overall score is greater than or equal to that of 75 percent of high schools.

For high schools, grades are based on cut scores determined prior to the release of the Progress Report. Further, schools with a four year graduation rate in the top third citywide cannot receive a grade lower than a C. Schools in their first year, without a graduating class or in phase out receive a report with no grade or score.

Progress Report Grades - High School

GRADE	SCORE RANGE	% OF SCHOOLS
A	70.0 or higher	35% of schools
B	58.0 - 69.9	37% of schools
C	47.0 - 57.9	20% of schools
D	40.0 - 46.9	5% of schools
F	39.9 or lower	3% of schools

QUALITY REVIEW

WD

Well Developed (2007-08)

The rating is based on three major categories of school performance: instruction that prepares students for college and careers, school organization and management, and quality of the learning environment.

A school that receives a Well Developed rating earned the highest grade for highly effective teaching and learning practice, strategic school management, and an excellent quality learning environment. For more information, see: http://schools.nyc.gov/Accountability/tools/review

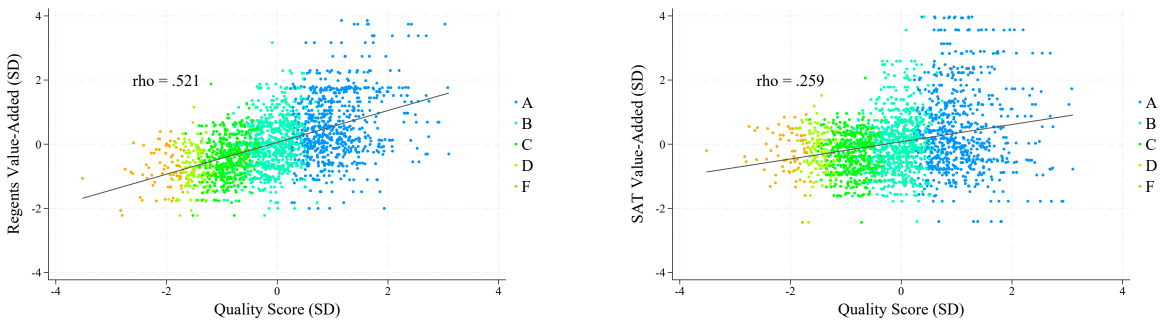
Overview Each school's Progress Report (1) measures student year-to-year progress, (2) compares the school to peer schools, and (3) rewards success in moving all children forward, especially children with the greatest needs.

CATEGORY	SCORE	GRADE	DESCRIPTION
Student Progress	33.0 out of 55	B	Student Progress measures the annual progress students make toward meeting the state's graduation requirements by earning course credits and passing state Regents exams.
Student Performance	14.4 out of 20	A	Student Performance measures how many students graduated within 4 and 6 years of starting high school, and the types of diplomas they earned.
School Environment	10.9 out of 15	A	School Environment measures student attendance and a survey of the school community rating academic expectations, safety and respect, communication, and engagement.
College and Career Readiness	7.3 out of 10	A	College and Career Readiness measures how well students are prepared for life after high school on the basis of passing advanced courses, meeting English and math standards, and enrolling in a post-secondary institution.
Closing the Achievement Gap	8.0 (16 max)		Schools receive additional credit for exceptional graduation and college/career readiness outcomes of students with disabilities, English Language Learners, and students who enter high school at a low performance level.
Overall Score	73.6 out of 100	A	The overall grade is based on the total of all scores above. Category scores may not add up to total score because of rounding.

Notes: This figure shows the 2011/12 progress report for East Side Community School as an example of how a school progress report looked like. Source: www.crpe.org

Figure A.1.2: Correlation Between School VA and the Bloomberg Quality Score

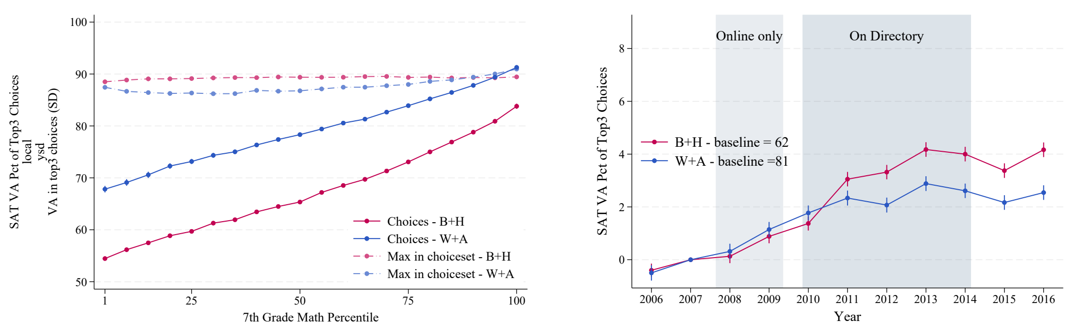
(a) Regents VA (b) SAT VA



Notes: This figure shows scatter plots of the quality score used in progress reports (x-axis) against OLS measures of Regents VA (panel a)) and SAT VA (panel b)), together with the corresponding correlation coefficient. Each dot is a school-year. Different colors indicate the letter grade received by each observation.

Figure A.1.3: The Racial School Quality Choice Gap - SAT VA

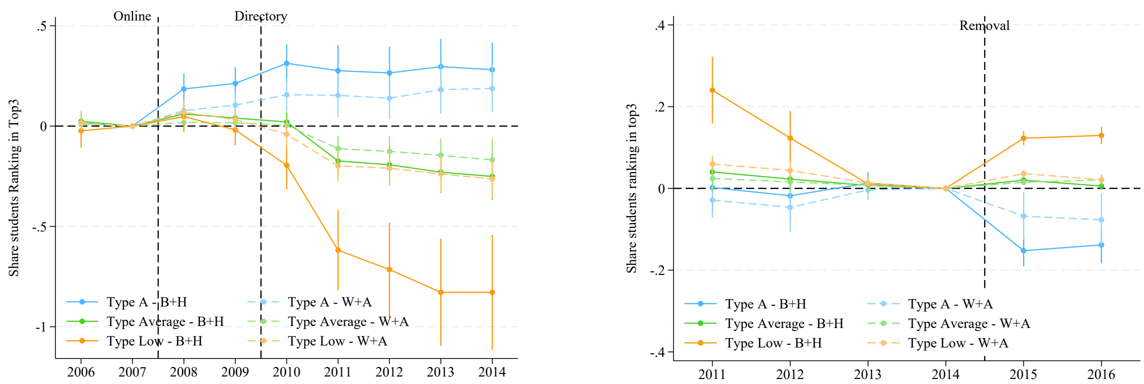
(a) By Baseline Achievement (b) Across Cohorts



Notes: This figure describes cross-race differences in chosen school quality, as measured by SAT value-added. It is analogous to Figure 1-1, but uses SAT value-added to measure school quality, rather than Regents value-added.

Figure A.1.4: Event Study Estimates - Separate Regressions by Student Race

(a) Introduction of Letters (b) Removal of Letters



Notes: The figure plots event study estimates of the coefficient β_L^t of equation (1.3), from separate regressions by race. Panel (a) considers changes relative to 2007, the year before the introduction of letters, using applicant cohorts of 2006-2014. Panel (b) considers changes around the removal of letters, normalizing share differences to 0 in 2014, and using cohorts of 2011-2016. Blue lines are for changes in shares for Type A schools, orange ones for shares of Type Low schools and the green ones for Type Average schools. Dashed lighter lines are for changes in shares of white and Asian students, solid ones for changes in shares of Black and Hispanic students.

Table A.1.1: Changes in Information Provided by the DOE

Year (fall 9th grade)	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Letter grades			○	○	✓	✓	✓	✓	✓		
Letter grade subcategories			○	○	○	○	○	○	✓		
Graduation %	✓	✓	○	○	○	✓	✓	✓	✓	✓	✓
College %	○	○	○	○	○	○	○	✓	✓	✓	✓
Regents performance	✓	✓	○	○	○	○	○	○	○	○	○
State quality review			○	○	✓	✓	✓	✓	✓	✓	✓
Quality measures - de Blasio											○
Survey-based measures										✓	✓
(feel safe, satisfaction, variety of classes)											

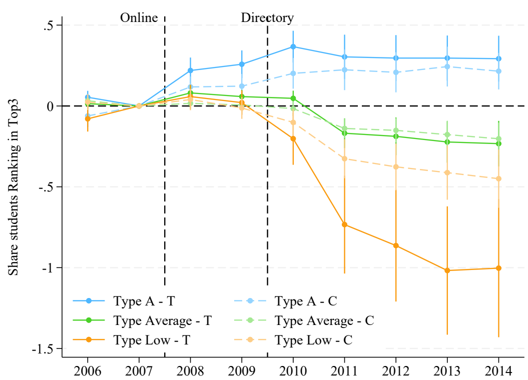
[1] ✓- information provided on the school directory (and online)
[2] ○- information provided online only

Notes: This table summarizes which type of information about school performance was shown on the printed high school directory and online (denoted with ✓) and which was only shown online (denoted with ○). Years denote applicant cohorts and refer to the fall of their enrollment in 9th grade. Information is distributed (and applicants apply) in the preceding year.

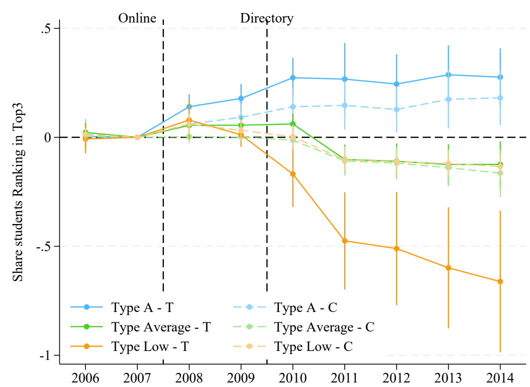
Figure A.1.5: Demand Responses to Introduction and Removal of Quality Signals - Heterogeneity by Exposure to New Information

Introduction of Letters

(a) Minority Applicants

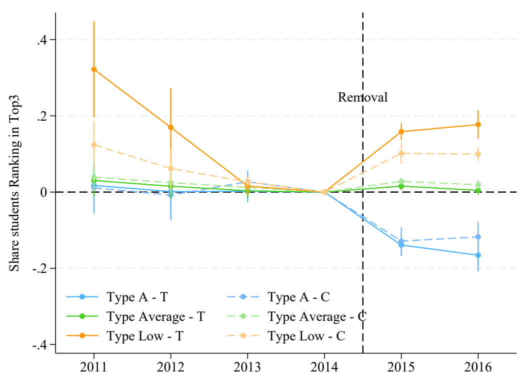


(b) Non-Minority Applicants

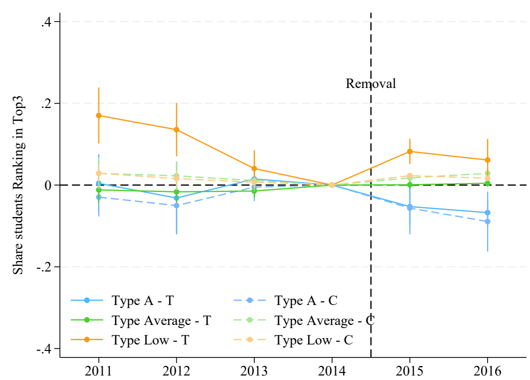


Removal of Letters

(c) Minority applicants



(d) Non-Minority applicants

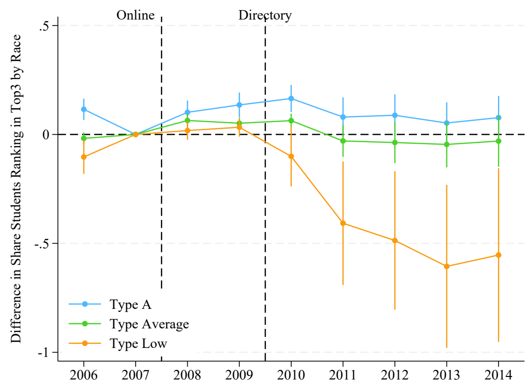


Notes: The figure plots event study estimates of the coefficient β_L^t of equation (1.3), from separate regressions by race and values of the dummy $Treated_i$. Panels (a) and (b) consider changes relative to 2007, the year before the introduction of letters, using applicant cohorts of 2006-2014. Panels (c) and (d) changes around the removal of letters, normalizing share differences to 0 in 2014, and using cohorts of 2011-2016. Blue lines are for changes in shares for Type A schools, orange ones for shares of Type Low schools and the green ones for Type Average schools. Dashed lighter lines are for changes in choice shares of students for whom $Treated_i = 0$, solid ones for choice shares among students with $Treated_i = 1$.

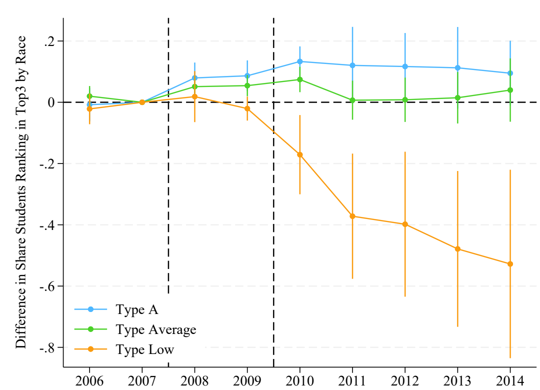
Figure A.1.6: Event Study Estimates of Demand Responses to Introduction and Removal of Quality Signals - Heterogeneity by Exposure to New Information

Introduction of Letters

(a) Minority Applicants

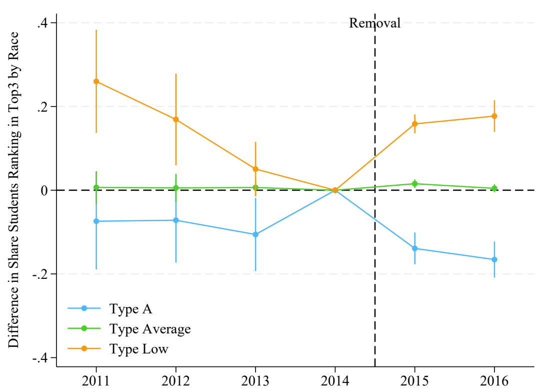


(b) Non-Minority Applicants

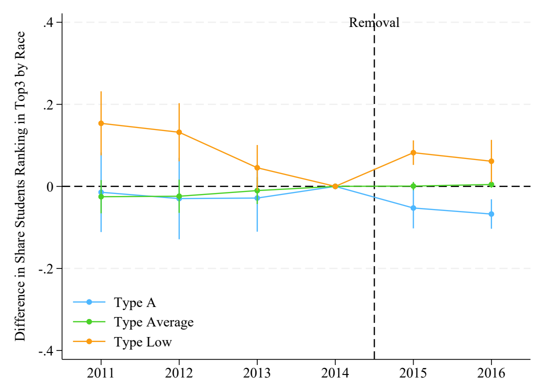


Removal of Letters

(c) Minority Applicants



(d) Non-Minority Applicants

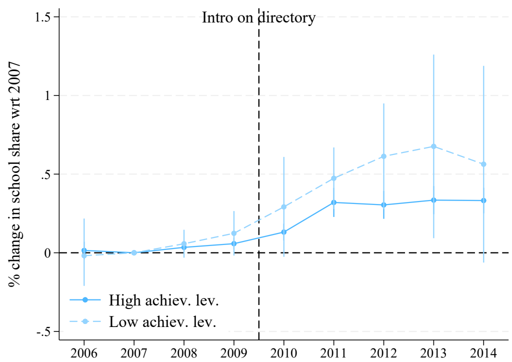


Notes: The figure plots event study estimates of the coefficient δ_L^t of a variant of equation (1.5) that considers differences in choice responses to the introduction and removal of letter grades along values of the dummy $Treated_i$ (rather than across race). Panels (a) and (b) consider differential changes relative to 2007, using applicant cohorts of 2006-2014, separately for minority and white students. Panels (c) and (d) consider changes around the removal of letters, normalizing share differences to 0 in 2014, and using cohorts of 2011-2016. Blue lines are for changes in shares for Type A schools, orange lines for shares of Type Low schools and the green lines are for Type Average schools.

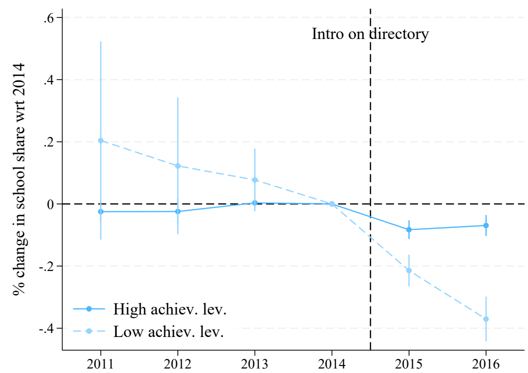
Figure A.1.7: Heterogeneity in Responses to Introduction and Removal of Quality Signals by School Peer Quality

Introduction and Removal of Positive Quality Signals

(a)

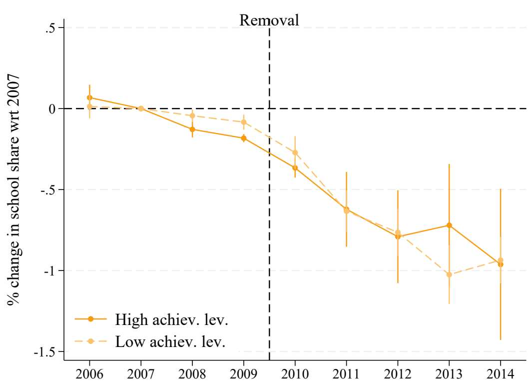


(b)

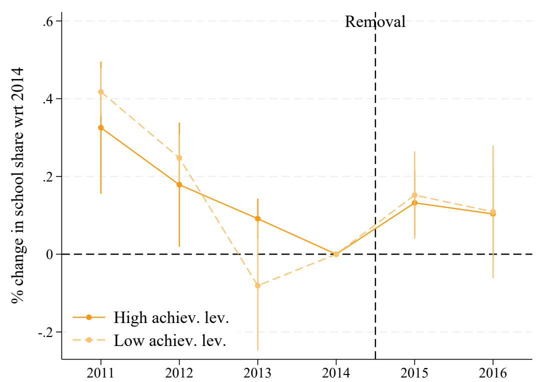


Introduction and Removal of Negative Quality Signals

(c)



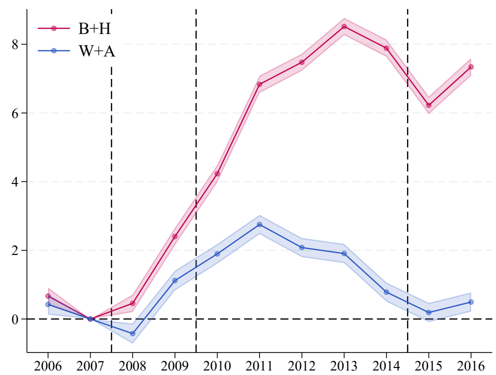
(d)



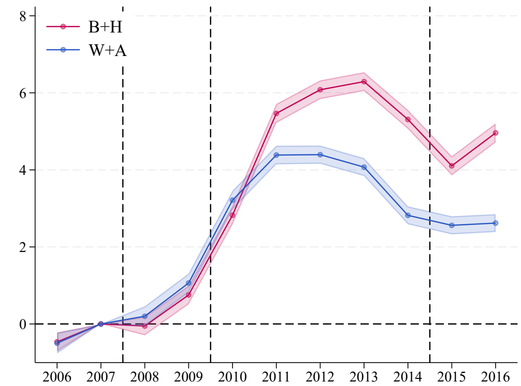
Notes: The figure plots event study estimates of the coefficient β_L^t of equation (1.3), from separate regressions for schools of different types and different peer quality. Panels (a) and (c) consider share changes relative to 2007, respectively for Type A and Type Low schools, using applicant cohorts of 2006-2014. Panels (b) and (d) share changes around the removal of letters, normalizing shares to 0 in 2014, and using cohorts of 2011-2016. Blue lines are for changes in shares of Type A schools, orange ones for shares of Type low schools. Dashed lighter lines are for changes in choice shares of schools enrolling lower achieving students, solid ones for schools enrolling higher achieving students.

Figure A.1.8: Evolution of Ranked School Characteristics by Race

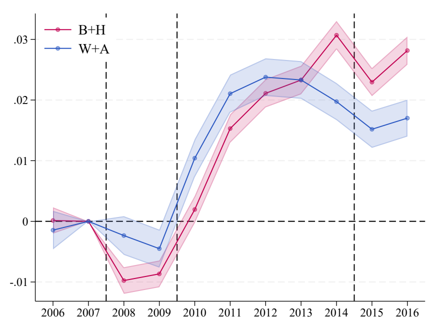
(a) Math VA (pct)



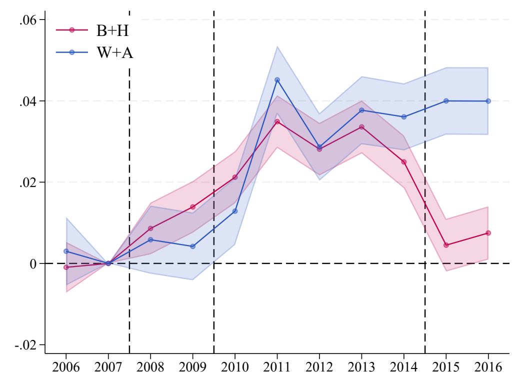
(b) Peer Quality (pct)



(c) White share

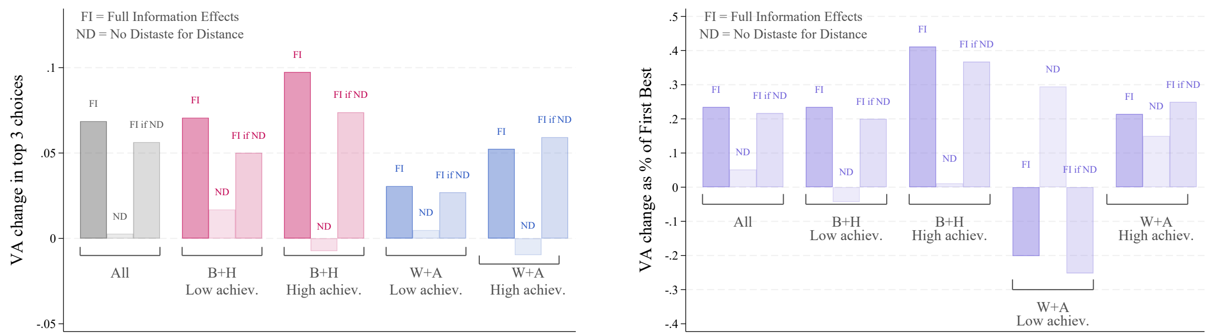


(d) I(applies out of borough)



Notes: The figure plots regression estimates of changes in average characteristics of applicants' first three high school choices over time with respect to 2007, by applicant race. Blue lines are for choices of white and Asian students, pink lines for choices of Black and Hispanic students. The lines plot coefficients of year dummies, normalizing the 2007 coefficient to 0. Controls include gender, ell status, subsidized lunch status and fixed effects for combinations of student borough and baseline test score terciles. Panel (a) shows trends in Regents value-added and panel (b) in peer quality of a student's first three choices, panel (c) in the share of white students enrolled in the student's first three school choices and panel (d) in the probability of applying to a school outside one's borough.

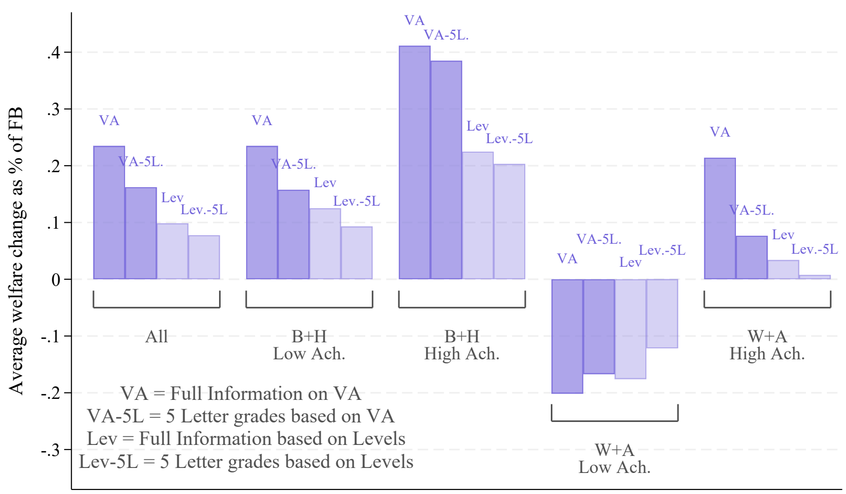
Figure A.1.9: Role of Distance in Explaining Choice Gaps and Effects of Information



(a) Average ΔVA in Choices (b) Welfare Gains

Notes: Panel (a) shows changes in VA of top 3 choices across three different simulations, taking averages within student groups defined by race and baseline achievement (above or below median). Panel (b) does the same thing for the resulting change in offered VA, expressed as a percentage of the average first-best achievement gains. Within each students subgroup, the first bar corresponds to the full-information benchmark that uses the real model estimates. The second bar corresponds to differences between the status-quo and a simulation in which students do not have distaste for commuting but the information environment is as in the status-quo. The third bar simulates changes with respect to the status-quo of providing full information if students do not have a distaste for commuting.

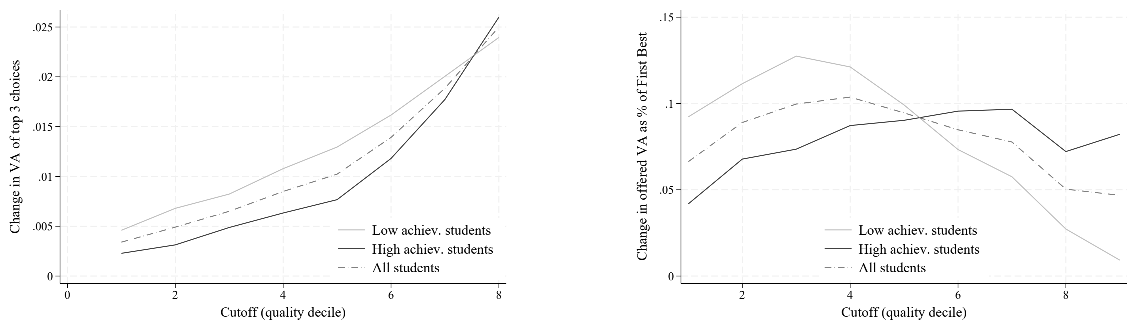
Figure A.1.10: Information About VA vs. Information About Achievement Levels



Notes: This figure plots the average welfare change, as defined by the average change in student test scores with respect to the status quo, by student subgroups for four different counterfactual simulations of student assignment. Welfare gains are expressed as a percentage of the average first-best achievement gains. Student subgroups are defined by combinations of race and baseline achievement. “VA" denotes the simulated student assignment under full information about school value added, “VA-5L" a counterfactual in which schools are rated from 1 to 5 based on their VA quintile, “Lev" a counterfactual in which students are told about differences in school achievement levels and these are presented as differences in VA, while “VA-5L" the counterfactual in which schools are rated from 1 to 5 based on their achievement level quintile.

Figure A.1.11: School Quality of Choices and Offers as Signal Precision Increases at the Top

(a) Δ VA in Choices | (b) Welfare Gains



Notes: This figure plots how value added of the top three school choices and of school offers changes as a function of the cutoff used to assign schools to a low or a high quality rating. Panel (a) plots changes in the average value-added of students top three choices, while Panel (b) plot achievement gains as a share of first-best gains. The dotted lines are for averages across all students, dark gray lines are for students with above median baseline test scores, light gray lines for students with below median baseline test scores.

Table A.1.2: School Progress Report Score Components

Component	**Description**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**Average**
Progress	*Students on track for graduation (credits), Students in school lowest 3rd on track for graduation, Regents pass rate*	51%	56%	56%	57%	56%	50%	47%	53%
Performance	*Graduation rate, Regents Diploma rate*	31%	25%	24%	24%	25%	20%	19%	24%
Environment	*Attendance rate, answers from school environment survey*	13%	14%	14%	14%	14%	14%	16%	14%
College and Career readiness	*College readiness index, college enrollment rate*	0%	0%	0%	0%	0%	10%	10%	3%
Extra points	*ELL diploma rate, city lowest 3rd diploma rate, sped regents pass rate*	5%	6%	6%	5%	5%	7%	8%	6%
Total		100%	100%	100%	100%	100%	101%	101%	100%

Notes: This table describes the components of the quality score used to assign letter grades, their year-specific weight and the outcomes used to create them. The last column reports the average weight of each component across years. The year refers to the fall of the school year of the progress report. For example, 2006 refers to the 2006-2007 school progress report, which graded schools existing in the 2006-2007 school year. This progress report was made available to the public during the 2007-2008 school year, and therefore would have been used by the 2008 high school enrollment cohort to decide where to apply.

Table A.1.3: Demand Responses to Introduction and Removal of School Quality Signals - Pooled Pre-Post Estimates

	School share			School log share		
	minority	white	difference	minority	white	difference
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: effect of introduction of information						
Type A $\cdot Post2010$	0.25***	0.17***	0.09**	0.26***	0.16**	0.10**
	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.03)
Average $\cdot Post2010$	-0.13**	-0.08*	-0.05**	-0.10	-0.18**	0.07**
	(0.05)	(0.04)	(0.02)	(0.07)	(0.06)	(0.02)
Type Low $\cdot Post2010$	-0.47***	-0.28**	-0.20***	-0.66***	-0.71***	0.05
	(0.12)	(0.09)	(0.05)	(0.15)	(0.13)	(0.05)
Never graded $\cdot Post2010$	0.08	0.07	0.02	0.39	0.17	0.22
	(0.07)	(0.05)	(0.02)	(0.34)	(0.22)	(0.19)
Graduation % (SD) · Visible	-0.03	-0.02	-0.01	0.00	-0.04	0.04
	(0.03)	(0.02)	(0.01)	(0.04)	(0.04)	(0.02)
College % (SD) · Visible	0.02*	0.01*	0.01	0.02	0.02*	-0.01
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)
N	54855	54855	109710	27896	15267	43163
N schools	463	463	463	446	432	446
Panel B: effect of removal of letters						
Type A $\cdot Post2015$	-0.13**	-0.10*	-0.03	-0.15***	-0.11**	-0.04
	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)
Type Average $\cdot Post2015$	-0.00	0.01	-0.01	-0.03	0.01	-0.04
	(0.01)	(0.01)	(0.01)	(0.03)	(0.02)	(0.02)
Type Low $\cdot Post2015$	0.02	0.00	0.02	0.02	0.06	-0.04
	(0.03)	(0.02)	(0.02)	(0.10)	(0.09)	(0.07)
Never graded $\cdot Post2015$	0.15**	0.10**	0.05**	0.20**	0.15**	0.05
	(0.04)	(0.03)	(0.02)	(0.05)	(0.04)	(0.05)
N	38,370	38,370	76,740	18,596	10,411	29,007
N schools	453	453	453	432	427	432

Notes: This table presents regression estimates of changes in demand for schools after the introduction (panel A) or after the removal (panel B) of letter grades for different categories of schools. The dependent variable is the share of students (or log share in columns (4)-(6)) in demographic cell c and application cohort t ranking the school among their first three choices. Demographic cells are defined by the interaction of student race, residential borough and baseline test score tercile. Schools are divided into mutually exclusive categories, fixed over time: *Type A* indicates schools receiving a grade of A in most years, *Type Low* indicates schools receiving a grade of C, D or F in most years, *Never graded* indicates schools that were never graded, while *Type Average* is a residual category for the remaining schools. Columns (1), (2), (4) and (5) report changes in the school shares over time separately by applicant race, pooling the event study coefficients β_L^t of equation (1.3) into pre-post differences. Columns (3) and (6) report estimates of the race-difference in changes in demand over time, pooling the event study coefficients δ_L^t of equation (1.5) into pre-post differences. Panel A uses application cohorts of 2006-2014, while panel B uses the 2011-2016 cohorts.

Table A.1.4: Demand Responses to Quality Signals - Heterogeneity by Applicant Race

	School share				School log share			
	Black+Hispanic students		White+Asian students		Black+Hispanic students		White+Asian students	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A	0.19**	0.19***	0.08*	0.04	0.18**	0.29***	0.22	0.25***
	(0.04)	(0.04)	(0.03)	(0.02)	(0.07)	(0.06)	(0.11)	(0.05)
B	0.10*	0.09**	0.05	0.01	0.05	0.17***	0.07	0.13**
	(0.04)	(0.02)	(0.03)	(0.01)	(0.05)	(0.03)	(0.09)	(0.04)
C	-0.03		0.01		-0.19**		-0.14*	
	(0.04)		(0.02)		(0.05)		(0.06)	
D	-0.12	-0.07	-0.01	-0.01	-0.37***	-0.12	-0.34**	-0.18*
	(0.07)	(0.04)	(0.02)	(0.02)	(0.05)	(0.06)	(0.08)	(0.07)
F	-0.30**	-0.23*	-0.01	0.01	-0.44**	-0.26**	0.12	0.07
	(0.10)	(0.10)	(0.03)	(0.02)	(0.12)	(0.09)	(0.14)	(0.06)
Graduation % (SD)	0.00	-0.09**	-0.01	-0.04**	0.06	0.01	0.05	0.01
	(0.06)	(0.03)	(0.02)	(0.01)	(0.08)	(0.04)	(0.08)	(0.04)
College % (SD)	0.03	-0.00	-0.00	-0.02	0.01	-0.02	-0.01	-0.05
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)	(0.05)
Graduation % (SD) · Visible	0.02**	0.21***	0.01**	0.09***	0.05***	0.26***	0.05***	0.26***
	(0.01)	(0.03)	(0.00)	(0.02)	(0.01)	(0.04)	(0.01)	(0.04)
College % (SD) · Visible	0.00	0.03*	-0.00	0.01	0.02	0.06**	0.04	0.06**
	(0.02)	(0.01)	(0.01)	(0.02)	(0.03)	(0.02)	(0.03)	(0.01)
Only graded schools		X		X		X		X
N	32,190	22,815	32,190	22,815	15,213	11,936	8,266	6,597
N schools	458	338	458	338	429	334	409	319
Average school share	0.625	0.782	0.579	0.745	0.625	0.782	0.579	0.745

Notes: This table presents regression estimates of letter grade effects on demand for schools, separately measuring effects on the school choices of Black and Hispanic students and of white and Asian students. The dependent variable is the share (or log share in columns (4)-(6)) of students in demographic cell c and application cohort t ranking the school among their first three choices. Demographic cells are defined by the interaction of student race, residential borough and baseline test score tercile. The first 5 rows report estimates of the coefficients β_g in equation (1.4) for each letter grade. The other rows the coefficients of a school graduation or college rates in the year prior to when cohort t applies and of their interaction with an indicator (*Visible*) for years when these statistics were printed on the school directories. Other controls include school-cell fixed effects, year-cell fixed effects, a school average Regents performance and the share of white and Asian students enrolled at the school in the year prior cohort t applies to school. Standard errors are clustered at the school-year level. Estimates use applicant cohorts from 2010 to 2014 included. Even columns restrict the observations in the preceding columns to schools receiving a grade, so that the omitted category is receiving a grade of C.

Table A.1.5: Demand Responses to Quality Signals - Robustness to Using Both Letters

	All (1)	Minority (2)	White (3)	All (4)	Minority (5)	White (6)	All (7)	Minority (8)	White (9)
A	0.14***	0.19***	0.03	0.15***	0.20***	0.04			
	(0.02)	(0.04)	(0.02)	(0.02)	(0.03)	(0.02)			
B	0.07**	0.10**	0.01	0.07**	0.10**	0.02			
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)			
D	-0.06	-0.08	-0.01	-0.05	-0.07	-0.00			
	(0.04)	(0.04)	(0.02)	(0.04)	(0.04)	(0.02)			
F	-0.18*	-0.27*	0.01	-0.15	-0.24*	0.02			
	(0.07)	(0.10)	(0.03)	(0.08)	(0.11)	(0.04)			
A - 2				0.09**	0.12*	0.02			
				(0.03)	(0.04)	(0.02)			
B - 2				0.05	0.07*	0.02			
				(0.02)	(0.03)	(0.02)			
D - 2				-0.01	-0.03	0.01			
				(0.03)	(0.04)	(0.01)			
F - 2				-0.20	-0.25	-0.08			
				(0.11)	(0.14)	(0.05)			
Two As							0.13**	0.17**	0.02
							(0.03)	(0.04)	(0.04)
One A							0.07**	0.09**	0.01
							(0.02)	(0.02)	(0.03)
Graduation %	-0.07**	-0.08**	-0.04*	-0.10**	-0.12**	-0.05**	-0.07**	-0.08**	-0.04*
	(0.02)	(0.02)	(0.01)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)
College %	-0.01	0.00	-0.01	-0.00	0.00	-0.01	0.00	0.01	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Graduation % · Visible	0.20***	0.24***	0.10***	0.20***	0.24***	0.11***	0.20***	0.24***	0.10***
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.03)	(0.01)
College % · Visible	0.02	0.03	0.01	0.03	0.03	0.01	0.03	0.04	0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
N	20,685	20,685	20,685	20,685	20,685	20,685	20,685	20,685	20,685
N schools	316	316	316	316	316	316	316	316	316
Average school share	0.766	0.782	0.745	0.766	0.782	0.745	0.766	0.782	0.745

Notes: This table presents robustness checks on estimates of letter grade effects presented in table 1.5 and A.1.4 by separately estimating the effect of the two letter grades (one for each of the two preceding years) printed on the directory received by cohort t. The dependent variable is the share of students from a given demographic group listing the school among their first three choices. The sample includes applicant cohorts of 2010 - 2014. Columns (1) - (3) report for comparison the benchmark estimates of the effects of the most recent letter grade printed on the directory in equation 1.4. The equation estimated in columns (4) - (6) extends equation 1.4 by adding letter grade dummies for the additional grade printed on the directory, corresponding to that received two years prior to when cohort t applies to high school. Columns (7) - (9) substitute letter grade indicators in equation 1.4 with indicators for receiving two consecutive As or only one A (in one out of the two years), leaving as omitted category the event of not receiving an A in any of the two years considered in the directory of cohort t. Controls in columns (1), (4) and (7) are the same as in table 1.5 and those in the remaining columns are the same as in table A.1.4. These always include school-cell and year-cell fixed effects.

Table A.1.6: Demand Responses to Quality Signals - Robustness to Using Additive or Interactive Models

	All (1)	Minority (2)	White (3)	All (4)	Minority (5)	White (6)
A	0.07**	0.10***	0.02			
	(0.02)	(0.02)	(0.03)			
Low	-0.08**	-0.11**	-0.02			
	(0.02)	(0.03)	(0.02)			
A - 2	0.03	0.05	-0.00			
	(0.02)	(0.02)	(0.02)			
Low - 2	-0.07*	-0.09*	-0.03			
	(0.03)	(0.04)	(0.02)			
A-A				0.10**	0.14**	0.01
				(0.03)	(0.04)	(0.05)
A-B				0.06*	0.09**	0.01
				(0.03)	(0.03)	(0.04)
A-Low				-0.06	-0.08	-0.02
				(0.06)	(0.07)	(0.05)
B-A				0.02	0.04	-0.01
				(0.02)	(0.02)	(0.03)
B-Low				-0.09*	-0.11*	-0.05
				(0.04)	(0.05)	(0.03)
Low-A				-0.09	-0.12*	-0.03
				(0.05)	(0.05)	(0.04)
Low-B				-0.10**	-0.14**	-0.04
				(0.03)	(0.04)	(0.03)
Low-Low				-0.15**	-0.20**	-0.04
				(0.05)	(0.06)	(0.03)
Graduation %	-0.09**	-0.11**	-0.04*	-0.09**	-0.11**	-0.04*
	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)
College %	-0.00	0.00	-0.01	-0.00	0.00	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Graduation % · Visible	0.20***	0.24***	0.10***	0.20***	0.24***	0.10***
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)
College % · Visible	0.03	0.04	0.00	0.03	0.04	0.00
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Constant	-0.06	-0.44	0.56	-0.07	-0.46	0.55
	(0.43)	(0.51)	(0.53)	(0.42)	(0.51)	(0.53)
N	20,685	20,685	20,685	20,685	20,685	20,685
N schools	316	316	316	316	316	316
Average school share	0.766	0.782	0.745	0.766	0.782	0.745

Notes: This table presents robustness checks on benchmark estimates of letter grade effects by considering models that estimate the effect of the two letter grades (one for each of the two preceding years) printed on the directory received by cohort t and that allow the two grade effects to be either additive (columns (1)-(3)) or interactive (columns (4)-(6)). The dependent variable is the share of students from a given demographic group listing the school among their first three choices. The sample includes applicant cohorts of 2010 - 2014. Grades of C, D, and F are pooled in one "Low grade" category. Controls in columns (1) and (4) are the same as in table 1.5 and those in the remaining columns are the same as in table A.1.4. These always include school-cell and year-cell fixed effects.

Table A.1.7: Demand Responses to Quality Score and Its Subcomponents

	School share in top 3 choices					
	(1)	(2)	(3)	(4)	(5)	(6)
Progress score (SD)	0.03**	-0.00	-0.00			
	(0.01)	(0.01)	(0.01)			
Performance score (SD)	0.04***	0.02*	0.02**			
	(0.01)	(0.01)	(0.01)			
Environment score (SD)	0.02	0.01	0.01			
	(0.01)	(0.01)	(0.01)			
Progress score^2 (SD)			0.00			
			(0.00)			
Performance score^2 (SD)			0.00			
			(0.01)			
Environment score^2 (SD)			-0.01			
			(0.01)			
Quality score (SD)				0.06**	-0.01	-0.18*
				(0.01)	(0.02)	(0.08)
Quality score^2 (SD)						0.02*
						(0.01)
A		0.12***	0.12***		0.15***	0.18***
		(0.02)	(0.02)		(0.03)	(0.04)
B		0.06**	0.06**		0.07**	0.10**
		(0.02)	(0.02)		(0.02)	(0.02)
D		-0.04	-0.04		-0.06	-0.09**
		(0.03)	(0.02)		(0.03)	(0.03)
F		-0.13	-0.12*		-0.15	-0.22*
		(0.06)	(0.06)		(0.07)	(0.08)
N	22,815	22,815	22,815	22,815	22,815	22,815
N schools	338	338	338	338	338	338
Average school share	0.766	0.766	0.766	0.766	0.766	0.766

Notes: This table presents regression estimates of the effect of the quality score components on demand for schools, with and without controlling for letter grade fixed effects. It shows that demand does not respond to changes in the quality score and its components, beyond the variation controlled for by letter grade fixed effects in columns (2)-(4) and (5)-(6). The dependent variable is the share of students (or log share in columns (4)-(6)) in demographic cell c and application cohort t ranking the school among their first three choices. Demographic cells are defined by the interaction of a student residential borough and baseline test score tercile. The set of controls is the same as for regressions in table 1.5. The estimation sample includes cohorts from 2010 to 2014 included.

Table A.1.8: Applicants Descriptive Statistics by Information Exposure Treatment Status

	All		Minority		Non-Minority	
	Control	Treated	Control	Treated	Control	Treated
	(1)	(2)	(3)	(4)	(5)	(6)
Black	0.15	0.40	0.31	0.48	0.00	0.00
Hispanic	0.34	0.42	0.69	0.52	0.00	0.00
white	0.24	0.07	0.00	0.00	0.46	0.41
Asian	0.27	0.10	0.00	0.00	0.53	0.54
Subsidized lunch	0.69	0.86	0.78	0.88	0.60	0.77
Ell	0.07	0.12	0.08	0.12	0.07	0.14
7th grade Math	0.47	-0.10	0.09	-0.19	0.83	0.34
7th grade English	0.38	-0.13	0.12	-0.18	0.63	0.11
Bronx	0.05	0.38	0.08	0.42	0.02	0.18
Brooklyn	0.30	0.37	0.27	0.38	0.32	0.31
Manhattan	0.14	0.03	0.17	0.04	0.12	0.01
Queens	0.41	0.19	0.40	0.15	0.43	0.33
Staten Island	0.10	0.03	0.08	0.00	0.12	0.17
Share of mostly grade A seats in neighborhood	0.17	0.10	0.16	0.11	0.18	0.07
Share of mostly low grade seats in neighborhood	0.16	0.23	0.16	0.21	0.15	0.31
Average 7th grade math in neighborhood	0.00	-0.15	-0.06	-0.17	0.06	-0.07
Minimum distance to A (minutes)	24.09	23.73	22.92	22.52	25.24	29.90
Minimum distance to Log grade (minutes)	29.98	23.28	28.52	22.33	31.42	28.08
% of 2006 MS students applying to mostly A	0.67	0.42	0.66	0.42	0.69	0.39
% of 2006 MS students applying to mostly low	0.12	0.39	0.16	0.41	0.09	0.30
N	333,938	254,810	162,358	209,571	171,580	45,239

Notes: This table provides student descriptive statistics across values of the indicator $Treated_i$ defined in section 1.3.2. Columns (1)-(2) report mean statistics considering all students, while columns (3)-(6) split students by race. The term “Minority" refers to Black and Hispanic students, while “Non-Minority" includes both white and Asian students.

Table A.1.9: Consequences of Letter Grade Introduction on Simulated Offers

	(1) Grade A	(2) Low grade	(3) Regents VA σ	(4) Regents VA pct	(5) Peer quality pct	(6) White and Asian %	(7) Screened	(8) P(matched) or P(enrolls)
Panel A: simulated offers under no screening								
$Post2010 \cdot M_i$	0.034***	-0.041***	0.026***	3.716***	0.615***	-0.011***	-0.002	-0.012***
	(0.003)	(0.002)	(0.001)	(0.158)	(0.135)	(0.001)	(0.003)	(0.002)
$Post2010$	0.017***	-0.049***	0.028***	2.464***	4.893***	0.028***	0.033***	-0.025***
	(0.003)	(0.002)	(0.001)	(0.126)	(0.102)	(0.001)	(0.002)	(0.002)
N	431,526	431,526	431,443	431,443	431,526	431,373	422,654	503,150
Black+Hispanic mean	0.154	0.198	-0.0119	48.94	56.67	0.190	0.157	0.914
White+Asian mean	0.263	0.0979	0.111	68.15	77.68	0.496	0.214	0.763
Panel B: offers								
$Post2010 \cdot M_i$	0.026***	-0.052***	0.029***	4.492***	1.582***	-0.006***	-0.005*	-0.000
	(0.003)	(0.002)	(0.001)	(0.142)	(0.118)	(0.001)	(0.003)	(0.002)
$Post2010$	0.017***	-0.042***	0.025***	1.967***	4.108***	0.021***	0.036***	-0.031***
	(0.002)	(0.001)	(0.001)	(0.108)	(0.084)	(0.001)	(0.002)	(0.001)
N	459,617	459,617	459,617	459,617	459,617	459,617	459,617	502,923
Black+Hispanic mean	0.144	0.211	-0.0264	46.75	53.93	0.167	0.224	0.929
White+Asian mean	0.276	0.0857	0.123	69.78	79.49	0.503	0.443	0.919

Notes: This table presents pooled differences in differences estimates of the differential changes in the attributes of school offers (panel B) and of school offers simulated using admission rules that remove all priorities based on residential address and academic screening (panel A). The sample includes students applying to enroll in 9th grade between 2006 and 2014. Controls include gender, ell status, subsidized lunch status and fixed effects for combinations of student borough and baseline test score terciles.

Table A.1.10: Consequences of Letter Grade Introduction on Achievement Inequality

	Regents Math (1)	SAT Math (2)	Graduates in time (3)	College in time (4)
Panel A: pooled diff-in-diff estimates by race				
$Post2010 \cdot M_i$	0.06***	0.01	0.05***	0.07***
	(0.00)	(0.00)	(0.00)	(0.00)
N	339,182	292,828	428,789	426,937
Black+Hispanic mean	-0.189	-0.283	0.660	0.452
White+Asian mean	0.628	0.656	0.874	0.768
Panel B: pooled diff-in-diff estimates by exposure to new information (Black and Hispanic students)				
$Post2010 \cdot Treated_i$	0.02***	-0.02***	0.01***	0.02***
	(0.01)	(0.01)	(0.00)	(0.00)
N	210,817	143,634	244,386	243,302
Treated Black+Hispanic mean	-0.304	-0.395	0.617	0.394
Control Black+Hispanic mean	-0.0170	-0.147	0.720	0.534
Panel C: pooled diff-in-diff estimates by exposure to new information (White and Asian students)				
$Post2010 \cdot Treated_i$	0.07***	0.01	0.03***	0.05***
	(0.01)	(0.01)	(0.00)	(0.01)
N	85,391	108,516	129,746	129,551
Treated White+Asian mean	0.294	0.264	0.802	0.645
Control White+Asian mean	0.728	0.750	0.894	0.802

Notes: This table presents pooled differences in differences estimates of the differential changes in the achievement outcomes by student race (panel A), and by values of the variable $Treated_i$ defined in section 1.3.2 within race (panels B and C) after the introduction of letter grades. The sample includes students from cohorts between 2006 and 2014, who enroll in the district and have non-missing achievement outcomes. Controls include gender, ell status, subsidized lunch status and fixed effects for combinations of student borough and baseline test score terciles.

Table A.1.11: Model Estimates

	Student Demographic Cell (Race x Baseline Tercile)								
Race:		Black			Hispanic			White	
Baseline tercile:	Low	Median	High	Low	Median	High	Low	Median	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
γ_c	3.823	6.044	7.000	3.084	4.433	6.193	2.050	4.072	6.874
	(0.033)	(0.049)	(0.047)	(0.027)	(0.027)	(0.039)	(0.036)	(0.037)	(0.034)
β_c^{white}	1.820	1.670	3.102	1.118	1.414	2.315	-0.209	0.875	1.531
	(0.049)	(0.081)	(0.086)	(0.043)	(0.053)	(0.078)	(0.063)	(0.066)	(0.073)
$\beta_c^{peerquality}$	3.899	4.427	5.953	4.091	3.910	5.743	2.172	4.003	4.013
	(0.027)	(0.049)	(0.055)	(0.026)	(0.031)	(0.045)	(0.04)	(0.043)	(0.049)
μ_{cL}	-0.107	-0.228	-0.086	-0.096	-0.089	-0.180	-0.125	-0.082	-0.060
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.011)	(0.006)	(0.004)
μ_{cH}	-0.146	0.084	0.143	-0.078	0.026	0.289	0.008	0.056	0.283
	(0.006)	(0.007)	(0.006)	(0.006)	(0.005)	(0.006)	(0.014)	(0.008)	(0.005)
σ_{cL}^{-1}	2.069	2.150	2.225	1.942	1.976	2.218	1.878	2.204	2.605
	(0.054)	(0.057)	(0.046)	(0.05)	(0.035)	(0.045)	(0.094)	(0.058)	(0.041)
σ_{cH}^{-1}	1.815	2.822	2.327	2.362	2.702	2.129	2.099	2.722	3.655
	(0.043)	(0.067)	(0.039)	(0.069)	(0.054)	(0.028)	(0.102)	(0.074)	(0.045)
$\tilde{\xi}_{cj}$									
mean	480	522	423	451	444	411	410	371	233
within-cell SD	(46)	(52)	(57)	(37)	(42)	(50)	(46)	(53)	(66)
Corr($\tilde{\xi}_{cj}$, VA)	0.015	0.240	0.377	0.034	0.304	0.417	0.367	0.485	0.600
p-value	[0.76]	[0]	[0]	[0.49]	[0]	[0]	[0]	[0]	[0]
Corr($\tilde{\xi}_{cj}$, Peer quality)	0.137	0.407	0.519	0.067	0.429	0.555	0.502	0.623	0.737
p-value	[0]	[0]	[0]	[0.17]	[0]	[0]	[0]	[0]	[0]
Corr($\tilde{\xi}_{jc}$, % white)	0.050	0.202	0.251	0.145	0.337	0.381	0.635	0.651	0.700
p-value	[0.31]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]
N school	423	423	422	423	423	422	423	423	420
N students	36,433	27,521	13,279	44,676	38,949	21,485	13,120	25,938	53,816

Notes: This table presents the model estimates by student demographic cells defined by the interaction of student race and baseline test score tercile. Asymptotic standard errors in parenthesis take into account the first stage sampling error and rely on numerical approximations when necessary. Square brackets report the p value of a test of the significance of the correlation coefficient in the row above.

Table A.1.12: Model Fit

	All students		Minority students		Non-Minority students		Below median Math		Above median Math	
	Real	Simulated	Real	Simulated	Real	Simulated	Real	Simulated	Real	Simulated
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Average in top 3 choices										
Regents VA (σ)	0.66	0.67	0.54	0.52	0.88	0.93	0.43	0.43	0.88	0.89
Regents VA (percentile)	67.5	67.4	63.3	62.4	75.0	76.3	60.6	60.4	74.0	74.0
Regents VA (σ) left unexploited	1.74	1.74	1.97	1.98	1.35	1.29	1.96	1.96	1.54	1.53
SAT VA (σ)	0.89	0.89	0.60	0.59	1.40	1.40	0.42	0.46	1.33	1.29
SAT VA (percentile)	72.9	72.9	66.6	66.3	84.0	84.7	63.3	64.0	81.9	81.4
SAT VA (σ) left unexploited	1.55	1.55	1.78	1.78	1.13	1.12	1.70	1.66	1.40	1.44
Peer quality	0.26	0.25	0.12	0.11	0.52	0.50	0.03	0.05	0.48	0.44
White+Asian %	0.38	0.38	0.27	0.28	0.57	0.57	0.27	0.28	0.48	0.48
Commuting time	39.87	37.40	40.47	37.97	38.79	36.39	38.92	36.53	40.76	38.22
Panel B: 2016-2014 changes in application behavior										
P(applies to A) as 1st	-0.038	-0.060	-0.049	-0.066	-0.023	-0.054	-0.050	-0.056	-0.036	-0.072
P(applies to A) in top3	-0.028	-0.051	-0.036	-0.064	-0.014	-0.029	-0.039	-0.064	-0.023	-0.044
P(applies to A) ever	-0.007	-0.018	-0.010	-0.022	-0.002	-0.010	-0.010	-0.023	-0.007	-0.015
P(applies to C/D/F) as 1st	0.006	0.022	0.010	0.031	0.000	0.009	0.013	0.028	0.002	0.020
P(applies to C/D/F) in top3	0.023	0.048	0.037	0.065	0.003	0.024	0.041	0.058	0.014	0.048
P(applies to C/D/F) ever	0.064	0.078	0.086	0.098	0.033	0.053	0.087	0.081	0.055	0.090
Panel C: simulated school offers										
Regents VA (σ)	0.39	0.29	0.20	0.11	0.73	0.62	0.08	-0.01	0.68	0.56
Regents VA (percentile)	59.7	56.1	53.4	50.2	71.0	66.9	50.1	46.5	68.8	64.9
SAT VA (σ)	0.51	0.40	0.17	0.07	1.12	1.01	0.03	-0.09	0.96	0.85
SAT VA (percentile)	64.9	61.7	56.8	53.5	79.4	76.7	53.6	49.5	75.6	72.8
Peer quality	0.07	0.01	-0.09	-0.15	0.38	0.32	-0.18	-0.24	0.31	0.25
White+Asian %	0.31	0.27	0.20	0.16	0.52	0.47	0.20	0.15	0.42	0.38
Commuting time	38.00	33.70	38.90	33.50	36.38	34.08	37.68	32.29	38.30	35.00
Share matched	0.94	0.96	0.94	0.98	0.93	0.94	0.94	0.95	0.93	0.98
N	53,014		33,896		19,118		25,698		27,316	

Notes: This table assesses the model fit. It compares summary statistics of the characteristics of students' first three school choices (Panel A), school offers (Panel C) and changes in the probability of applying to high or low letter grade schools in the real data with those simulated using the model estimates (Panel B). The sample is the 2016 applicant cohort in panels A and C, and applicants in 2014 and 2016 for panel B. Simulations of the school match are based on priorities that are reconstructed on the basis of the admission rules used in the 2016 general education high school match and real school capacities.

Table A.1.13: Robustness of Model Estimates - More Discrete School Types

	By Race			By 7th Grade Math Tercile		
	Black	Hispanic	White	Low	Median	High
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: second step - preferences						
γ_c	5.7	4.9	4.5	3.8	5.5	5.8
β_c^{white}	1.3	0.5	0.6	0.8	-0.3	1.9
$\beta_c^{peerquality}$	4.5	4.3	3.0	3.5	4.7	3.6
$\tilde{\xi}_{cj}$ SD	17.1	14.4	21.1	14.4	16.7	21.5
$\tilde{\xi}_{cj}$ range	106.2	88.4	110.8	93.8	100.6	109.0
$\tilde{\xi}_{cj}$ skewness	0.1	0.3	0.7	0.2	0.4	0.6
Corr($\tilde{\xi}_{cj}$, VA)	0.07	0.13	0.33	0.05	0.19	0.31
Corr($\tilde{\xi}_{cj}$, Peer quality)	0.31	0.33	0.69	0.19	0.49	0.68
Corr($\tilde{\xi}_{cj}$, % white)	0.17	0.32	0.69	0.22	0.44	0.56
Panel B: second step - beliefs						
μ_{cj} 1^{st} quartile of R_j	-0.07	-0.01	0.13	-0.23	0.03	0.27
μ_{cj} 2^{nd} quartile of R_j	-0.20	-0.08	-0.41	-0.06	-0.05	-0.58
μ_{cj} 3^{rd} quartile of R_j	-0.03	-0.05	0.07	-0.04	0.02	0.00
μ_{cj} 4^{th} quartile of R_j	0.14	0.31	0.52	0.02	0.38	0.62
σ_{cj}^{-1} 1^{st} quartile of R_j	1.33	1.04	0.77	1.56	0.87	0.65
σ_{cj}^{-1} 2^{nd} quartile of R_j	1.49	1.32	1.13	1.38	1.56	0.95
σ_{cj}^{-1} 3^{rd} quartile of R_j	0.21	0.88	0.52	0.56	0.67	0.47
σ_{cj}^{-1} 4^{th} quartile of R_j	1.19	1.49	1.47	1.55	1.06	1.60

Notes: This table summarizes the second step model estimates when prior means and precision are a non parametric function of four discrete school types based on quartiles of average achievement levels. Panel A reports estimates of the preference parameters $\gamma_c, \beta_c, \xi_{cj}$ and panel B of the first and second moments of priors for each school type, taking a weighted average of cell-specific estimates across cells sharing the same covariate (race or baseline test score), using weights proportional to cell size.

Table A.1.14: Robustness of Model Estimates - Beliefs as a Linear Function of Achievement Levels

	By race			By 7th grade Math tercile		
	Black	Hispanic	White	Low	Median	High
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: second step - preferences						
γ_c	5.2	4.1	5.2	3.3	4.9	6.1
β_c^{white}	2.1	1.0	0.5	0.8	1.2	1.4
$\beta_c^{peerquality}$	4.8	4.6	4.7	3.7	4.7	5.6
$\tilde{\xi}_{cj}$ SD	17.0	14.2	20.2	14.4	16.3	20.6
$\tilde{\xi}_{cj}$ range	103.9	88.2	108.9	93.7	97.5	108.1
Corr($\tilde{\xi}_{cj}$, VA)	0.15	0.22	0.53	0.09	0.32	0.51
Corr($\tilde{\xi}_{cj}$, Peer quality)	0.29	0.31	0.66	0.18	0.46	0.64
Corr($\tilde{\xi}_{cj}$, % white)	0.13	0.29	0.68	0.21	0.38	0.56
Panel B: second step - beliefs						
μ_{c0}	0.03	-0.70	-0.54	-0.18	0.08	-1.26
μ_{c1}	-0.01	0.07	0.07	0.01	-0.01	0.15
σ_{c0}	0.97	0.88	1.14	1.18	0.67	1.13
σ_{c1}	0.13	0.12	0.14	0.10	0.18	0.11
μ_{cj} 1^{st} quartile of R_j	-0.08	-0.12	0.04	-0.11	0.00	-0.06
μ_{cj} 2^{nd} quartile of R_j	-0.09	-0.07	0.09	-0.10	0.00	0.05
μ_{cj} 3^{rd} quartile of R_j	-0.09	-0.03	0.13	-0.10	-0.01	0.13
μ_{cj} 4^{th} quartile of R_j	-0.11	0.05	0.21	-0.09	-0.02	0.30

Notes: This table summarizes the second step model estimates when prior means and precision are a continuous linear function of school average achievement levels. Panel A reports estimates of the preference parameters $\gamma_c, \beta_c, \xi_{cj}$ and panel B of parameters defining the first and second prior moments, taking a weighted average of cell-specific estimates across cells sharing the same covariate (race or baseline test score), using weights proportional to cell size. The last rows report the average quality priors for schools in each quartile of the distribution of the average school performance on Regents exams, R_j, based on μ_0 and μ_1 estimates.

Table A.1.15: Model Estimates - Beliefs as a Linear Function of Achievement Levels and Value Added

	By race			By 7th grade Math tercile		
	Black	Hispanic	White	Low	Median	High
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: second step - preferences						
γ_c	5.6	4.5	5.4	3.7	5.2	6.5
β_c^{white}	0.9	0.5	-1.0	0.5	-0.7	0.6
$\beta_c^{peerquality}$	4.9	4.3	4.3	3.7	4.7	5.1
$\tilde{\xi}_{cj}$ SD	22	17	27	15	22	29
$\tilde{\xi}_{cj}$ range	123	103	144	100	120	149
$\tilde{\xi}_{cj}$ skewness	0.00	0.26	0.45	0.24	0.26	0.26
Corr($\tilde{\xi}_{cj}$, VA)	0.15	0.21	0.47	0.09	0.32	0.45
Corr($\tilde{\xi}_{cj}$, Peer quality)	0.25	0.28	0.57	0.16	0.42	0.53
Corr($\tilde{\xi}_{cj}$, % white)	0.15	0.28	0.60	0.21	0.39	0.46
Panel B: second step - beliefs						
μ_{c0}	-0.09	-0.07	0.07	-0.11	-0.03	0.06
μ_{c1}	-0.01	0.04	0.13	0.02	0.01	0.14
μ_{c2}	0.15	0.15	0.14	0.14	0.16	0.14
σ_{c0}	2.88	2.52	3.18	2.79	3.10	2.63
σ_{c1}	-0.12	-0.10	0.25	-0.17	0.17	0.04
σ_{c2}	0.08	0.11	0.10	0.09	0.11	0.09
Absolute Bias	0.65	0.62	0.55	0.63	0.64	0.54
μ_{cj} below med. R_j, below med. Q_j	-0.20	-0.22	-0.14	-0.24	-0.15	-0.16
μ_{cj} above med. R_j, below med. Q_j	-0.19	-0.16	0.02	-0.19	-0.12	0.00
μ_{cj} below med. R_j, above med. Q_j	-0.04	-0.05	0.03	-0.08	0.01	0.00
μ_{cj} above med. R_j, above med. Q_j	0.04	0.11	0.33	0.05	0.13	0.32

Notes: This table summarizes the second step model estimates when prior means and precision are a continuous linear function of school average achievement levels R_j and school quality Q_j. That is: $\mu_j = \mu_0 + \mu_1 \cdot R_j + \mu_2 \cdot Q_j$, $\sigma_j^{-1} = \sigma_0 + \sigma_1 \cdot R_j + \sigma_2 \cdot Q_j$. Panel A reports estimates of the preference parameters $\gamma_c, \beta_c, \xi_{cj}$ and panel B of parameters defining the first and second prior moments, taking a weighted average of cell-specific estimates across cells sharing the same covariate (race or baseline test score), using weights proportional to cell size. The last rows report the average quality priors for four types of schools, depending on whether they have above or below median value-added and above or below median average achievement levels, which are based on the estimates of μ_0 , μ_1 and μ_2 in the rows above.

Table A.1.16: Robustness of Model Estimates to Different Functional Forms

	By Race						By 7th Grade Math Tercile					
	Location and scale shift			Log Normal			Location and scale shift			Log Normal		
	Black	Hispanic	White	Black	Hispanic	White	Low	Median	High	Low	Median	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: second step - preferences												
γ_c	5.0	5.0	4.9	4.7	4.6	4.7	4.9	5.0	5.1	4.5	4.7	4.8
β_c^{white}	1.1	1.0	1.0	1.1	1.0	1.0	1.0	1.0	1.1	1.0	1.0	1.0
$\beta_c^{peerquality}$	2.3	2.2	2.1	2.1	2.1	2.0	2.1	2.1	2.3	2.1	2.1	2.1
$\tilde{\xi}_{cj}$ SD	18.1	15.1	21.7	18.1	15.1	21.5	14.5	17.4	22.9	14.3	17.3	23.0
$\tilde{\xi}_{cj}$ range	111.6	90.5	110.5	112.6	89.9	108.5	93.9	101.7	114.5	92.9	101.1	114.3
$\tilde{\xi}_{cj}$ skewness	0.2	0.3	0.7	0.2	0.3	0.7	0.3	0.4	0.7	0.3	0.3	0.7
Corr($\tilde{\xi}_{cj}$, VA)	0.12	0.17	0.33	0.13	0.17	0.33	0.07	0.22	0.34	0.07	0.23	0.35
Corr($\tilde{\xi}_{cj}$, Peer quality)	0.41	0.41	0.70	0.43	0.42	0.70	0.25	0.56	0.73	0.26	0.57	0.74
Corr($\tilde{\xi}_{cj}$, % white)	0.25	0.36	0.69	0.26	0.36	0.69	0.26	0.45	0.62	0.26	0.46	0.63
Panel B: second step - beliefs												
μ_{cL}	-0.10	-0.03	-0.02	-0.46	-0.47	-0.48	0.04	-0.06	-0.12	-0.44	-0.45	-0.53
μ_{cH}	0.24	0.15	0.21	-0.41	-0.38	-0.34	0.08	0.20	0.29	-0.44	-0.37	-0.31
σ_{cL}^{-1}	1.63	1.92	1.52	1.67	1.73	1.31	1.96	1.66	1.47	1.98	1.54	1.16
σ_{cH}^{-1}	1.74	1.81	1.74	1.69	1.78	1.45	1.91	1.69	1.71	2.00	1.63	1.28

Notes: This table summarizes the second step model estimates for models using alternative functional forms for prior distributions. “Location and scale shift" refers to a model in which priors have the same distribution as quality in the city up to a location and a scale shift that varies across school types. “Log Normal" refers to a model in which priors have a log normal distribution. Parameters of the log normal distribution are re-scaled to make them comparable to other estimates, so that value-added is always rescaled to have mean zero and standard deviation 1 across schools.

Table A.1.17: Robustness of Model Estimates - Strategic Reporting

	By race			By 7th grade Math tercile		
	Black	Hispanic	White	Low	Median	High
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: first step						
δ_{cjt} SD	53	31	33	37	38	39
δ_{cjt} range	302	182	169	217	211	206
Corr(δ_{cjt}, VA)	0.29	0.41	0.59	0.35	0.42	0.55
Corr(δ_{cjt}, Peer quality)	0.41	0.55	0.74	0.48	0.56	0.69
Corr(δ_{cjt}, % white)	0.28	0.44	0.68	0.38	0.46	0.59
Panel B: second step - preferences						
γ_c	5.6	4.3	5.9	2.7	5.7	7.4
β_c^{white}	3.2	3.7	2.5	2.5	4.3	2.7
$\beta_c^{peerquality}$	7.3	4.6	4.3	3.7	4.8	7.4
$\tilde{\xi}_{cj}$ SD	26	19	25	22	22	25
$\tilde{\xi}_{cj}$ range	159	118	122	134	127	131
$\tilde{\xi}_{cj}$ skewness	0.27	0.24	0.54	0.47	0.13	0.45
Corr($\tilde{\xi}_{cj}$, VA)	0.25	0.32	0.54	0.33	0.32	0.47
Corr($\tilde{\xi}_{cj}$, Peer quality)	0.41	0.45	0.69	0.47	0.47	0.62
Corr($\tilde{\xi}_{jc}$, % white)	0.21	0.30	0.64	0.36	0.31	0.52
Panel C: second step - beliefs						
μ_{cL}	-0.56	-0.01	-0.10	-0.03	-0.25	-0.31
μ_{cH}	0.64	0.95	0.41	1.20	0.36	0.46
σ_{cL}^{-1}	3.45	1.70	2.17	2.78	2.25	2.00
σ_{cH}^{-1}	2.45	1.49	2.52	1.08	2.80	2.47
Absolute Bias	0.41	0.41	0.37	0.42	0.31	0.46

Notes: This table summarizes the model estimates when students are allowed to report preferences strategically. Specifically, students are assumed to only consider schools where they have a non-zero probability in admission and to rank schools truthfully within this set. Panel A reports summary statistics for the estimates of the mean school utility δ_{cjt} obtained in the first step. Panel B reports the second step estimates of the preference parameters $\gamma_c, \beta_c, \xi_{cj}$ and panel C of the prior moments $\boldsymbol{\mu_c}, \boldsymbol{\sigma_c^{-1}}$ taking a weighted average of cell-specific estimates across cells sharing the same covariate (race or baseline test score), using weights proportional to cell size.

Table A.1.18: Full-Information Benchmark

	All students		Black + Hispanic		White+Asian		Below median Math		Above median Math	
	No info	Full info	No info	Full info	No info	Full info	No info	Full info	No info	Full info
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
					Panel A: Top 3 choices					
ΔW		0.069		0.081		0.047		0.064		0.073
VA - pct	68	74	62	71	76	81	60	68	74	81
Peer math - pct	76	79	70	74	88	89	67	71	85	87
White+Asian %	0.383	0.403	0.276	0.298	0.572	0.588	0.278	0.297	0.481	0.502
					Panel B: Offers					
ΔW| offered		0.011		0.013		0.008		0.010		0.013
ΔW		0.009		0.012		0.005		0.008		0.011
ΔW as % of first best		24%		30%		12%		19%		28%
VA - pct	55	57	50	52	65	66	46	47	64	66
Peer math - pct	63	63	55	55	77	77	49	49	75	75
White+Asian %	0.263	0.265	0.159	0.161	0.449	0.449	0.146	0.147	0.374	0.375
Offered	0.964	0.961	0.975	0.971	0.944	0.943	0.952	0.949	0.975	0.973
N	52,997		33,901		19,096		25,706		27,291	

Notes: This table compares summary statistics of the characteristics of students' first three school choices and school offers in the simulated status-quo ("No info") and in the full-information benchmark ("Full info"). In the status quo students receive no additional information about school quality from the policy maker, and form beliefs about quality only based on their priors. In the full information counterfactual, students are perfectly informed about the VA of each school. Welfare is measured by the student average Regents Math test scores. The first two columns report averages for the entire set of applicants, while the remaining columns split applicants by race or by baseline achievement (above and below the median 7th grade math test score).

Table A.1.19: Targeted Outreach

	Targeted students		Non- targeted students	
	Outreach	Full info	Outreach	Full info
	(1)	(2)	(3)	(4)
Panel A: Targeted = students from lowest performing middle schools				
ΔW - choices	0.074	0.074	0.000	0.065
ΔW - offers	0.033	0.011	-0.008	0.008
ΔW - offers under no screening	0.037	0.016	-0.012	0.005
% B+H in top 20% schools	0.027	0.026	0.027	0.026
N	17197		35800	
Panel B: Targeted = top performing Black and Hispanic students				
ΔW - choices	0.105	0.105	0.000	0.064
ΔW - offers	0.053	0.018	-0.006	0.008
ΔW - offers under no screening	0.054	0.012	-0.006	0.008
% B+H in top 20% schools	0.023	0.026	0.023	0.026
N	6089		46908	

Notes: This table compares changes in average value-added (ΔW) of students' first three school choices and school offers in the targeted outreach counterfactual ("Outreach") and in the full-information benchmark ("Full info"), relative to the status-quo. In the outreach counterfactuals only students denoted with "targeted" receive perfect information about school quality, while everyone is perfectly informed in the full information benchmark. Panel A considers an outreach intervention that provides information only to students in the bottom half of middle school performance, while panel B one providing information to Black and Hispanic students with test scores in the top tercile of the 7th grade test score distribution. The last row of each panel also reports the change in the share of Black and Hispanic students receiving offers to the best 20% of schools with respect to the status quo.

Table A.1.20: Best and Worst 5 Letter Rules

	Cutoff percentile				Offered VA change				
	D	C	B	A	All	Low achiev.	High achiev.	Black+Hispanic	White+Asian
					(1)	(2)	(3)	(4)	(5)
Full information					23.4%	15.9%	30.6%	30.0%	11.8%
Naïve	20^{th}	40^{th}	60^{th}	80^{th}	17.2%	11.4%	22.6%	25.1%	3.3%
Best on average	10^{th}	30^{th}	70^{th}	90^{th}	20.1%	13.5%	26.2%	28.4%	5.0%
Worst on average	70^{th}	80^{th}	90^{th}	95^{th}	11.3%	2.2%	20.0%	15.3%	4.5%
Best for low achieving	5^{th}	10^{th}	30^{th}	70^{th}	18.5%	17.0%	20.0%	27.6%	2.0%
Worst for low achieving	70^{th}	80^{th}	90^{th}	95^{th}	11.3%	2.2%	20.0%	15.3%	4.5%
Best for high achieving	10^{th}	40^{th}	70^{th}	90^{th}	20.1%	13.2%	26.4%	28.4%	5.0%
Worst for high achieving	5^{th}	10^{th}	20^{th}	30^{th}	14.2%	15.2%	13.3%	21.0%	2.2%
Best for Black and Hispanic	10^{th}	35^{th}	70^{th}	95^{th}	20.1%	16.1%	23.9%	29.4%	3.5%
Worst for Black and Hispanic	70^{th}	80^{th}	90^{th}	95^{th}	11.3%	2.2%	20.0%	15.3%	4.5%
Best for white and Asian	5^{th}	30^{th}	90^{th}	95^{th}	17.2%	11.4%	22.8%	22.0%	8.9%
Worst for white and Asian	30^{th}	40^{th}	50^{th}	60^{th}	14.2%	11.7%	16.7%	21.5%	1.2%

Notes: This table compares changes in welfare relative to the status-quo in the full-information benchmark (top row) with those induced by information disclosure policies that rate school quality with five letters, from A to F, varying the position of the cutoffs along the quality distribution. Welfare gains are expressed as a percentage of the average first-best achievement gains. The naïve intervention places the cutoffs evenly apart, while the other rating policies reported are those that maximize or minimize the test scores of a given subgroup of students. The unnumbered columns describe the position of the letter cutoffs in terms of value added percentile ranking. The remaining columns report changes in test scores by student subgroup.

A.2 Robustness to Using Alternative Measures of Value-Added

Here I consider alternative ways of measuring school value-added than those used in the main analysis and provide evidence that it makes a little difference for the results of this paper.

A.2.1 Robustness to Using Race-Specific Measures of Value-Added

I relax the assumption that school value-added is constant across students embedded in the model in equation (1.1) and allow school effectiveness to vary by student race as captured by OLS estimates of α_{jr} in the following regression:

$$Y_i = \alpha_0 + \sum_{j=1}^{J} \alpha_{jr(i)} D_{ij} + X_i' \Gamma_{t(i)} + \epsilon_i \tag{A.1}$$

where $r(i)$ indicates student i's binary race (pooling Black and Hispanic student into the “minority" category and white and Asian into the “white" category). I provide evidence that the constant-effect model used in the main analysis is already a good approximation of reality because value-added changes little across student race, therefore using a race-specific measure of value-added would not change the main results.

Figure A.2.12 shows that race-specific measures of VA are highly correlated (correlation coefficient of about 0.7) for both Regents and SAT test scores. Table A.2.21, instead, compares lottery-based tests of bias for these two VA models. The idea behind these tests is to use random variation in school offers embedded in the centralized school match to test whether the VA estimates in (1.1) predict student outcomes (Angrist et al., 2016, 2021, 2022b).

Figure A.2.12: Correlation of Race-Specific Measures of School Quality

(a) Regents VA (b) SAT VA

Notes: This figure shows that race-specific estimates of school value-added are strongly correlated within schools by presenting scatter plots of value-added estimates for white and Asian students (x-axis) against estimates of value-added for Black and Hispanic students (y-axis) for both SAT and Regents test scores.

The forecast test captures the extent to which the estimated value added α_j predicts causal school effectiveness on average. In practice it is conducted by instrumenting estimates of the value added $\alpha_{d(i)} = \sum_j \alpha_j D_{ij}$ of the school where i enrolls (D_{ij} denotes enrollment indicators), with random school lottery offers. That is, it test the null hypothesis that the IV estimate $\hat{\psi}$ of the following second stage equation is equal to 1

$$Y_i = \tau_0 + \psi \alpha_{d(i)} + X_i' \tau + \nu_i \tag{A.2}$$

, meaning that a one-unit increase in α_j translates into a one-unit increase in Y_i. The second stage parameter ψ is often referred to as a forecast coefficient and deviations from the null that $\psi = 1$ are called forecast bias. The omnibus test provides another way of testing the CIA by testing that the regression residuals ϵ_i are unrelated to any randomness in school offers. This is a joint test of l orthogonality restrictions $E[(Z_{il} - p_{il})\epsilon_i] = 0$, one for each of L school lotteries available to the econometrician. Z_{il} indicate offers in lottery l, while p_{il} is an assignment propensity score measuring student i's probability of receiving an offer in lottery l. In practice, these restrictions are tested by asking whether $\tau_1 = ... = \tau_L = 0$ in the residual regression equation:

$$\hat{\epsilon}_i = \tau_0 + \sum_{l=1}^{L} \tau_l Z_{il} + \sum_{l=1}^{L} \mu_l p_{il} + X_i' \Delta + \nu_i \tag{A.3}$$

. Angrist et al. (2016) show how this test can be decomposed into two separate test statistics. The first is equivalent to the one used in the forecast test, while the second is the Sargan LM statistic for a test of 2SLS overidentifying restrictions, which checks whether VAM estimates are equally predictive within every lottery. In practice, in all the tests reported in this appendix, schools are classified into 10 bins defined by deciles of the distribution of the estimated conventional value-added in equation (1.1). The testing equation (A.3) is estimated using bin-level (rather than single-school) offers and propensity scores. Propensity scores at the school level are computed using the method derived in (Abdulkadiroglu et al., 2022) and then aggregated at the bin level taking a sum over the propensity scores of schools in the bin. Offers are random conditional on propensity scores and running variable controls defined and constructed in Abdulkadiroglu et al. (2022).

The lottery-based tests of bias show that measures of VA that do not vary by race ("Pooled VA") have a good predictive validity for student Regents scores of both races. Black and Hispanic students Regents test scores are equally well predicted by the pooled and by the race-specific measures of VA. White and Asian student outcomes are instead better predicted by the race-specific VA (the forecast bias test of the pooled VA rejects the null, unlike the one of the race-specific VA) although the forecast coefficient of the pooled VA is relatively high even for this student subgroup.Moreover, overidentification test results of race-specific VA are similar to those for the pooled VA, which further supports the existence of little heterogeneity in school effects across student races (Angrist et al., 2017). As noted above, SAT OLS VA is instead more biased. Forecast bias tests always reject the null, but this is true regardless of whether VA is estimated by race or on the pooled sample, suggesting that bias in not related to heterogeneity in treatment effects by race.

Finally, I directly show that Black and Hispanic students choose worst schools even when considering measures of race-specific value added, indicating that the reason behind cross-race gaps in choices is not that students are choosing schools that are best for their own demographic group while constant VAM models are failing to capture race-specific school match effects As shown by comparisons of cross-race gaps in the table below with those reported in the main text, if anything, cross-race gaps are larger when considering measures of value-added that vary with student race.

Table A.2.21: Pooled and Race-Specific VAM Bias Tests for Regents and SAT scores

	Pooled VAM			Race-specific VAM		
	All	Black+Hispanic	White+Asian	All	Black+Hispanic	White+Asian
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Regents math VA					
Forecast coefficient	0.966	1.02	0.837	0.968	0.983	0.970
	(0.032)	(0.037)	(0.072)	(0.032)	(0.036)	(0.082)
First stage F statistic	1771	1369	309	1620	1374	223
Bias tests						
Forecast	1.11	0.217	5.19	0.968	0.229	0.136
	[0.292]	[0.641]	[0.023]	[0.325]	[0.632]	[0.712]
Overidentification (9 d.f.)	12.1	13.1	5.32	12.0	13.2	4.90
	[0.208]	[0.158]	[0.805]	[0.213]	[0.154]	[0.843]
Omnibus (10 d.f.)	13.2	13.3	10.5	13.0	13.4	5.03
	[0.213]	[0.207]	[0.396]	[0.225]	[0.201]	[0.889]
N (testing)	49322	35739	13583	49322	35739	13583
N (estimation)	179978	130281	49697	179978	130281	49697
	Panel B: SAT math VA					
Forecast coefficient	0.756	0.736	0.636	0.794	0.821	0.564
	(0.050)	(0.055)	(0.131)	(0.052)	(0.061)	(0.113)
First stage F statistic	1203	1267	121	901	1020	117
Bias tests						
Forecast	23.8	23.3	7.78	15.5	8.63	14.8
	[0.000]	[0.000]	[0.005]	[0.000]	[0.003]	[0.000]
Overidentification (9 d.f.)	6.28	9.83	10.4	6.58	8.70	9.44
	[0.712]	[0.364]	[0.317]	[0.681]	[0.466]	[0.398]
Omnibus (10 d.f.)	30.1	33.1	18.2	22.1	17.3	24.2
	[0.001]	[0.000]	[0.052]	[0.015]	[0.067]	[0.007]
N (testing)	46679	34693	11986	46679	34693	11986
N (estimation)	179978	130281	49697	179978	130281	49697

Notes: This table reports tests for bias in OLS value-added models (VAMs). The pooled VAM uses all students to estimate school value-added as measured by the coefficient α_j in equation (1.1), regardless of their race or ethnicity. The Race-specific VAM instead estimates school value-added on separate sub-samples of students, dividing students according to their race or ethnicity. Both VAMs control for cubic functions of baseline math and ELA scores and indicators for sex, race, subsidized lunch, special education, limited English proficiency, each interacted with application year. Forecast coefficients are from instrumental variables regressions of test scores on VAM fitted values, instrumenting fitted values with binned assignment indicators. Assignments are binned by decile of the estimated conventional VAM. IV models control for propensity scores, running variable controls, and baseline demographics and achievement. Test scores for outcomes and VAMs are standardized to be mean zero and standard deviation one in the student-level test score distribution, separately by year. The forecast bias test checks whether the forecast coefficient equals 1; the overidentification test checks overidentifying restrictions implicit in the procedure used to estimate the forecast coefficient. The omnibus test combines tests for forecast bias and overidentification. Standard errors are reported in parentheses; test p-values are reported in brackets. Different columns use different samples of students for testing: columns (1) and (4) pool all students together, while the remaining columns split students by race.

Table A.2.22: Gap in Choice of School Quality - Robustness to Using Race-Specific Value-Added

	N	Race gap (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable: school race-specific value-added (test score σ)									
Regents VA in top 3 choices	734,854	-0.11***	-0.10***	-0.07***	-0.09***	-0.08***	-0.07***	-0.05***	-0.06***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
SAT VA in top 3 choices	734,853	-0.15***	-0.14***	-0.11***	-0.11***	-0.11***	-0.11***	-0.09***	-0.09***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
borough FE			X				X		
zipcode FE				X				X	
test score controls						X	X	X	X
mean and max in choice-set					X				X

Notes: This table consider whether estimates in table 1.3 are robust to using race-specific estimates of VA. It reports race differences in the quality of school choices as estimated by the coefficient β in equation (1.2), using measures of VA that vary by applicant race. The regressions in the first column correspond to raw race gaps, while columns (2)-(8) progressively add controls for residential location, test scores and quality available in the students' feasible set.

A.2.2 Robustness to Using Risk-Controlled Value-Added

Next, I relax the CIA by estimating risk-controlled (RC) VAM, as introduced by Angrist et al. (2021). RC VAM supplements the vector of controls with applicant characteristics integral to school matching, such as where they apply and the priority status that a school assigns them. This requires restricting the sample to the subset of applicants cohorts for which I have the necessary information to replicate the high school match, that is, starting from 2012 applicants. Because student typically take these SAT tests and Regents exams after their sophomore year, these measures rely on tests taken between 2014 and 2019. As a consequence, RC VAM estimates are not available for a subset of schools in my sample that were phased out before these dates, and rely on a much shorter time span. For these reasons, in the main analysis I rely on OLS VAM estimates of school quality and I provide evidence that conventional and risk-controlled VAM measures in this setting are largely equivalent.

Figure A.2.13 shows that conventional and risk-controlled measures of value added (VA) are highly correlated, with correlation coefficients very close to 1. Table A.2.23 compares lottery-based tests of bias for these two VA models, which confirm that the predictive validity of conventional estimates of Regents VA is incredibly similar to that of risk-controlled measures, and that both are very good. Adding risk-related controls, however, substantially improves the predictive validity of conventional SAT VA measures, that are otherwise substantially more biased. For this reason, I use Regents VA as the primary measure of school quality unless otherwise noted.

Table A.2.23: Conventional and Risk-Controlled VAM Bias Tests for Regents and SAT scores

	Coventional VAM			Risk-controlled VAM		
	All	Black+Hispanic	White+Asian	All	Black+Hispanic	White+Asian
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Regents math VA			
Forecast coefficient	0.966	1.02	0.837	0.927	0.986	0.817
	(0.032)	(0.037)	(0.072)	(0.031)	(0.037)	(0.067)
First stage F statistic	1771	1369	309	1839	1391	351
Bias tests						
Forecast	1.11	0.217	5.19	5.53	0.154	7.48
	[0.292]	[0.641]	[0.023]	[0.019]	[0.694]	[0.006]
Overidentification (9 d.f.)	12.1	13.1	5.32	9.78	5.84	7.75
	[0.208]	[0.158]	[0.805]	[0.369]	[0.756]	[0.560]
Omnibus (10 d.f.)	13.2	13.3	10.5	15.3	5.99	15.2
	[0.213]	[0.207]	[0.396]	[0.121]	[0.816]	[0.124]
N (testing)	49322	35739	13583	49291	35595	13696
N (estimation)	179978	130281	49697	179978	130281	49697
			Panel B: SAT math VA			
Forecast coefficient	0.756	0.736	0.636	0.960	0.925	0.939
	(0.050)	(0.055)	(0.131)	(0.061)	(0.066)	(0.169)
First stage F statistic	1203	1267	121	1007	1069	90.6
Bias tests						
Forecast	23.8	23.3	7.78	0.430	1.30	0.130
	[0.000]	[0.000]	[0.005]	[0.512]	[0.255]	[0.718]
Overidentification (9 d.f.)	6.28	9.83	10.4	9.68	9.82	7.35
	[0.712]	[0.364]	[0.317]	[0.377]	[0.365]	[0.601]
Omnibus (10 d.f.)	30.1	33.1	18.2	10.1	11.1	7.48
	[0.001]	[0.000]	[0.052]	[0.431]	[0.348]	[0.680]
N (testing)	46679	34693	11986	47008	34763	12245
N (estimation)	179978	130281	49697	179978	130281	49697

Notes: This table reports tests for bias in OLS value-added models (VAMs). The conventional VAM controls for cubic functions of baseline math and ELA scores and indicators for sex, race, subsidized lunch, special education, limited English proficiency, each interacted with application year. Risk-only VAM adds propensity score and running variable controls to the uncontrolled specification. RC VAM adds propensity score and running variable controls to the controls in the conventional VAM. Forecast coefficients are from instrumental variables regressions of test scores on VAM fitted values, instrumenting fitted values with binned assignment indicators. Assignments are binned by decile of the estimated conventional VAM. IV models control for propensity scores, running variable controls, and baseline demographics and achievement. Test scores for outcomes and VAMs are standardized to be mean zero and standard deviation one in the student-level test score distribution, separately by year. The forecast bias test checks whether the forecast coefficient equals 1; the overidentification test checks overidentifying restrictions implicit in the procedure used to estimate the forecast coefficient. The omnibus test combines tests for forecast bias and overidentification. Standard errors are reported in parentheses; test p-values are reported in brackets. Different columns use different samples of students for testing: columns (1) and (4) pool all students together, while the remaining columns split students by race.

Figure A.2.13: Correlation of Conventional and Risk-Controlled Measures of School Quality

(a) Regents VA   (b) SAT VA

Notes: This figure shows that conventional and risk-controlled estimates of school value-added are strongly correlated within schools by presenting scatter plots of risk controlled value-added estimates (x-axis) against conventional estimates of value-added (y-axis) for both SAT and Regents test scores.

A.3 Model and Counterfactuals Appendix

A.3.1 Model Identification

Separating Preferences for Quality from Priors This proof is analogous to the argument used in the proof of proposition 1 in Vatter (2022), modified to the case in which quality is scalar and identification comes from changes in letter grades or their absence within schools. In what follows, for simplicity, I focus on variation within a school over time and thus I drop the school subscript j to write: $\delta_t = X_t'\beta + \gamma E[q|s_t = r] + \xi$. The argument developed here can be directly applied to the demand of different demographic cells and for schools (or school types) that receive at least 3 different quality ratings, or 2 ratings and no rating. For simplicity, I also assume no variation over time in X_t to focus only on identification of beliefs from preferences for quality but the argument is easily extended to consider preferences for other time-varying school attributes X_t as long as these are not perfectly collinear with letter grades. Throughout, I assume the identification of $\delta_t = \gamma[q|s_t = r] + \xi$ up to a constant. In this simplifying case in which I have dropped the dependence on X_t, these are effectively only letter grade fixed effects δ_r (including the case of lack of grades) for each school type $h(j)$.

Let f and g be two distinct, strictly positive. densities, supported over $Q = [\underline{q}, \bar{q}]$. Then there exists $\underline{x} < \tilde{x} < \bar{x}$ s.t. ${}_f[x|x \in (\underline{x}, \tilde{x})] \leq_g [x|x \in (\underline{x}, \tilde{x})]$ and ${}_f[x|x \in (\tilde{x}, \bar{x})] \geq_g [x|x \in (\tilde{x}, \bar{x})]$ with one of the inequalities strict.

Also, there exists $x' < x'' \in [\underline{q}, \bar{q}]$ such that ${}_f[x|x \in (x', x'')] =_g [x|x \in (x', x'')]$.

Proof. Because f and g are distinct and continuous over a common support, they must cross at an interior point $\tilde{x} \in (\underline{q}, \bar{q})$. By continuity, $\exists \epsilon > 0$ such that $f(x) > g(x) \forall x \in (\tilde{x}, \tilde{x} + \epsilon)$ and

$f(x) \leq g(x) \forall x \in (\tilde{x}-\epsilon, \tilde{x})$ where there role of f and g is wlog. Define $h_f(x,\epsilon) = \frac{f(x)}{F(\tilde{x}+\epsilon)-F(\tilde{x})}$ and analogously for g, where F and G denote the cumulative distribution functions of f ad g respectively. Both $h_f(\cdot,\epsilon)$ and $h_g(\cdot,\epsilon)$ are continuous and integrate to 1 over $(0,\epsilon)$ and therefore intersect at an interior point in $(0,\epsilon)$. Moreover, note that $\forall \tilde{\epsilon} \in (0,\epsilon)$, $h_f(\cdot,\tilde{\epsilon}) < h_g(\cdot,\tilde{\epsilon})$. Pick $\tilde{\epsilon} \in (0,\epsilon)$ such that $h_f(\cdot,\tilde{\epsilon})$ and $h_g(\cdot,\tilde{\epsilon})$ intersect only once at a point $\hat{x}$ and denote $\bar{x} = \tilde{x} + \tilde{\epsilon}$. Then we have that $E_f[x|x \in (\tilde{x},\bar{x})] - E_g[x|x \in (\tilde{x},\bar{x})]$:

$$\begin{aligned}\int_{\tilde{x}}^{\bar{x}} (h_f(q,\tilde{\epsilon}) - h_g(q,\tilde{\epsilon}) q dq) &= \int_{\tilde{x}}^{\hat{x}} \underbrace{(h_f(q,\tilde{\epsilon}) - h_g(q,\tilde{\epsilon})}_{<0} q dq) + \int_{\hat{x}}^{\bar{x}} \underbrace{(h_f(q,\tilde{\epsilon}) - h_g(q,\tilde{\epsilon})}_{>0} q dq) \\ &> \hat{x}\Big[\int_{\tilde{x}}^{\hat{x}} (h_f(q,\tilde{\epsilon}) - h_g(q,\tilde{\epsilon}) dq) + \int_{\hat{x}}^{\bar{x}} (h_f(q,\tilde{\epsilon}) - h_g(q,\tilde{\epsilon}) dq)\Big] = 0\end{aligned}$$

which proves the first inequality. The proof for the second part of the statement is analogously applied to $(\tilde{x}-\epsilon, \tilde{x})$. The proof of the equality follows from defining $w(\lambda) =_f [x|x \in (\underline{x}, \tilde{x}+\lambda\tilde{\epsilon})] -_g [x|x \in (\underline{x}, \tilde{x}+\lambda\tilde{\epsilon})$. Because $w(0) \leq 0$, $w(1) > 0$ and $w(\cdot)$ is a continuous function, by the intermediate value theorem, $\exists \lambda^*$ such that $w(\lambda^*) = 0$ and thus the interval cutoffs for the second part are $x' = \underline{x}$ and $x'' = \tilde{x} + \lambda^* \tilde{\epsilon}$. □

(Variation in signals) Let r_t be the quality rating received by the school in year t, which is defined by a compact interval of possible quality values $[\underline{c}_t^r, \bar{c}_t^r]$. In the absence of letter grades, this is degenerate and coincides with the entire quality space $Q = [\underline{q}, \bar{q}]$. The possible quality partitions r_t are drawn from a distribution over intervals of quality and at least $N \geq 3$ are observed in the data.

(Knowledge of the rating design) Consumers know the quality cutoffs $\underline{c}_t^r, \bar{c}_t^r$ that define the quality rating s_t. Because s_t may also include a degenerate signal, i.e. its absence, this means they also know the boundaries of the quality space Q. Moreover, they use the ratings and Bayes' rule to update a continuous prior density $f: Q \rightarrow R+$.

Let assumptions 3 and 4 hold and let student preferences over quality be linear, i.e., $v(q) = \gamma q$. Then $(\gamma, f(\cdot))$ are identified.

Proof. By contradicion, suppose there exist two distinct elements in the identified set $(\gamma_0, f_0), (\gamma_1, f_1)$. By the above lemma and assumption 2, there exists three possible quality ratings r, r' and r'' which may be drawn with positive probability, such that: $_{f_0}[q|r] =_{f_1} [q|r]$, $_{f_0}[q|r'] <_{f_1} [q|r']$ and $_{f_0}[q|r''] \geq_{f_1} [q|r'']$. Therefore, we have that:

$$\gamma_0(_{f_0}[q|r'] -_{f_0} [q|r]) = \delta_{r'} - \delta_r = \gamma_1(_{f_1}[q|r'] -_{f_1} [q|r]) \implies \gamma_0 > \gamma_1$$

$$\gamma_0({}_{f_0}[q|r''] - {}_{f_0}[q|r]) = \delta_{r''} - \delta_r = \gamma_1({}_{f_1}[q|r''] - {}_{f_1}[q|r]) \implies \gamma_0 \leq \gamma_1$$

, a contradiction.

□

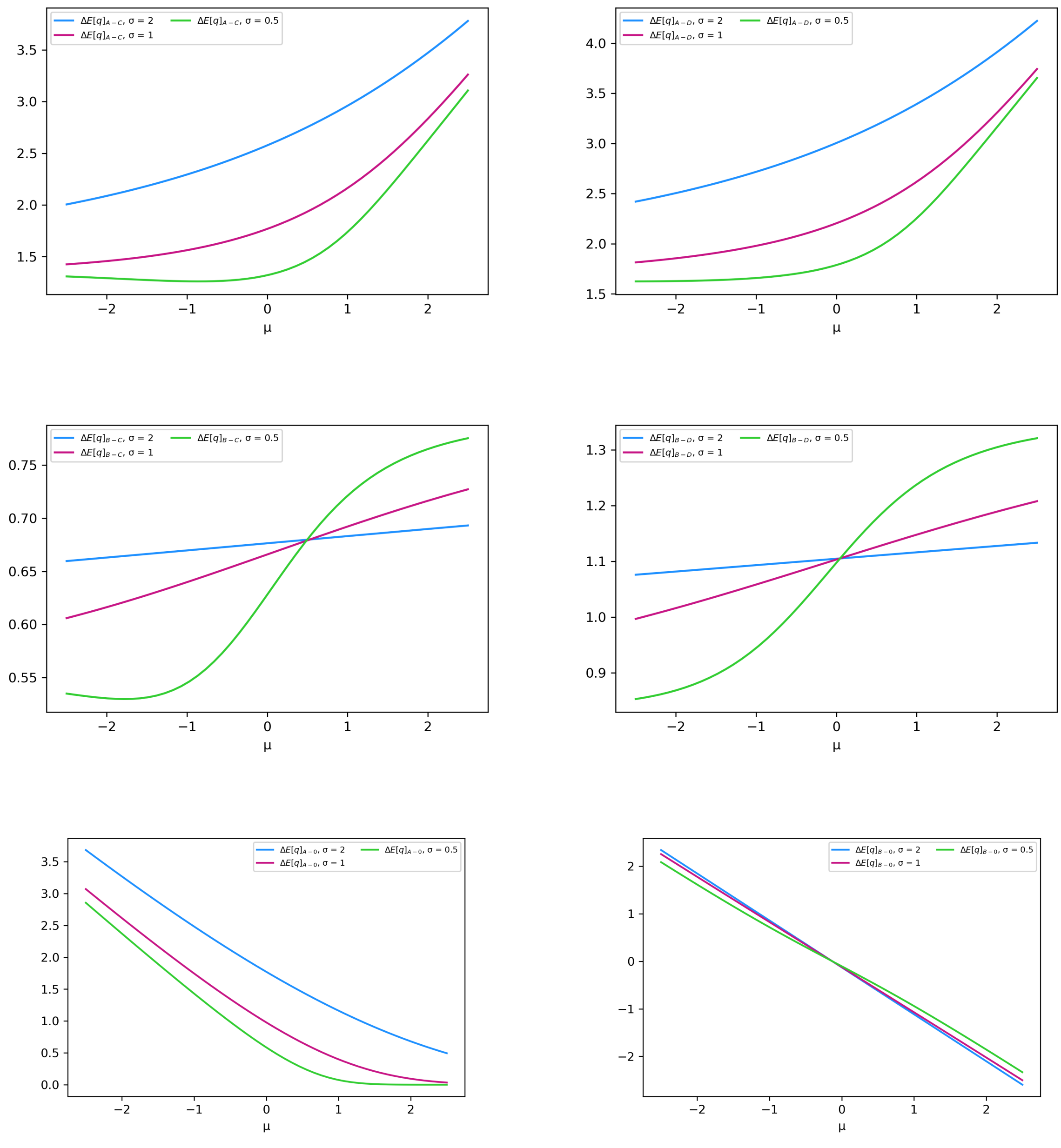
In theory, point identification without a functional form restriction for priors requires a lot of variation in partitioning cutoffs of ratings r_t over the years. In practice, in my case, quantiles used to assign letter grades vary little from year to year, but I still have variation from the presence or absence of letter grades, which is equivalent to a degenerate partitional design. Moreover, the fact that the same school may receive different grades in different years and the normal functional form assumption together allow to identify the model parameters.

To build intuition for how this argument is used in practice given the observed letter grades cutoffs, figure A.3.14 below plots the change in expected quality for pairs of different letter grades as a function of the prior parameters μ and σ. While the positive updating in beliefs between receiving a D and and A is always larger the larger the variance of the prior, σ, this is no longer the case when looking at the belief updating between receiving a D and a B. This variation is what allows to separate the prior precision from preferences for quality. Changes in demand between years with and without letter grades help pin down the prior mean μ , as they vary most strongly in this parameter as compared to the precision of the prior. Intuitively, if following the removal of positive (negative) signals the fall (increase) in utility is larger, this means that the mean prior of quality was lower (higher), as people were more surprised by the positive quality signals. Changes in demand when the school receives different letter grades, help identifying preferences for quality and prior uncertainty.

School Quality Distribution I assume that each school in the city has quality q_j that is fixed over time and that is not observed by students. For the empirical estimation I re-center value added around the city-wide average without loss of generality, since value added is always measured relative to the value-added of some arbitrarily picked school. Moreover, I rescale it by its across-school standard deviation which simply changes the interpretation of the preference parameter γ to measure willingness to travel for 1 additional standard deviation of quality in the cross-school distribution. Denoting with $[q, \bar{q}]$ the space of possible quality values observed in the city, the empirical distribution of quality in NYC is well approximated by a truncated normal distribution over $[q, \bar{q}]$.[1] Given the normalization, I have that the mean and standard deviation

[1]While the number of schools and the distribution of quality in the city are naturally discrete, this continuous approximation is a convenient simplifying assumption because it allows to work with continuous probability densities. This assumption is more innocuous in a setting like NYC, in which the number of high schools families can choose from is large (~400) compared to the average public school district in the U.S.

Figure A.3.14: Belief Updating as a Function of Prior Parameters and Changes in Letter Grades



Notes: This figure plots the change in expected school quality $\Delta E[q]_{s'-s''}$ due to a change in quality signal from s'' to s' as a function of the prior mean μ and variance σ. $s = 0$ denotes the absence of letter grades.

of the quality distribution are $\mu_q = 0$ and $\sigma_q = 1$.

Figure A.3.15 shows that the normal functional form is a very good approximation of reality. The figure compares the distribution of value added in the city (standardized across schools) and the distribution of the standardized quality scores underlying letter grades used in one random year against the probability density function of a standard normal. The three densities are similarly bell-shaped and approximately symmetric.

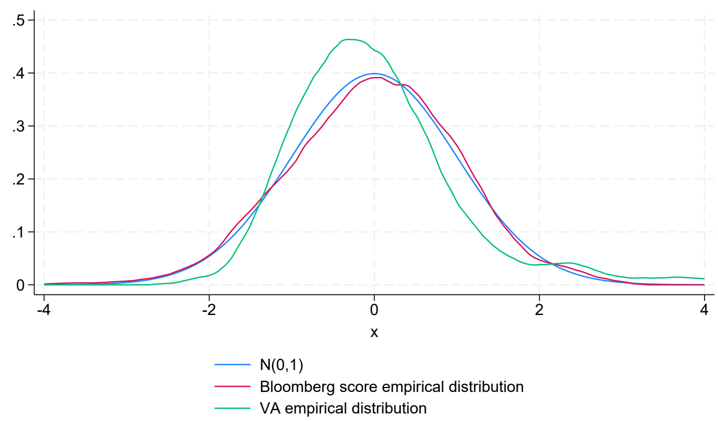
The point of the figure is to show that, while the underlying quality score was only imperfectly correlated with value added, quantiles of the two measures correspond roughly everywhere. This is important in light of my assumption that students observed quantile cutoffs of the quality score distribution used to assign letter grades and that they believed these were the cutoffs of the underlying value-added distribution. Because the two coincide in practice, this makes it easier to believe the assumption and also easier to argue that changes in demand following letter signals based on quantiles of the quality score can be more easily generalized to counterfactual scenarios in which signals of quality are based on value added. The fact that quantiles of the two distributions coincide (and that their shape is also similar to the normal form that I picked for student priors) is nice but not necessary for my estimation argument to be valid. Under my assumption, I would still be measuring preferences for value-added even if the quantiles of the two distributions were different.

Additionally, the fact that the empirical distribution of quality is approximately normal motivates the adoption of a truncated normal functional form for student priors. Priors are assumed to be normally distributed over the possibly quality space but are not constrained in any additional way, that is, they may be incorrect on average. This allows students to believe that certain schools are of better or worse quality than the average school and be more or less certain about that, but the shape of their beliefs is still realistic, in the sense that it is similar to what is observed in the city.

A.3.2 Counterfactual Simulations Details

In all simulations, I restrict the pool of general education 9th grade applicants to students for whom I know baseline test scores, residential address and race. These restrictions drop 20% of the applicant pool. I restrict capacities of all schools uniformly by this same factor. I simulate choices at the school rather than at the program, since all observable schools characteristics are measured at this level, and I use the admission rules (priority design and screening method) of the school largest program, in case the school has more than one, which is the case for approximately 30% of the schools. I am able to reconstruct student priorities and ranking

Figure A.3.15: Comparison of Quality Distribution Relative to a Standard Normal



Notes: This figure plots the distribution of standardized value-added in NYC high schools (green line), of the standardized quality score used in the last year of the letter grade policy (pink line) and the distribution of a standard normal (blue line).

at all schools based on student residential address, home language, ell status, middle school test scores, standardized state test scores and the middle school the student attended. Several schools give higher priority to students attending an information session at the school, which is a piece of information I do not have and that I cannot counterfactually estimate. Therefore the simulated priorities abstract from this.

Finally, I keep students' lottery number and preferences shocks constant across counterfactual simulations. The length of the rank ordered list is model-driven and coincides with all schools with utility above zero, the utility of the outside option. This typically results in longer lists, which explains why in my counterfactual simulations the probability of receiving an offer in the first round is slightly larger (by approximately 2 p.p.) than what is observed in reality.

A.4 Data Appendix

A.4.1 NYCDOE Data

Lists of high school applicants, their rank ordered lists, priorities, lottery numbers, and assignments are constructed from annual records from the New York City Department of Education (NYCDOE) school assignment system. Information on student demographic characteristics and schools attended comes from the NYCDOE's Office of School Performance and Accountability (OSPA). Baseline middle school achievement in taken from the New York State Assessment. High school achievement outcomes come from Regents exams, graduation, SAT and college records that originate with different sources, and are collected by the NYCDOE. Geographic information on students comes from Zoned DBN data. All these data files were provided by NYCDOE. They include a unique student identification number that links records across files. More details on each data source are provided below.

NYCDOE Assignment Data Data on NYC high school applications are maintained by the Student Enrollment Office of the NYCDOE. I received all applications for the 2006-07 through 2018-2019 school years. Application records include students' rank-order lists of academic programs submitted in each round of the application process, eligibility, priority group and rank at each program listed, the admission procedure used at the respective program, and the program to which the applicant was assigned. Lottery numbers and details on assignments at Educational Option (Ed-Opt) programs are provided in separate data sets only for the high school match of 2012 to 2017. The NYC high school match is conducted in three rounds, and separately for 9th grade and 10th grade seats and for general education and special education seats. I focus on first-time applicants to general education 9th grade seats using data from the first round.

OSPA Data I received registration and enrollment files for the 2005-06 through 2018-2019 school years from NYCDOE's Office of School Performance and Accountability (OSPA). These data include every student's grade and school District Borough Number (DBN), as of June of each school year, as well as information on student demographic variables. I use this file to code school enrollment, special education status, subsidized lunch status, and limited English proficiency.

New York Regents Exam Data Regents Examinations are statewide-standardized exams used to determine the type of New York State Diploma students are eligible for and more broadly to determine graduation eligibility. I received data on all Regents examinations conducted between the 2005-2006 and the 2020-2021 school years by students enrolled in a NYC public school. These years cover all high school applicant cohorts in my analysis sample, since most students take these tests before or during their junior year. I use the first test in each subject for multiple takers. I only consider tests taken in the subjects of ELA and Algebra 1 (or corresponding denominations that may vary slightly over the years). During my sample period, Regents in ELA and Math have been redesigned. In the 2013-24 SY NYC began administering the new test, designated as Common Core aligned Regents. During the first year the Common Core was rolled out, students were allowed to take the old and the new test and the higher of the two was counted for grading and other purposes. I adopt the same convention for students taking both tests during the transition period. Scores are then normalized to have mean zero and standard deviation one within a subject-year. A very small subset of students takes the ELA test during 8th grade. I only keep records for tests take after high school.

SAT Data I received data on SAT scores for tests conducted between the 2006-2007 and the 2020-2021 school years by students enrolled in a NYC public school. These years should cover all cohorts in my analysis sample, since most students take these tests before or during their junior year. These data originate with the College Board and but are provided by the NYCDOE. I use the first test for multiple takers. For applicants tested in the same month, I use the highest score. During my sample period, the SAT has been redesigned. I re-scale scores of SAT exams taken prior to the reform according to the official re-scaling scheme provided by College Board. Scores are normalized to have mean zero and standard deviation one within a subject-year among all students taking SAT in that year.

Graduation Data The DOE Graduation file records the discharge status for public school students enrolled from the 2005-2006 to 2020-2021 SY. I do not have graduation records for the last cohort of applicants in my sample, who were expected to graduate during the 2021-2022 SY. These records are used to construct a dummy indicating students graduating within 4 years of their enrollment in 9th grade.

College Data College enrollment data are generated by the DOE School Performance office based on data collected from the National Student Clearinghouse (NSC). They contain one record per student from the graduating cohort of that school-year, indicating whether a student enrolled in college in the fall that immediately followed their on-time graduation. I received data covering years from 2005 to 2020. These would cover students graduating on time up to the 2016 cohort included. Records before 2016 do not distinguish between 2 year and 4 years higher education institutions. For this reason, I use this data to construct an indicator for enrollment in any college in the fall that immediately followed their expected on-time graduation.

New York State Assessment Data The New York State Assessment is the standardized state exam for New York, taken in grades 3-8. The NYCDOE provided scores for students taking the exam from the 2005-06 to 2018-19 school years. Each observation in the dataset corresponds to a single test record. I use 7th grade test scores from the 2003-04 to the 2017-18 SY to assign baseline math and English Language Arts (ELA) scores. Baseline scores are normalized to have mean zero and standard deviation one within a subject-year among all 7th grade NYC public school students taking the test.

Zoned DBN Data The Zoned DBN dataset provides geographic data for elementary, middle, and high school students in NYC based on the address provided to the DOE. In these files, there is a record for every student with an active address record during the school year. I use Zoned DBN files of school years from 2007-08 to 2018-19 to collect data on student residential districts and census tracts at the time of high school application.

A.4.2 Public Data

Commuting Time I collect commuting times between a student's residence and a school, estimated using the HERE Public Transit API. They are given by public transit travel time made of the shortest combination of walking, local and express bus, and subway modes, setting an arrival time of 8:00AM on September 9th, 2020. Residential addresses are approximate and given by the centroid of the census tract of residence.

High School Directories I collect the pdfs of the printed high school directories used by applicant cohorts of 2006-2018 and the corresponding digitized versions. I use this data to understand the information displayed on each directory and as a basis for my school-year panel.

School Progress Reports I obtain publicly available data included in the NYC DOE School Progress Reports rating school performance in the school years from 2006-07 to 2013-14 from the NYC Open Data website.[2] These files include the overall grade and quality score a school

[2]https://opendata.cityofnewyork.us/

received, as well as grades and sub-scores for each of the components of the overall quality score (e.g. school environment score, school progress score etc.) Grades referring to a school year (e.g. 2006-07) were typically made public during the following year (e.g. 2007-08).

A.4.3 Survey Data

In February and March of 2023, we surveyed parents and guardians of students who had applied to 9th grade seats during the 2023-2024 NYC high school admission cycle, in partnership with the NYCDOE. A more extensive analysis of the survey data is presented in a companion paper (Corradini and Idoux, 2023). The survey was designed to be sent after families applied to high school but before the offers were sent out. The timeline allowed parents to have at least two weeks to complete it, and the survey had no time constraints beyond this deadline. Incomplete surveys were automatically submitted by the deadline. Participants who answered at least one question by the deadline received a $10 Amazon gift card. The survey was sent electronically using the email addresses of families used in the high school application process in the top three most spoken languages in NYC: English, Spanish, and Simplified Chinese. The survey was designed on Qualtrics and it was available in those same three languages. All questions were marked as optional, except the consent to participate which included declaring to be older than 18.

The survey was sent to 21,401 families. This sample consisted of a subset of parents or guardians of students applying to start 9th grade in fall 2023 who satisfied the following conditions: 1) they had been enrolled in a NYC middle school since 6th grade 2) they had non-missing demographic records within the NYCDOE database and 3) they had taken the New York State Assessment standardized test in 4th grade and we observed that in our records. We received 3,628 responses (17% response rate).

Table A.4.24 compares descriptive statistics of survey receivers (columns 2-3) and respondents (column 4-7) to those of NYC applicants (column 1). Respondents were slightly more likely to be white and Asian, less likely to be eligible for a subsidized lunch and had students with higher baseline achievement, compared to the sample of eligible families. We only use responses of students completing at least 50% of the survey (descriptives in column 5).

To reduce the time it takes to complete the survey and increase participation, we devised eight different survey versions by creating different combinations of question subsets. Eligible participants were shown one version of the survey at random. In this paper , I only use the answers to one question asking respondents to situate a school in the distribution of school quality of their residential borough. The text of the question asked:“How well does *school name*

Table A.4.24: Descriptive Statistics of Survey Receivers and Respondents

	Applicants with baselines			Respondents			
	NYC	Survey receivers	Receiving belief Q.	All	Answers > 50%	Ans. > 50% and gets belief Q	Ans. > 50% and ans. belief Q
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Asian	20%	25%	25%	29%	29%	28%	29%
Black	19%	16%	16%	14%	14%	13%	13%
Hispanic	42%	39%	39%	33%	32%	32%	30%
White	16%	17%	17%	21%	22%	23%	25%
Subsidized Lunch	76%	73%	73%	66%	64%	64%	62%
Brooklyn	29%	31%	31%	31%	31%	33%	32%
Queens	32%	35%	34%	37%	36%	35%	36%
Manhattan	10%	9%	9%	10%	11%	12%	12%
Bronx	21%	18%	18%	14%	14%	14%	12%
7th grade Math	0.11	0.34	0.33	0.55	0.58	0.59	0.61
7th grade ELA	0.11	0.35	0.35	0.51	0.54	0.55	0.58
N	47618	21401	7946	3628	3099	1142	849

Notes: This table provides descriptive statistics about the sample of households receiving (column (2)) and responding (column (3)) to our survey. Column (1) compares their characteristics to the entire sample of 9th grade applicants from which we drew the sample of survey receivers. Column (5) restricts the sample of respondents to those answering at least 50% of the survey questions, column (6) further conditions on respondents receiving the question about beliefs used in this paper and column (7) conditions on having responded to the question, which is the sample used in this paper.

- *(school code)* prepare students for their Regents exams compared to other schools in your borough?". Answers could vary from 1 to 4, with 1 corresponding to the worst 25% of schools and 4 to the best 25%. By design, 37.5% of the survey participants received this question in their questionnaire. Columns 6 and 7 restrict the sample of respondents to those receiving the question I study (N=1,142) and finally answering it (N=849). This last sub-sample is what I use in the analyses in section 2.2.2.

The schools that populated the question varied across respondents. Schools were assigned to respondents at random subject to a set of constraints. The set of high schools eligible for inclusion was determined on the basis of their proximity to the student's address and other criteria as follows:

1. we started from schools existing in the 2021-2022 high school directory, dropped specialized schools, special districts (75 and 79), and home schools and keep only schools participating in the 2023 high school match.

2. For each of the 32 residential districts, we took all schools that are located in the district borough

3. We then restricted to fairly popular schools in the district, as indicated by the fact that

they were ranked by at least 5% of students in the district.

This returns, on average, 55 schools per district. A school in this subset is ranked, on average, by 11% of students residing in the district. For each district, we then selected at random 4 schools within this subset, subject to the constraint that each of the four school corresponded to a possible combination of two dummies. The first dummy indicated schools enrolling a high share of white and Asian students (above 26% of white and Asian students, corresponding to the 25% most white schools). These are schools typically also enrolling higher achieving students. The second indicated schools with Regents value-added above the median in the borough. For the purpose of designing the question, I classify a school as having above median Regents value-added if it is above the median for both Regents Math and Regents ELA.

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

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

Each respondent was randomly assigned one of the four possible schools selected for her district. The balance table below confirms that the characteristics of schools assigned to the questionnaires did not differ by respondent race. It includes regression estimates of school characteristics on respondent's race, controlling for district fixed effects. The coefficient on respondent's race is never statistically different from zero.

Table A.4.25: Balance of School Characteristics Across Respondent Race

	Regents VA (SD)		Average Regents (SD)	
	(1)	(2)	(3)	(4)
Respondent is white or Asian	-0.003	-0.039	0.040	0.042
	(0.109)	(0.053)	(0.094)	(0.045)
Average Regents (SD)		0.912***		
		(0.024)		
Regents VA (SD)				0.734***
				(0.020)
N	849	849	849	849
Dep. var. mean	0.141	0.141	0.467	0.467
Dep. var s.d.	1.147	1.147	1.162	1.162

Notes: This table shows that the quality and mean achievement levels of the schools populating the survey questions are balanced across respondent's race. The table reports estimates of a regression of school characteristics on a dummy indicating whether respondents are white or Asian. Columns (2) and (4) also control for the school attribute not used in the left hand side (e.g. school achievement levels in the regression with school quality as a dependent variable) to check balance also conditional on other school attributes.

Appendix B

Appendix to Chapter 2

B.1 Appendix Tables and Figures

Table B.1.1: Sample Construction

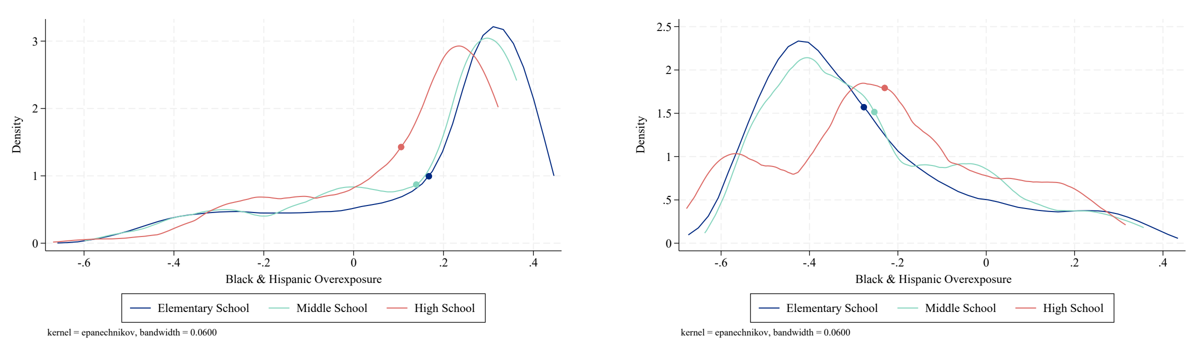
	N
Non-sped 6th grade applicants with demographics	255,876
Who enroll in 6th grade in the district	255,574
Who apply to 9th grade in the district	227,149
Ranked at least 2 programs, first is oversubscribed	153,193
Has non-degenerate risk of school assignment	88,471
Who are offered a 6th grade seat	73,957
Who have baseline scores	69,336

Notes: The initial administrative data sample includes students who applied for middle school enrollment in the school years from 2015-2016 to 2019-2020.

Figure B.1.1: Differences in Exposure to Black & Hispanic Peers by Race and School-Level

(a) Black and Hispanic students

(b) white and Asian students



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

Table B.1.2: Survey Summary Statistics

	Applicants with baselines			Respondents			
				Answered > 0%	Answered > 50%		
	NYC (1)	Wave 2 (2)	Wave 3 (3)	All (4)	All (5)	Wave 2 (6)	Wave 3 (7)
Asian	20%	27%	22%	29%	30%	31%	28%
Black	19%	15%	17%	14%	14%	12%	15%
Hispanic	42%	39%	39%	33%	31%	32%	31%
White	16%	16%	19%	21%	22%	22%	23%
Poverty	76%	73%	74%	66%	63%	61%	65%
Brooklyn	29%	33%	28%	31%	31%	33%	28%
Queens	33%	36%	32%	37%	37%	38%	34%
Manhattan	10%	7%	11%	10%	11%	9%	15%
Bronx	21%	16%	20%	14%	13%	11%	17%
Math 4^{th}	.06	.32	.29	.51	.56	.56	.55
ELA 4^{th}	.02	.28	.28	.46	.51	.52	.49
N	47,618	11,415	9,986	3,628	2,927	1,830	1,097
Response %				17%	14%	16%	11%

Notes: This table reports summary statistics for students applying to enroll in high school in the fall of 2023. The first column restricts the sample to high school applicants with non missing baseline demographics and achievement outcomes, while the second and third column to applicants selected to receive our survey. Columns (4) to (7) report summary statistics for the survey respondents.

Table B.1.3: Differences in Information Accuracy by Race

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

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

Table B.1.4: Attrition and Covariate Balance - Discrete Treatment

	Offered majority white+Asian MS				
	All	White	Minority	Black	Hispanic
Dependent variable	(1)	(2)	(3)	(4)	(5)
Panel A: Attrition					
Enrolls in district (6th grade)	0.000	-0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
mean	1.00	1.00	1.00	1.00	1.00
N	77,640	27,561	48,673	17,659	31,014
Has 9th grade application	-0.003	-0.015	0.007	0.011	0.010
	(0.009)	(0.012)	(0.013)	(0.023)	(0.015)
mean	0.89	0.89	0.89	0.89	0.89
N	77,571	27,554	48,612	17,635	30,977
Panel B: Covariates balance					
Black	0.005		0.004		
	(0.011)		(0.019)		
mean	0.22		0.36		
Hispanic	0.014		-0.004		
	(0.013)		(0.019)		
mean	0.40		0.64		
White+Asian	-0.023**				
	(0.011)				
mean	0.36				
Female	0.009	-0.015	0.024	0.013	0.029
	(0.015)	(0.022)	(0.021)	(0.034)	(0.027)
mean	0.51	0.50	0.53	0.53	0.53
English Language Learner	-0.000	-0.006	0.003	-0.004	0.006
	(0.007)	(0.009)	(0.011)	(0.009)	(0.017)
mean	0.06	0.04	0.07	0.01	0.11
FRPL	0.001	0.012	-0.018	-0.014	-0.019
	(0.012)	(0.020)	(0.015)	(0.028)	(0.019)
mean	0.75	0.62	0.84	0.80	0.86
Baseline English	0.031	0.031	0.034	-0.015	0.076*
	(0.023)	(0.032)	(0.033)	(0.054)	(0.043)
mean	0.21	0.61	-0.04	-0.02	-0.04
Baseline Math	-0.004	0.033	-0.031	-0.031	-0.031
	(0.023)	(0.031)	(0.032)	(0.053)	(0.042)
mean	0.20	0.76	-0.12	-0.18	-0.09
N	69,336	25,040	43,043	15,285	27,758

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

Table B.1.5: Attrition and Covariate Balance - Continuous Treatment

	Offered 10pp more White+Asian				
	All	White	Minority	Black	Hispanic
Dependent variable	(1)	(2)	(3)	(4)	(5)
Panel A: Attrition					
Enrolls in district (6th grade)	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
mean	1.00	1.00	1.00	1.00	1.00
N	77,640	27,561	48,673	17,659	31,014
Has 9th grade application	0.003	-0.000	0.003	0.003	0.004
	(0.002)	(0.003)	(0.002)	(0.004)	(0.003)
mean	0.89	0.89	0.89	0.89	0.89
N	77,571	27,554	48,612	17,635	30,977
Panel B: Covariates balance					
Black	-0.002		-0.002		
	(0.002)		(0.004)		
mean	0.22		0.36		
Hispanic	0.000		0.002		
	(0.003)		(0.004)		
mean	0.40		0.64		
White+Asian	0.000				
	(0.002)				
mean	0.36				
Female	0.004	-0.000	0.005	-0.006	0.011**
	(0.003)	(0.005)	(0.004)	(0.007)	(0.005)
mean	0.51	0.50	0.53	0.53	0.53
English Language Learner	-0.000	-0.003	0.002	-0.002	0.003
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
mean	0.06	0.04	0.07	0.01	0.11
FRPL	-0.008***	-0.003	-0.012***	-0.010*	-0.013***
	(0.002)	(0.005)	(0.003)	(0.005)	(0.003)
mean	0.75	0.62	0.84	0.80	0.86
Baseline English	0.005	0.015*	-0.003	-0.019*	0.008
	(0.005)	(0.007)	(0.006)	(0.010)	(0.008)
mean	0.21	0.61	-0.04	-0.02	-0.04
Baseline Math	-0.002	0.014**	-0.014**	-0.019*	-0.011
	(0.005)	(0.007)	(0.006)	(0.010)	(0.008)
mean	0.20	0.76	-0.12	-0.18	-0.09
N	69,336	25,040	43,043	15,285	27,758

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

Table B.1.6: OLS Estimates of Peer Effects on Students' Top 3 High School Choices

	% Black (1)	% Hispanic (2)	% white (3)	Peer Ela (4)	Peer Math (5)	Popularity (6)	Screened (7)	SAT math VA (8)	Length of rol (9)
					Panel A: Black & Hispanic students				
Majority white+Asian MS	-5.773***	-3.407***	9.295***	0.111***	0.138***	0.323**	-0.036**	0.118***	-0.161
	(0.638)	(0.695)	(0.675)	(0.013)	(0.012)	(0.100)	(0.017)	(0.022)	(0.201)
Share white+Asian (10pp)	-1.522***	-0.977***	2.488***	0.032***	0.036***	0.105***	0.008**	0.040***	-0.106**
	(0.118)	(0.104)	(0.134)	(0.002)	(0.003)	(0.026)	(0.003)	(0.005)	(0.041)
mean	24.70	45.85	27.70	-0.02	0.04	3.55	0.60	0.35	8.23
N	43,042	43,042	43,042	43,042	43,042	42,731	43,042	43,038	43,042
					Panel B: White & Asian students				
Majority Black+Hispanic MS	4.455***	2.820***	-7.552***	-0.087***	-0.108***	-0.024	0.044***	-0.118***	0.352*
	(0.479)	(0.681)	(0.808)	(0.015)	(0.018)	(0.102)	(0.013)	(0.029)	(0.185)
Share Black+Hispanic (10pp)	1.379***	1.097***	-2.520***	-0.027***	-0.037***	-0.058**	0.003	-0.035***	0.115**
	(0.100)	(0.133)	(0.176)	(0.002)	(0.003)	(0.022)	(0.003)	(0.005)	(0.041)
mean	13.44	26.48	58.10	0.38	0.48	5.10	0.85	1.01	7.58
N	25,040	25,040	25,040	25,040	25,040	25,014	25,040	25,022	25,040

Notes: This table reports OLS estimates of middle school demographic composition effects on the characteristics of top 3 high school choices. Panel A focuses on Black and Hispanic applicants, while panel B on White and Asian applicants. The sample and controls are as defined in the notes of Table 2.5. Standard errors clustered at the Middle school × year level in parenthesis.

Table B.1.7: 2SLS Estimates of Peer Share Effects on Black Students' High School Choices

	% Black (1)	% Hispanic (2)	% white (3)	Peer Ela (4)	Peer Math (5)	Popularity (6)	Screened (7)	SAT math VA (8)	Length of rol (9)
					Top 3 choices				
Majority white+Asian MS	-2.650	-1.809	4.760**	0.080**	0.137**	0.440	-0.098	0.106	-0.171
	(2.181)	(1.509)	(2.414)	(0.040)	(0.045)	(0.319)	(0.063)	(0.071)	(0.540)
Share white+Asian (10pp)	-1.211**	-0.029	1.268**	0.013*	0.020**	0.148**	0.006	0.028**	-0.051
	(0.461)	(0.280)	(0.455)	(0.008)	(0.008)	(0.071)	(0.011)	(0.014)	(0.101)
mean	34.00	37.11	26.88	-0.02	0.02	3.18	0.62	0.29	8.43
					All choices				
Majority white+Asian MS	-2.274	-2.781**	5.257**	0.087**	0.130***	0.433*	-0.004	0.133**	
	(1.687)	(1.336)	(1.873)	(0.036)	(0.039)	(0.252)	(0.040)	(0.059)	
Share white+Asian (10pp)	-0.933**	-0.384*	1.327***	0.018**	0.023**	0.146**	0.013	0.031**	
	(0.367)	(0.232)	(0.343)	(0.007)	(0.007)	(0.052)	(0.009)	(0.012)	
mean	34.61	38.51	24.90	-0.09	-0.05	2.81	0.81	0.20	
N	15,284	15,284	15,284	15,284	15,284	15,210	15,284	15,283	

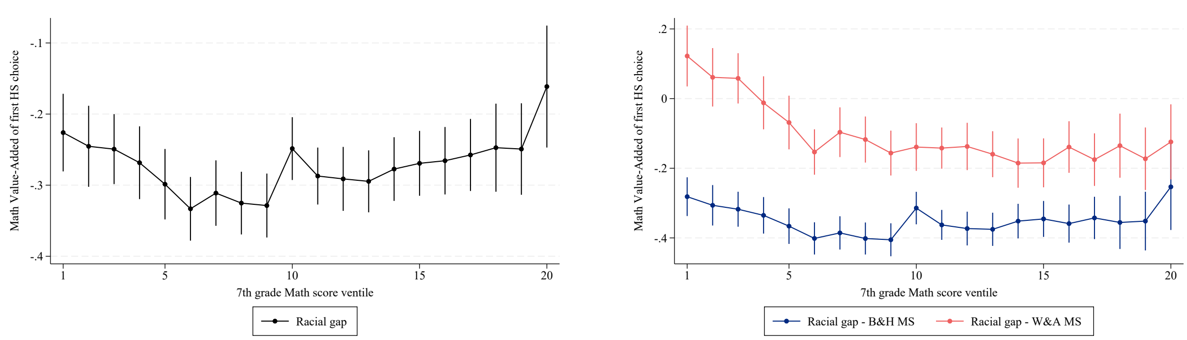
Notes: This table reports 2SLS estimates of middle school demographic composition effects on Black students' high school choices. Panel A focuses on each applicant's top 3 choices, panel B includes all the choices. The sample, controls and endogeneous variables are as defined in the notes of Table 2.5. Standard errors clustered at the Middle school × year level in parenthesis.

Table B.1.8: 2SLS Estimates of Peer Share Effects on Hispanic Students' High School Choices

	% Black (1)	% Hispanic (2)	% white (3)	Peer Ela (4)	Peer Math (5)	Popularity (6)	Screened (7)	SAT math VA (8)	Length of rol (9)
					Top 3 choices				
Majority white+Asian MS	-3.630***	-2.294*	6.178***	0.082**	0.147***	0.133	0.048	0.133*	0.035
	(0.995)	(1.382)	(1.806)	(0.038)	(0.043)	(0.327)	(0.045)	(0.072)	(0.482)
Share white+Asian (10pp)	-0.604**	-0.440	1.047**	0.019**	0.025**	0.077	0.012	0.015	-0.023
	(0.203)	(0.281)	(0.332)	(0.007)	(0.008)	(0.074)	(0.008)	(0.014)	(0.086)
mean	19.59	50.67	28.15	-0.01	0.05	3.76	0.58	0.38	8.12
					All choices				
Majority white+Asian MS	-2.513**	-1.393	4.255**	0.059**	0.114**	0.053	0.023	0.104**	
	(0.942)	(0.961)	(1.333)	(0.030)	(0.036)	(0.277)	(0.040)	(0.052)	
Share white+Asian (10pp)	-0.371**	-0.276	0.665**	0.012**	0.016**	0.027	0.006	0.015	
	(0.185)	(0.210)	(0.257)	(0.005)	(0.006)	(0.066)	(0.007)	(0.010)	
mean	20.49	51.29	26.65	-0.08	-0.01	3.38	0.77	0.29	
N	27,758	27,758	27,758	27,758	27,758	27,521	27,758	27,755	

Notes: This table reports 2SLS estimates of middle school demographic composition effects on Hispanic students' high school choices. Panel A focuses on each applicant's top 3 choices, panel B includes all the choices. The sample, controls and endogeneous variables are as defined in the notes of Table 2.5. Standard errors clustered at the Middle school × year level in parenthesis.

Figure B.1.2: Differences in Value-Added of High School Choices by Race, Controlling for District



Notes: This figure shows how the racial gap varies across student baseline achievement for all students (panel a), and separately for students enrolled in majority Black and Hispanic middle schools and majority white and Asian middle schools (panel b). The two panels plot the coefficient on a dummy indicating Black and Hispanic applicants in regressions of school value-added in first choices on race and controlling for district of residence. Each dot corresponds to a separate regression restricting to applicants with baseline achievement in a different vingtile. Vertical bars indicate 95% confidence intervals.

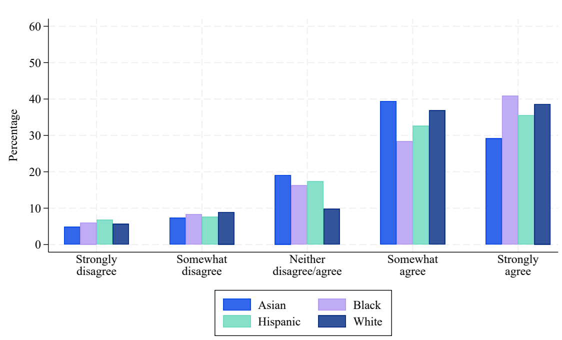
Table B.1.9: 2SLS Estimates of Peer Share Effects on Black & Hispanic Students' Top 3 High School Choices - Multiple Treatment

	% Black (1)	% Hispanic (2)	% white (3)	Peer Ela (4)	Peer Math (5)	Popularity (6)	Screened (7)	SAT math VA (8)	Length of rol (9)
					Panel A: Black applicants				
Share white (10pp)	-1.817**	-0.263	2.103***	0.020*	0.028**	0.230**	0.021	0.026	-0.020
	(0.640)	(0.372)	(0.591)	(0.010)	(0.011)	(0.095)	(0.015)	(0.020)	(0.143)
Share Hispanic (10pp)	-3.868***	4.273***	-0.173	-0.006	-0.002	-0.032	-0.041*	0.018	0.307
	(0.968)	(0.740)	(0.793)	(0.018)	(0.019)	(0.165)	(0.024)	(0.032)	(0.260)
mean	34.00	37.11	26.88	-0.02	0.02	3.18	0.62	0.29	8.43
N	15,284	15,284	15,284	15,284	15,284	15,208	15,284	15,283	15,284
					Panel B: Hispanic applicants				
Share white (10pp)	-0.329	-2.121***	2.429***	0.033**	0.040***	0.176	0.034**	0.012	-0.009
	(0.270)	(0.487)	(0.518)	(0.010)	(0.012)	(0.122)	(0.013)	(0.018)	(0.151)
Share Black (10pp)	3.082***	-3.416***	0.175	0.012	-0.011	-0.072	0.055**	-0.008	0.430
	(0.470)	(0.702)	(0.659)	(0.015)	(0.018)	(0.177)	(0.022)	(0.028)	(0.279)
mean	34.00	37.11	26.88	-0.02	0.02	3.18	0.62	0.29	8.43
N	27,758	27,758	27,758	27,758	27,758	27,517	27,758	27,753	27,758
					Panel C: White applicants				
Share Black (10pp)	2.247***	1.215*	-3.583***	-0.036*	-0.057**	-0.054	0.023	-0.066*	-0.216
	(0.644)	(0.626)	(0.816)	(0.019)	(0.021)	(0.139)	(0.021)	(0.034)	(0.256)
Share Hispanic (10pp)	-0.433	1.800***	-1.342**	-0.012	-0.021	0.157	0.001	-0.002	0.087
	(0.343)	(0.487)	(0.567)	(0.013)	(0.015)	(0.100)	(0.014)	(0.027)	(0.179)
mean	34.00	37.11	26.88	-0.02	0.02	3.18	0.62	0.29	8.43
N	25,040	25,040	25,040	25,040	25,040	24,994	25,040	25,022	25,040

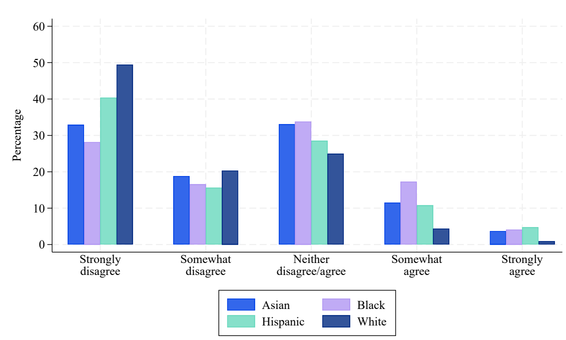
Notes: This table reports 2SLS estimates of middle school demographic composition effects on high school choices for models with two endogenous regressors, one for each race share different from own. Panel A focuses on Black applicants, panel B on Hispanic applicants, while panel C on white and Asian applicants. The sample, controls and endogeneous variables are as defined in the notes of Table 2.5. Standard errors clustered at the Middle school × year level in parenthesis.

Figure B.1.3: Perceptions of Fitness in School

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



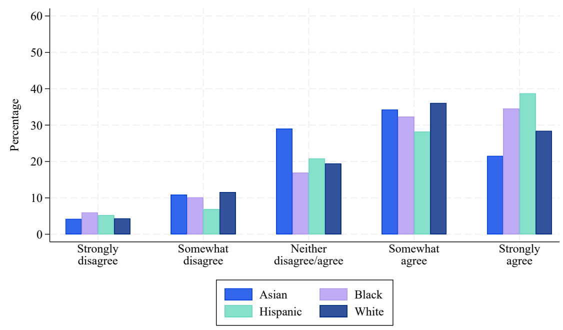
(b) Student would fit well with peers of different achievement



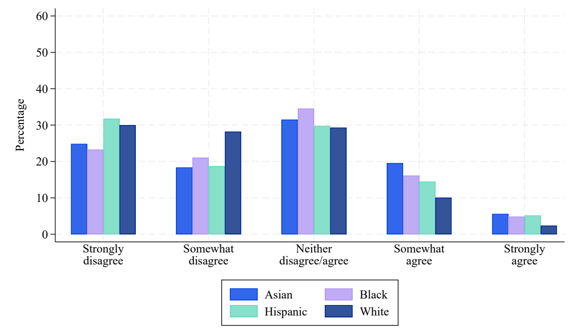
Notes: This figure shows the raw distribution of answers to survey questions asking respondents how well they feel they would fit within a school community, separately by respondent race.

Figure B.1.4: Perceptions of Racial Belonging and Discrimination

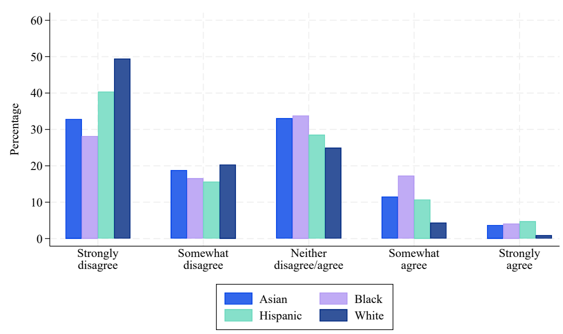
(a) Student would belong with peers of other race



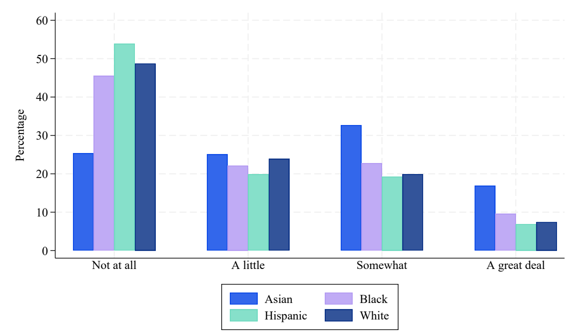
(b) Discrimination by classmates



(c) Discrimination by teachers



(d) Discrimination influenced choices



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

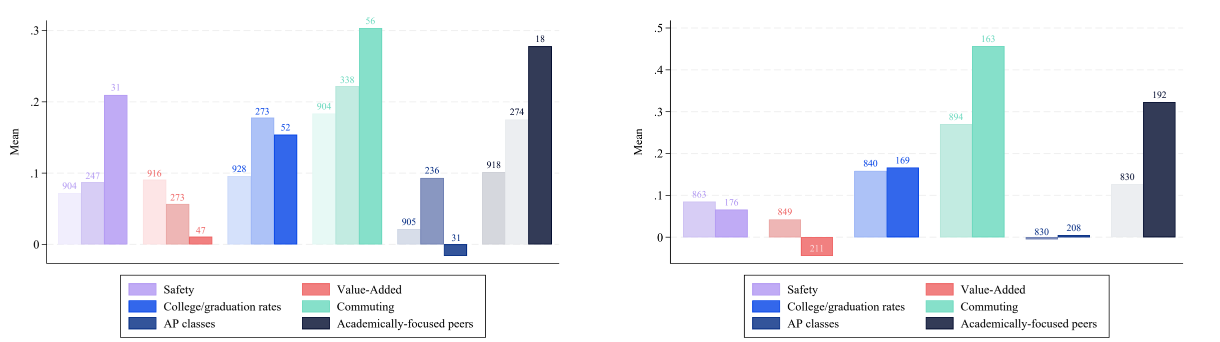
Table B.1.10: Peer Effects on Perceived Discrimination

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

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

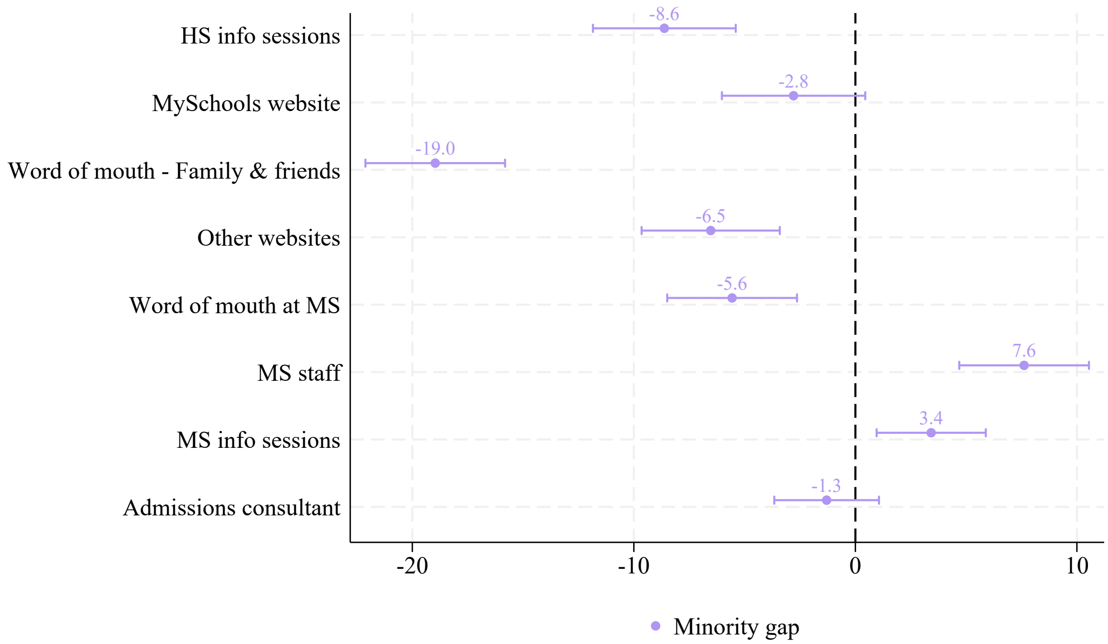
Figure B.1.5: Differences in Information Accuracy About School Characteristics if School is in Awareness Set

(a) Excess p(correct) - pairwise comparisons (b) Correlation between real quartile and answer



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

Figure B.1.6: Race Differences in Sources of Information Used



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

B.2 Survey Appendix

This appendix provides comprehensive details on the content and implementation of the post-application survey conducted in partnership with the New York City Department of Education during the 2023-2024 high school admission cycle. This appendix is organized as follows. Section B.2.1 explains the survey logistics, including timeline, the emails sent to participants, and the Qualtrics design. Section B.2.2 describes the selection of potential participants. Section B.2.4 describes the survey blocks and the randomization of participants to survey versions. Section B.2.5 explains the selection of randomized schools in the survey. Finally, section B.2.3 includes all survey questions as shown to participants.

B.2.1 Survey logistics

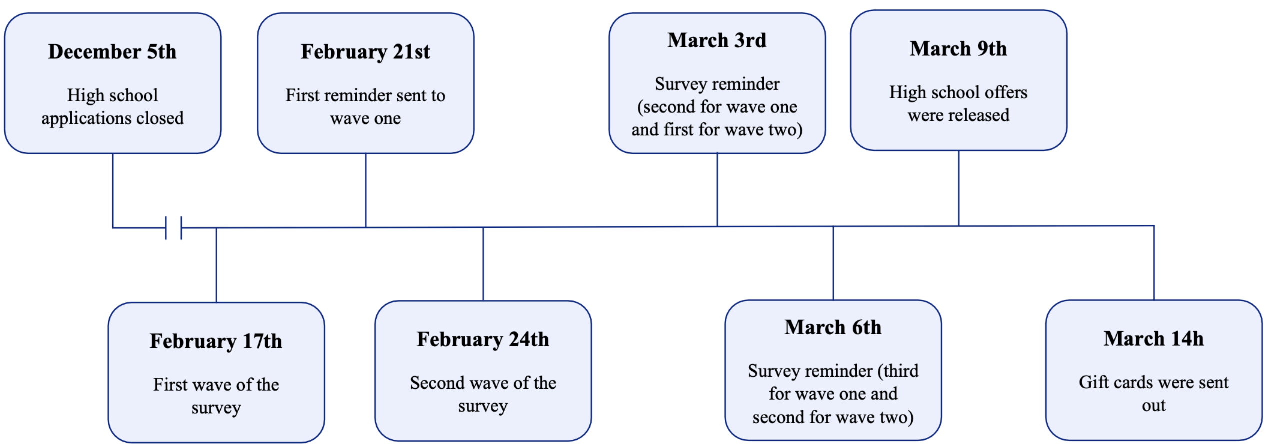
Survey timeline

High school applications in NYC closed on December 5th, 2023. The survey was designed to be sent after families applied to high school but before the offers were sent out on March 9th,

2023. The timeline allowed parents to have at least two weeks to complete it, and the survey had no time constraints beyond the March 6th deadline. Incomplete surveys were automatically submitted by the deadline. Participants who answered at least one question by the deadline received a $10 Amazon gift card.

The survey was sent electronically using the email addresses of families used in the high school application process. It was conducted in two waves between February 17th and March 6th, 2023. The first wave, including 11,415 families, was sent on Friday, February 17th. A week later, on Friday, February 24th, the second wave was sent out to 9,986 families. Wave one participants received three reminders: one on February 21st, the second on March 3rd, and the last on March 6th, 2023 (the last day to respond to the survey). Wave two participants received two reminders: one on March 3rd, and the second on March 6th. All gift cards were sent out on March 14th. Figure B.2.7 illustrates the survey timeline.

Figure B.2.7: Survey Timeline



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

Survey emails

The survey was sent by email in the top three most spoken languages in NYC[1]: English, Spanish, and Simplified Chinese. Figure B.2.8 and B.2.9 show the English version of the first email sent to the potential participants and the reminder email.

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

Figure B.2.8: Invitation to Participate. English Version

MIT Economics

Dear parent/guardian,

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

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

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

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

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

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

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

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

Thank you very much!

The research team

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

---

Figure B.2.9: Survey Reminder. English Version

MIT Economics

Dear parent/guardian,

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

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

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

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

Or copy and paste the URL below into your internet browser:
https://mit.co1.qualtrics.com/jfe/preview/previewId/81e99fbf-5fc8-43f4-a64b-fa87bf0f6fb4/SV_4OysVkOlp8xjaZM?Q_CHL=preview&Q_lang=EN

This survey is part of a research study conducted by researchers at the Massachusetts Institute of Technology in partnership with the New York City Department of Education. **This study is independent from the high school choice placement process** and the survey is **confidential** and **voluntary**. You must be **at least 18 years old** to participate in this study.

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

The research team

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

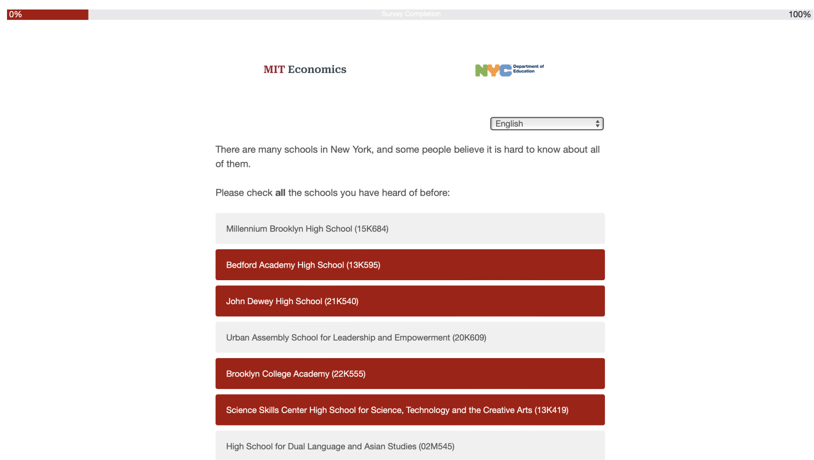
---

Qualtrics design

The survey was designed on Qualtrics and it was available in English, Spanish, and Simplified Chinese. All questions were marked as optional, except the consent to participate one: to access the survey, participants had to check a box stating that they were over 18 years old.

We personalized the survey by using JavaScript to present participants with different sets of schools. For instance, question 9 displayed a distinct set of high-demand and popular schools based on the participant's borough. See Figure B.2.10 and section B.2.5 of the main text for more details.

Figure B.2.10: Question 9. Variation in Schools Presented to Respondents Based on Their Characteristics



Notes: This figure shows one of the possible school sets presented to respondents in Q9. The set of schools in this question was tailored to each respondent, as explained in detail in section B.2.5.

B.2.2 Selection of participants

Assignment of participants to survey waves

The sample of all eligible participants consisted of parents or guardians of students applying to start 9th grade in fall 2023. Participant population data was selected as described in Section 3.3 of the paper. Summary statistics can be found in Table 2.1. We categorized participants into survey waves based on their informativeness by assigning each participant a priority, which sorted them into different waves. We only sent the survey to some of the waves. Eligible participants' priorities range from one to six and were determined as follows:

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

2. Survey priority is 2 if the student was enrolled in a DOE school in 6th grade, has all demographic information, and has all baseline scores.

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

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

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

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

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

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

The first wave of the survey included all families of students with a survey priority of 1. The second wave included the first 10,000 priority 2 students, sorted by their scrambled ID. Additional waves of participants were created for potential expansion of the survey, although they were not used. The final sample of potential participants comprised 21,401 parents or guardians.

B.2.3 Survey Questions

The complete survey had a total of 47 questions, including the consent question and the end-of-survey comment box. The questions are reported in the pictures below.

Figure B.2.11: Consent Question

MIT Economics

NYC Department of Education

English

Researchers at the Massachusetts Institute of Technology are conducting a research study in partnership with the New York City Department of Education (DOE) about the New York City public high school application process. You are receiving this survey because you recently applied to high school.

We are conducting this research study to **learn more about how families choose schools.** We are also interested in understanding whether families make different choices based on the middle school that their student attends. We hope that our results will generate new understanding about school choice and help the DOE improve the application process in the future.

This study is **separate from the high school application process**. The information you provide to us through this survey will be kept **completely confidential**. Your decision to participate and any answers you provide will **not** influence your offer in any way, nor will your answers be provided to anyone at your student's current or future school.

If you participate in this study, you will be asked to answer a **10-minute survey about your experience with the application process**. We are interested in hearing your **perspective as a parent or guardian**. There are no known risks associated with your participation in this research beyond those of everyday life. The **deadline** for filling the survey is **March 6th.**

The first 5,000 respondents will receive a **$10 Amazon gift card as compensation for their time**. The gift card will be sent to this same email address after the survey is closed.

Participation in this study is **voluntary**. You may choose not to participate or stop at any time. **Please read the rest of this consent form for more information about the study.** If you have any questions about this study, you may contact the investigator, Clemence Idoux at cidoux@mit.edu

☑ **By checking this box and completing the survey, you are consenting to participate in this study and certifying to be at least 18 years old.**

→

Figure B.2.12: Question 1

0% Survey Completion 100%

MIT Economics

NYC Department of Education

English

What is your relationship to the student?

- Parent
- Brother/Sister
- Grandparent
- Uncle/Aunt
- Other: please specify

Figure B.2.13: Question 2

Who played the most important role in deciding how to fill out the high school application?

- The student
- The parents/guardians
- Both the parents/guardians and the student
- School staff
- Other: please specify

→

Powered by Qualtrics

Figure B.2.14: Question 3

0% Survey Completion 100%

MIT Economics

English

In the past year, how many times did you talk to **other parents/guardians** from your student's middle school about which high schools to apply to?

- Never
- 1-5 times
- More than 5 times

Figure B.2.15: Question 3b

What were the **most important sources of information** for deciding which schools to include in your student's application? **Rank** up to three (1 should be the most important to you).

	1	2	3
Word-of-mouth from other parents/guardians at your student's middle school	○	○	○
Word-of-mouth from family and friends	○	○	○
School-admissions consultant	○	○	○
Attending information sessions at your student's middle school	○	○	○
Attending high school information sessions	○	○	○
Other websites (InsideSchools, GreatSchools etc.)	○	○	○
MySchools' high school directory	○	○	○
Staff at your student's middle school	○	○	○

Figure B.2.16: Question 3b.2

If you think of an important source of information that is not mentioned, please add it here:

→

Powered by Qualtrics

Figure B.2.17: Question 4

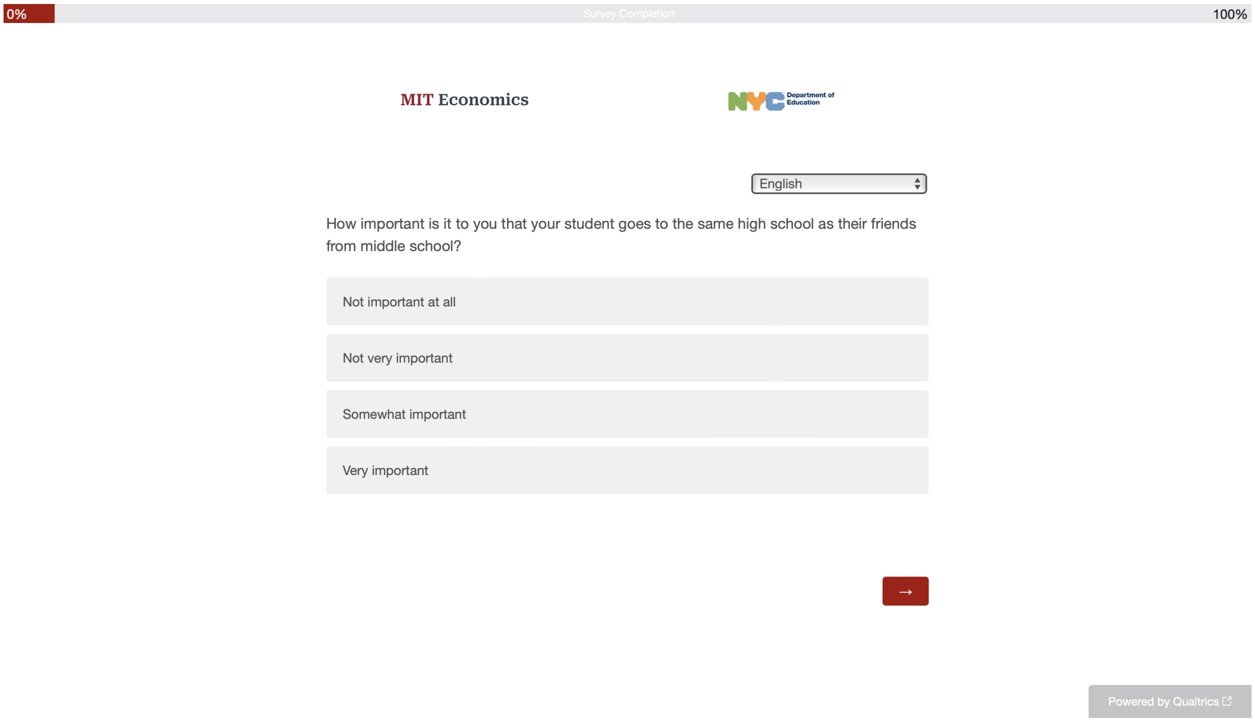


Figure B.2.18: Question 5

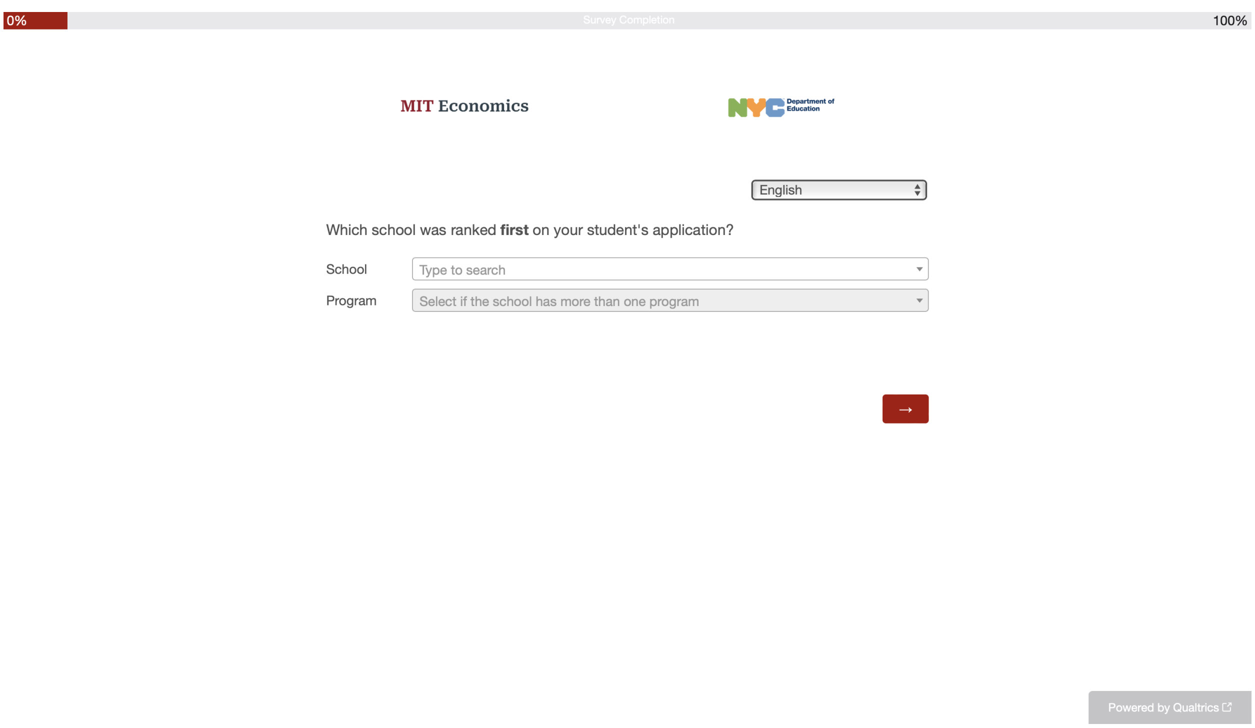


Figure B.2.19: Question 6

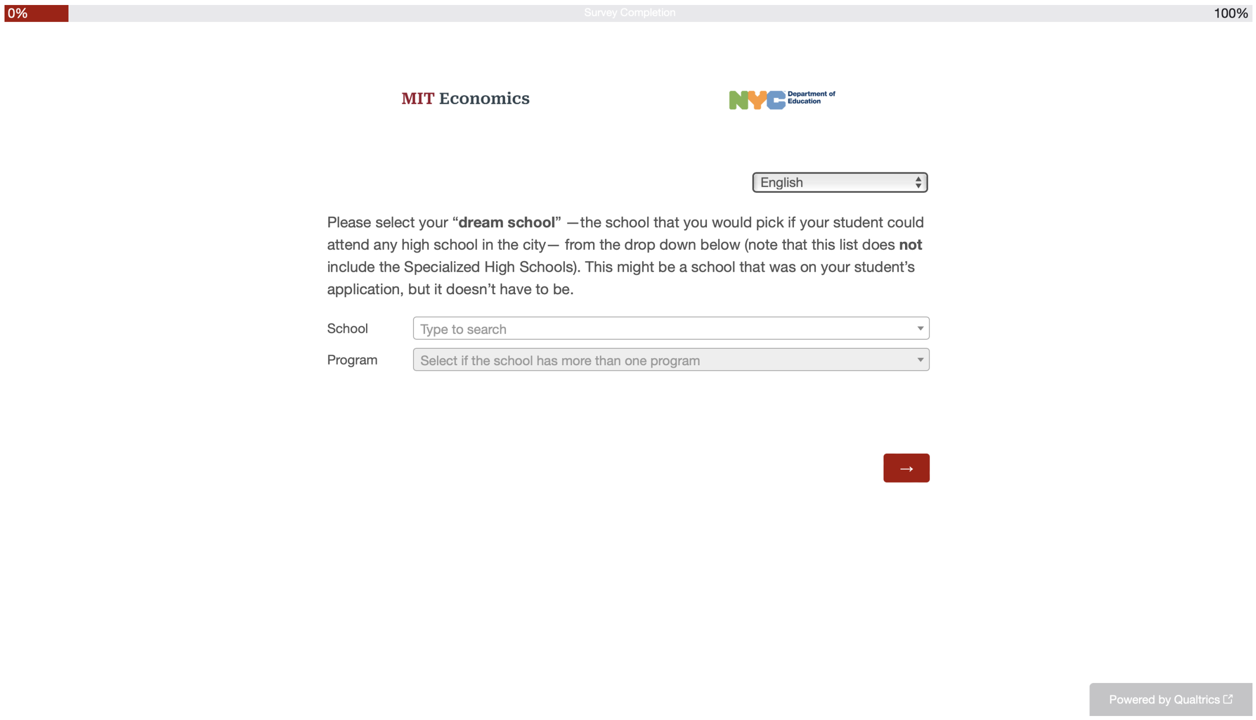


Figure B.2.20: Question 7a

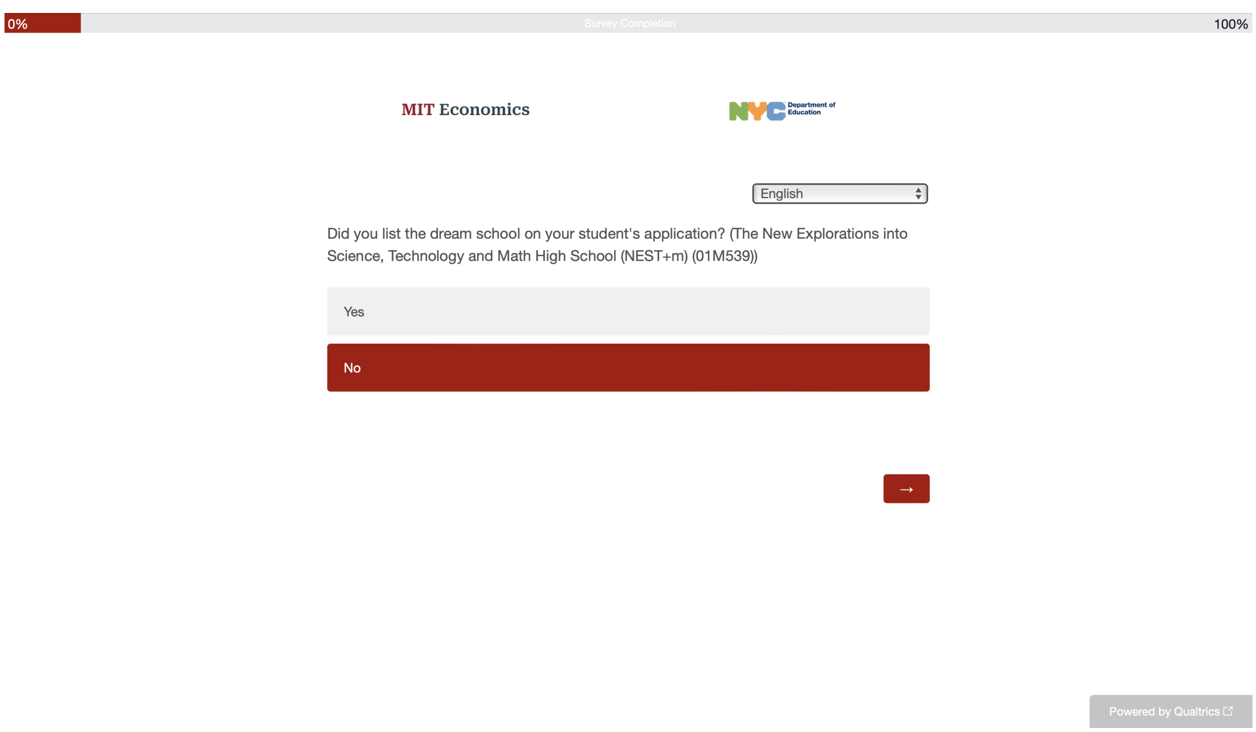
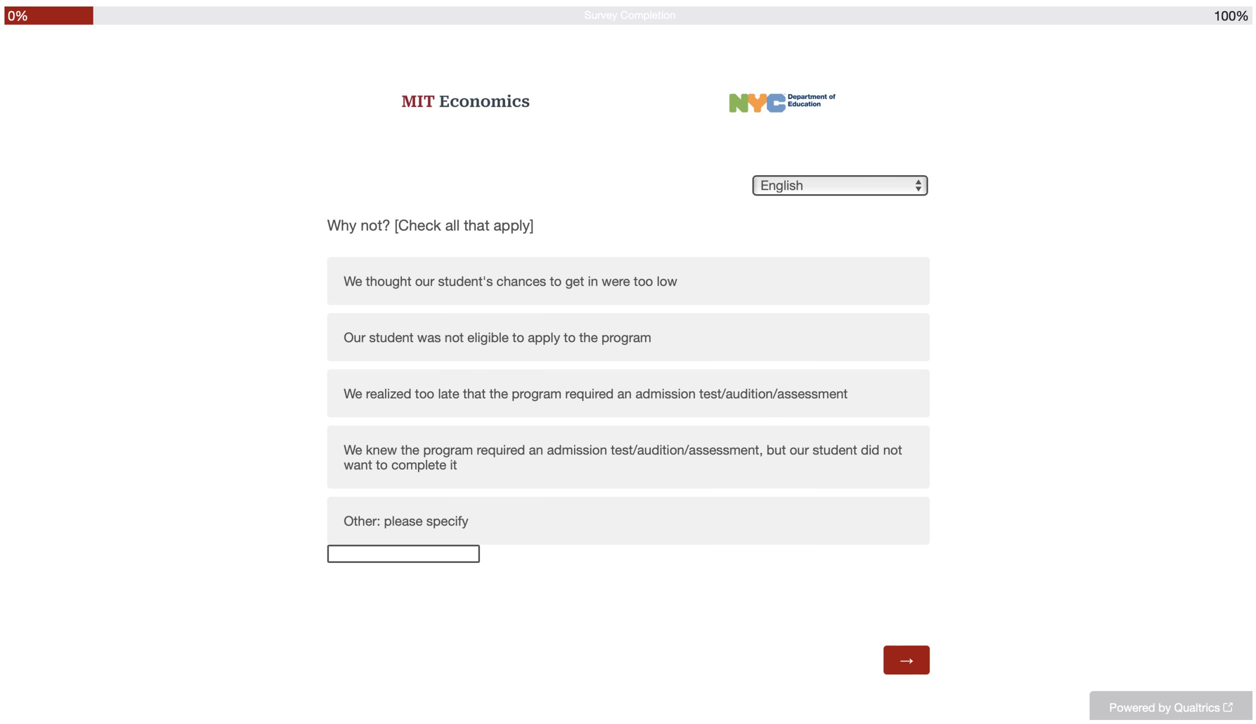


Figure B.2.21: Question 7b


0% Survey Completion 100%

MIT Economics

English

Why not? [Check all that apply]

We thought our student's chances to get in were too low

Our student was not eligible to apply to the program

We realized too late that the program required an admission test/audition/assessment

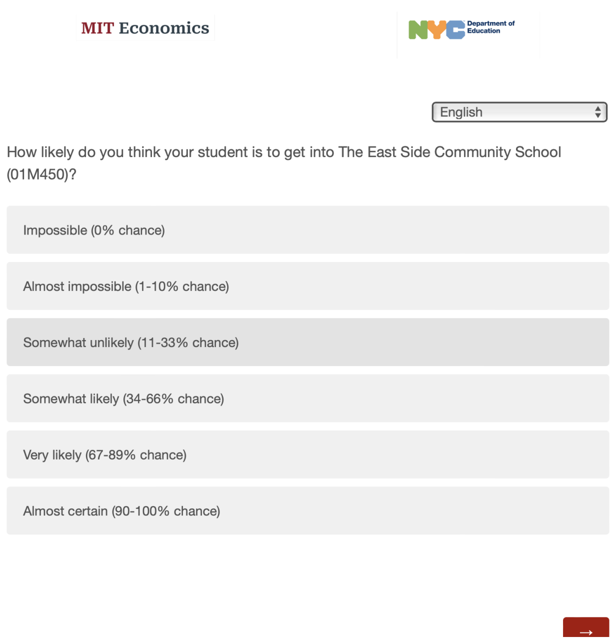
We knew the program required an admission test/audition/assessment, but our student did not want to complete it

Other: please specify

→

Powered by Qualtrics

Figure B.2.22: Question 7c


MIT Economics

English

How likely do you think your student is to get into The East Side Community School (01M450)?

Impossible (0% chance)

Almost impossible (1-10% chance)

Somewhat unlikely (11-33% chance)

Somewhat likely (34-66% chance)

Very likely (67-89% chance)

Almost certain (90-100% chance)

→

Figure B.2.23: Question 7c.2

0% Survey Completion 100%

MIT Economics

English

How likely do you think your student would be to get into The New Explorations into Science, Technology and Math High School (NEST+m) (01M539) if you had listed it first on your application?

- Impossible (0% chance)
- Almost impossible (1-10% chance)
- Somewhat unlikely (11-33% chance)
- Somewhat likely (34-66% chance)
- Very likely (67-89% chance)
- Almost certain (90-100% chance)

Figure B.2.24: Question 8a

0% Survey Completion 100%

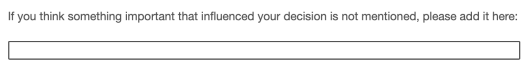
MIT Economics

English

What is **most important** to you when you choose a school? **Rank** up to three (1 should be the most important to you).

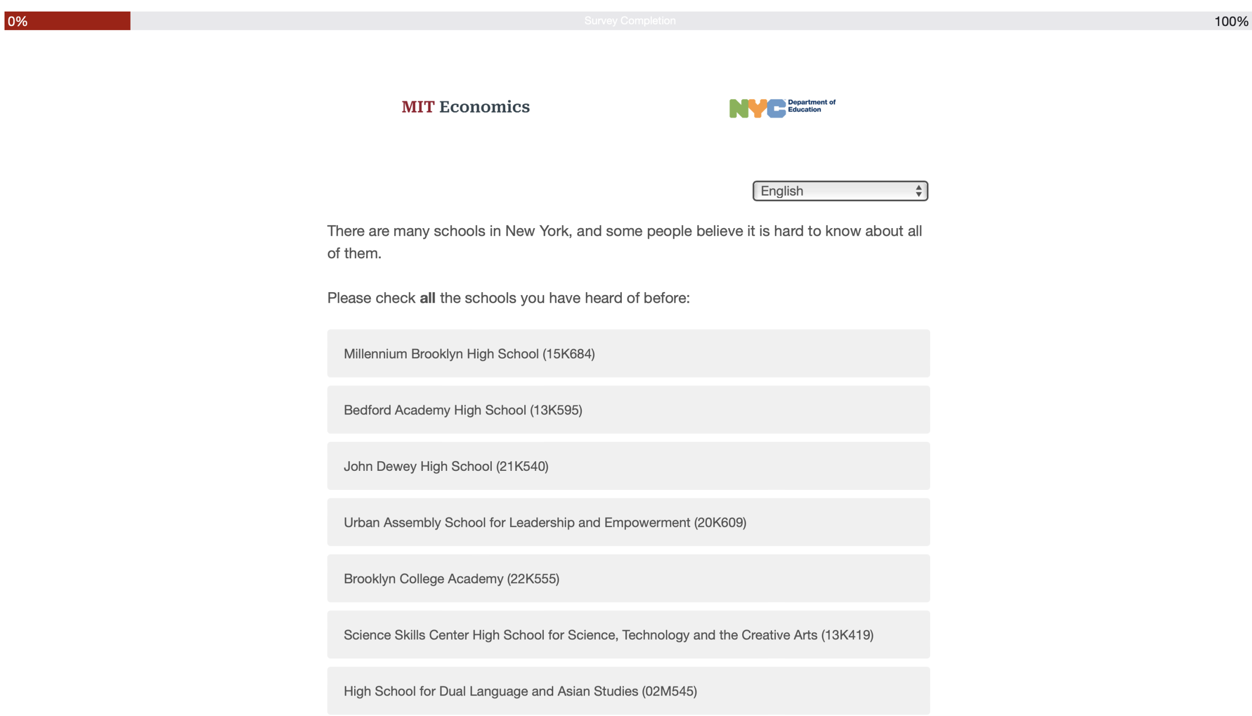
	1	2	3
Career and Technical education	○	○	○
Whether I think my student would feel like they belong	○	○	○
Graduation and/or college enrollment rate(s)	○	○	○
Demographics of the school	○	○	○
School interest areas	○	○	○
Advanced classes (for instance AP classes) offered	○	○	○
Extracurricular activities offered	○	○	○
Safe school environment	○	○	○
School campus and facilities	○	○	○
Other incoming students are academically-focused	○	○	○
Ease of commute	○	○	○
School's neighborhood	○	○	○
Academic progress of students at the school	○	○	○

Figure B.2.25: Question 8a.2



If you think something important that influenced your decision is not mentioned, please add it here:

Figure B.2.26: Question 9



0% Survey Completion 100%

MIT Economics

NYC Department of Education

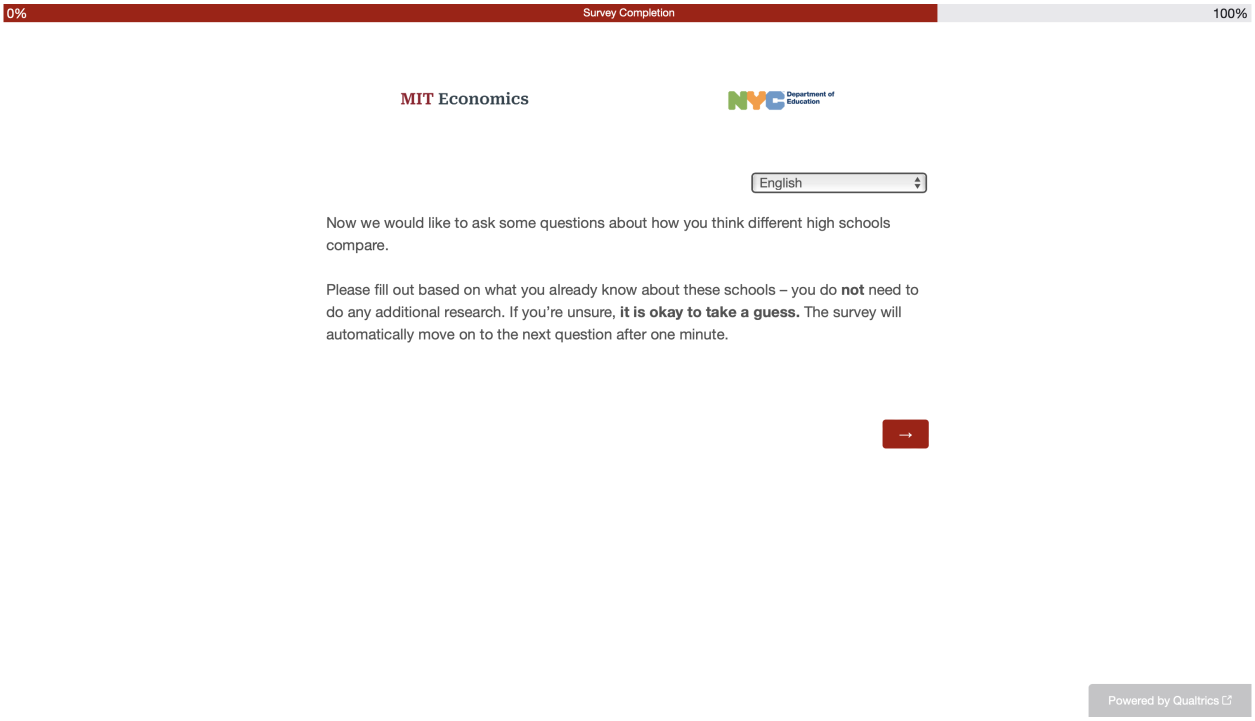
English

There are many schools in New York, and some people believe it is hard to know about all of them.

Please check **all** the schools you have heard of before:

- Millennium Brooklyn High School (15K684)
- Bedford Academy High School (13K595)
- John Dewey High School (21K540)
- Urban Assembly School for Leadership and Empowerment (20K609)
- Brooklyn College Academy (22K555)
- Science Skills Center High School for Science, Technology and the Creative Arts (13K419)
- High School for Dual Language and Asian Studies (02M545)

Figure B.2.27: Question 10



0% Survey Completion 100%

MIT Economics

NYC Department of Education

English

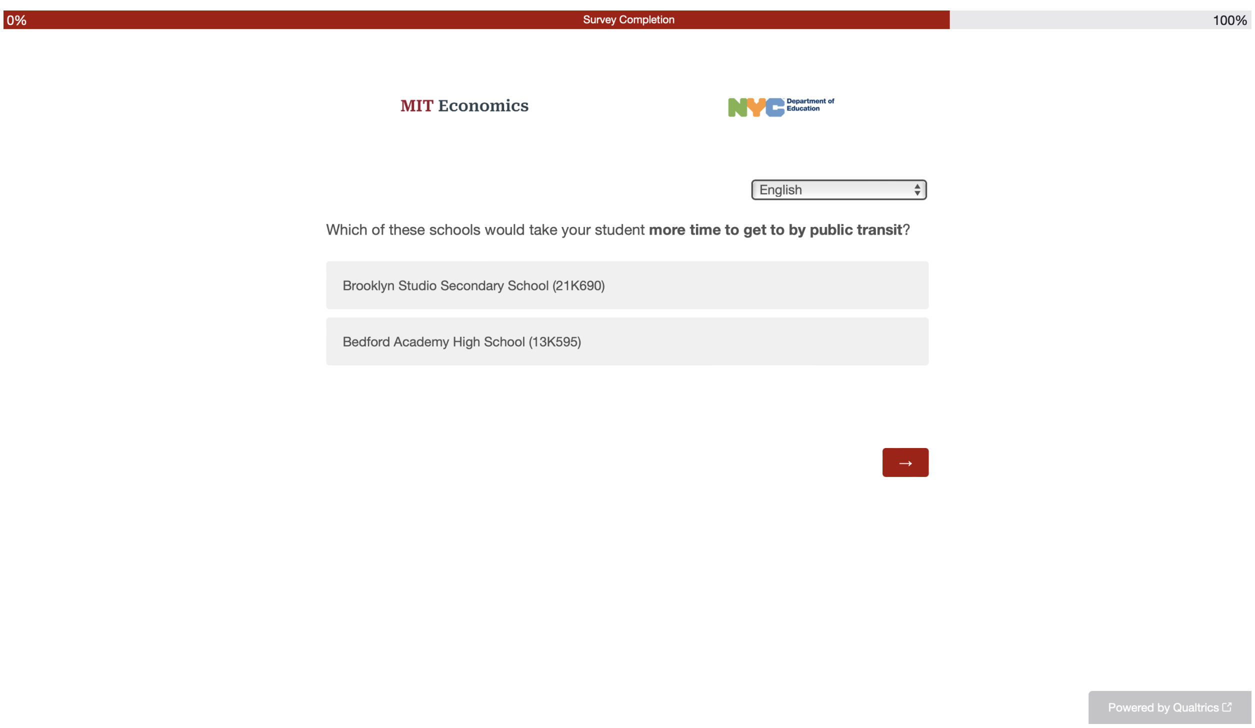
Now we would like to ask some questions about how you think different high schools compare.

Please fill out based on what you already know about these schools – you do **not** need to do any additional research. If you're unsure, **it is okay to take a guess.** The survey will automatically move on to the next question after one minute.

→

Powered by Qualtrics

Figure B.2.28: Question 10a



0% Survey Completion 100%

MIT Economics

NYC Department of Education

English

Which of these schools would take your student **more time to get to by public transit**?

Brooklyn Studio Secondary School (21K690)

Bedford Academy High School (13K595)

→

Powered by Qualtrics

Figure B.2.29: Question 10b

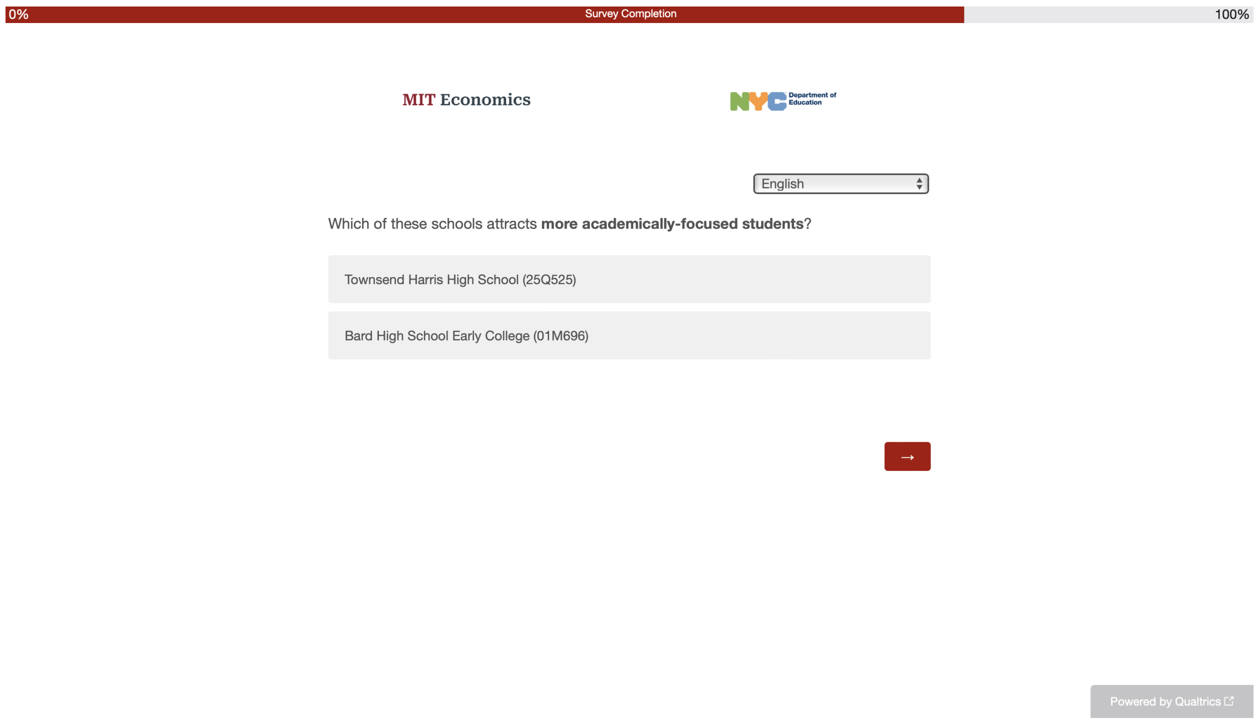


Figure B.2.30: Question 10c

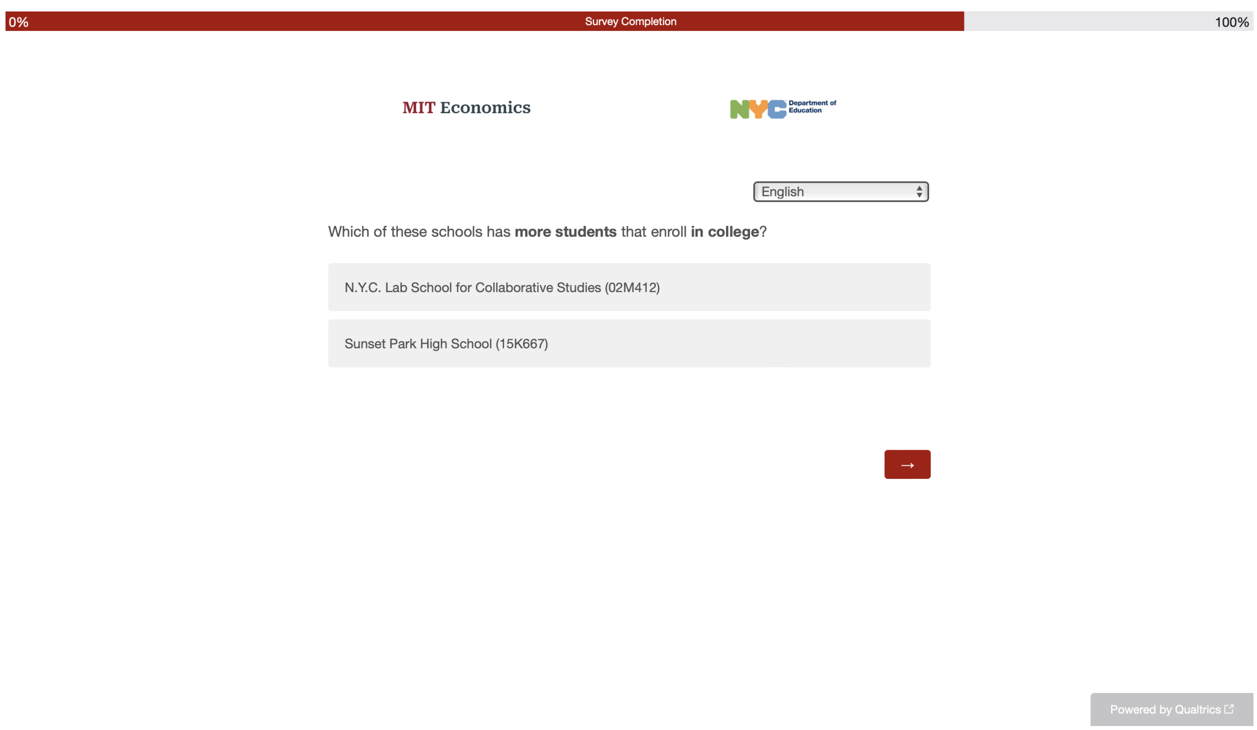


Figure B.2.31: Question 10d

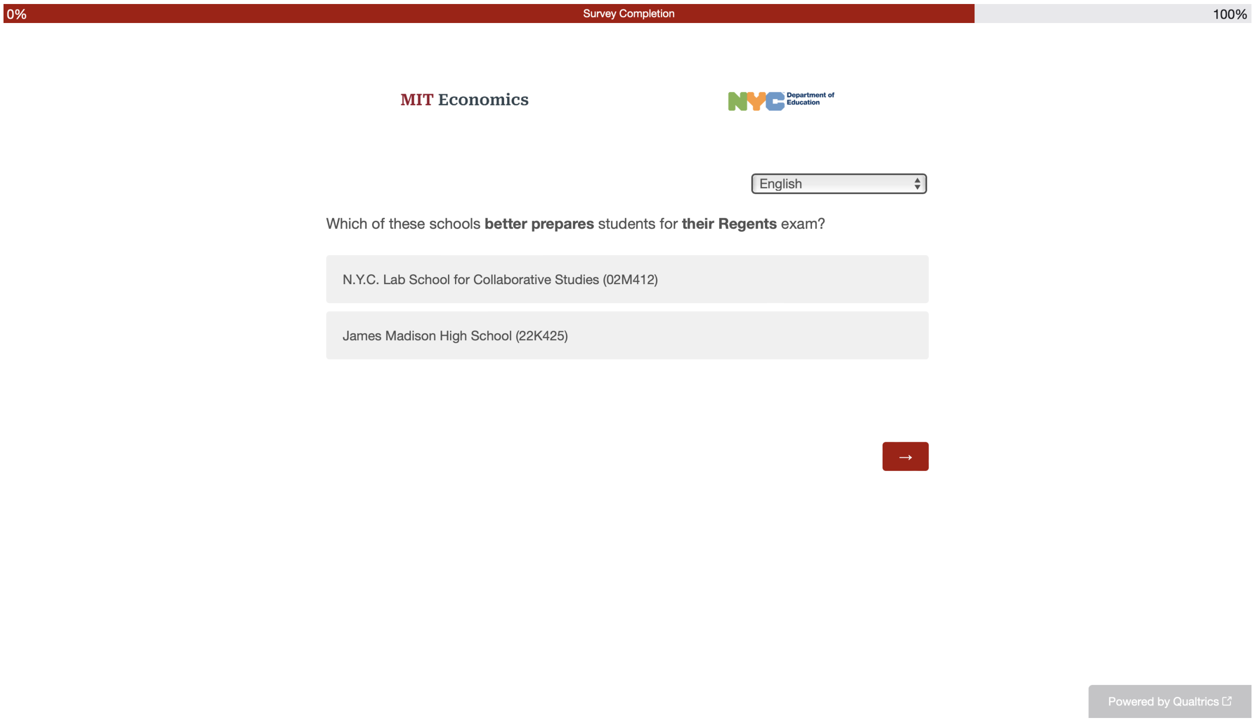


Figure B.2.32: Question 10f

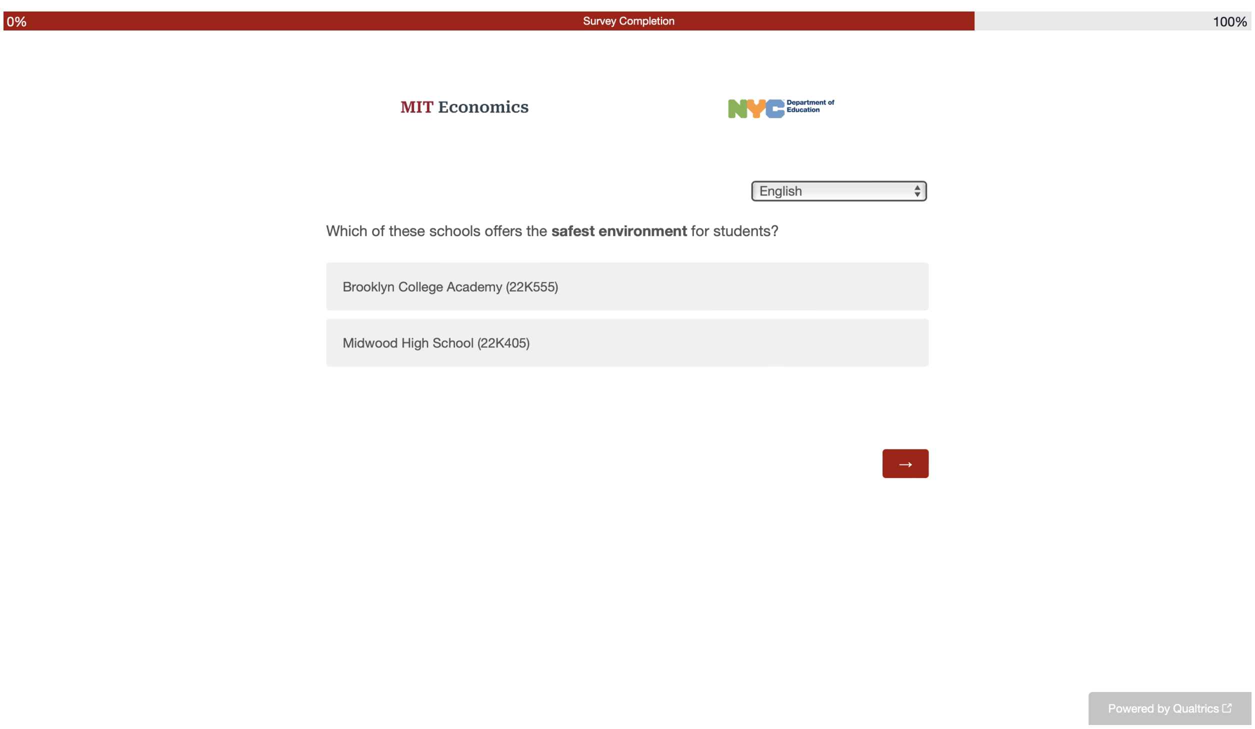


Figure B.2.33: Question 10g

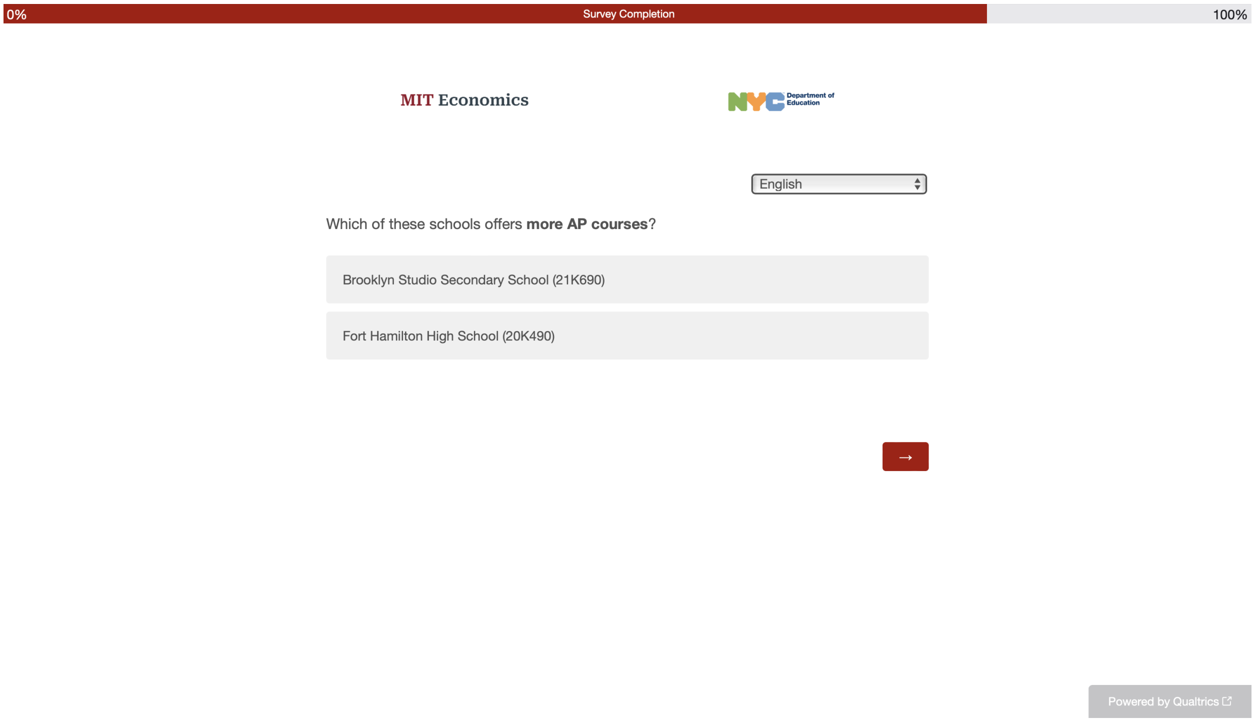


Figure B.2.34: Question 11a

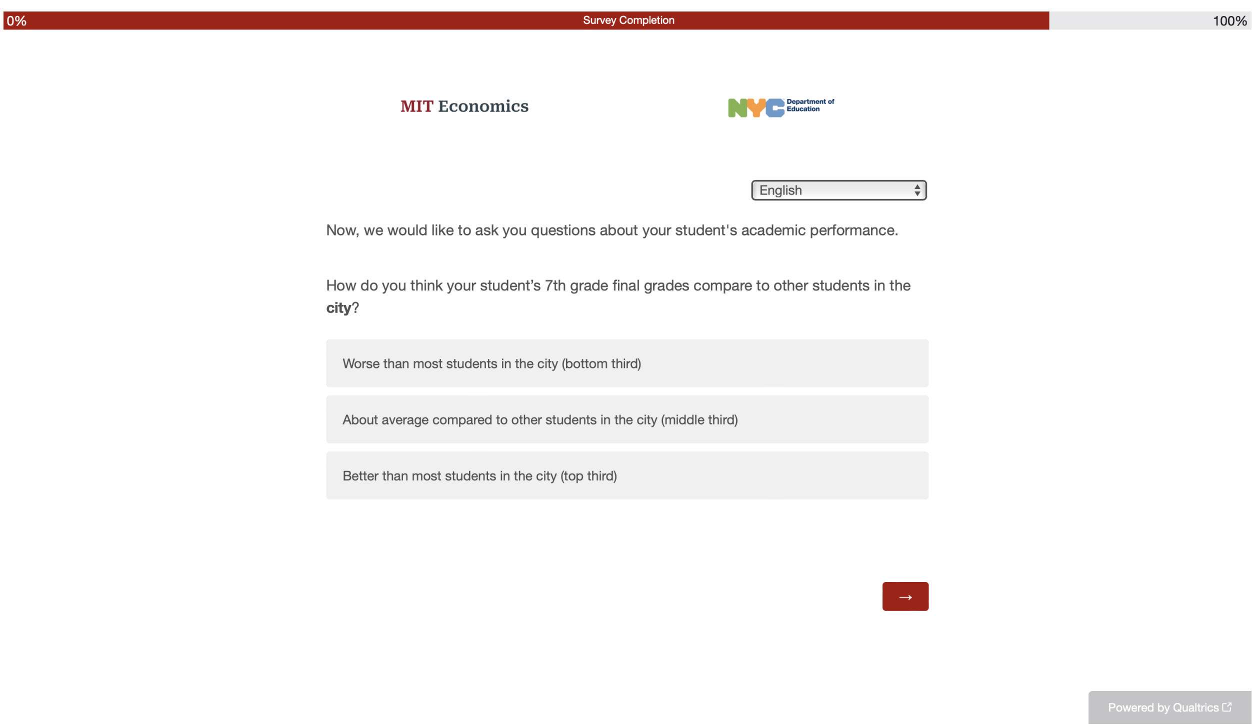


Figure B.2.35: Question 12

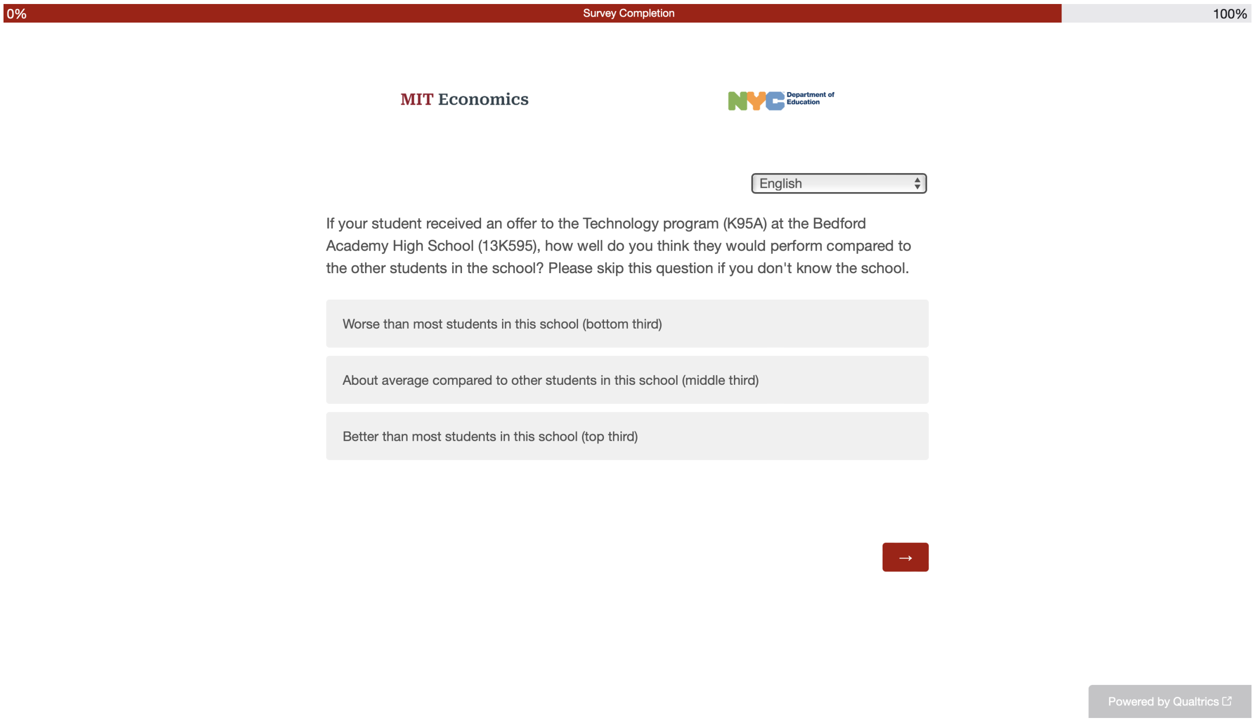


Figure B.2.36: Question 13

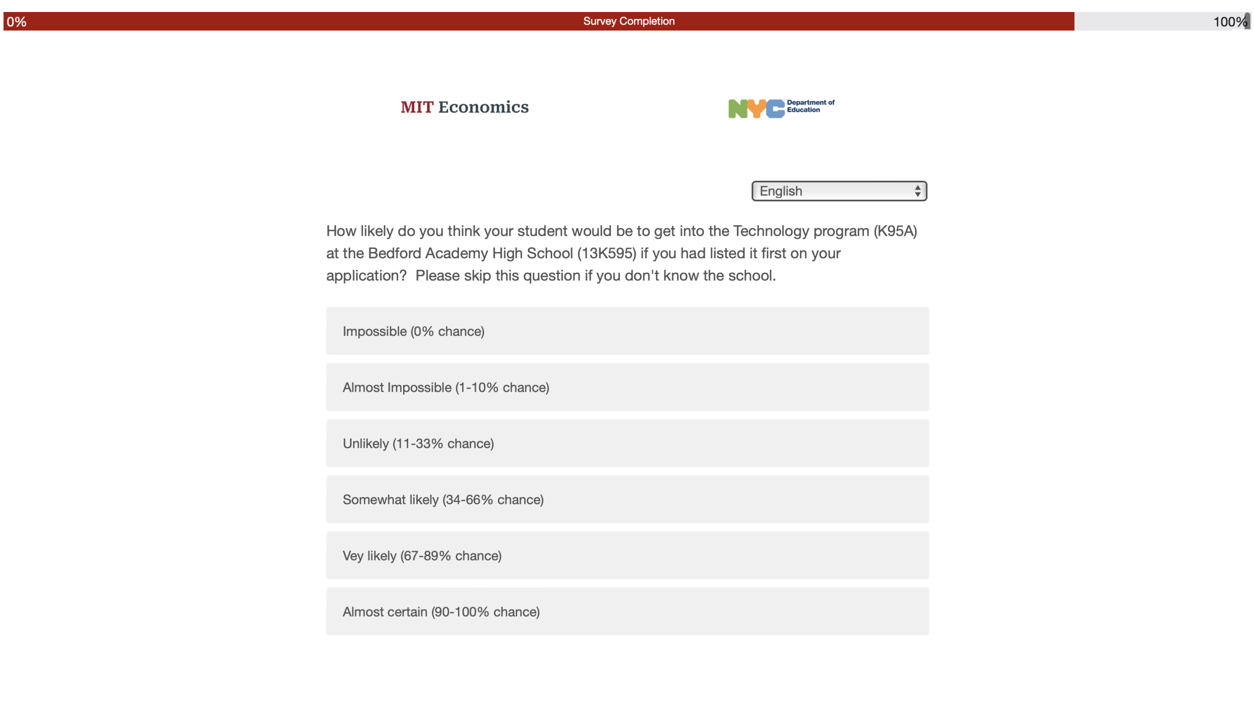


Figure B.2.37: Question 14a

0% Survey Completion 100%

MIT Economics

English

Now we would like to ask you questions about your **aspirations** for your student.

Do you believe that it is important to go to college in order to do well in life?

- Strongly disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Strongly agree

Figure B.2.38: Question 15

What is the **minimum** level of education that you would like your student to complete?

- It does not matter
- Complete high school
- 2 years of college
- 4 years of college
- Graduate school (master, PhD, law or medical school etc.)

Powered by Qualtrics

Figure B.2.39: Question 16a

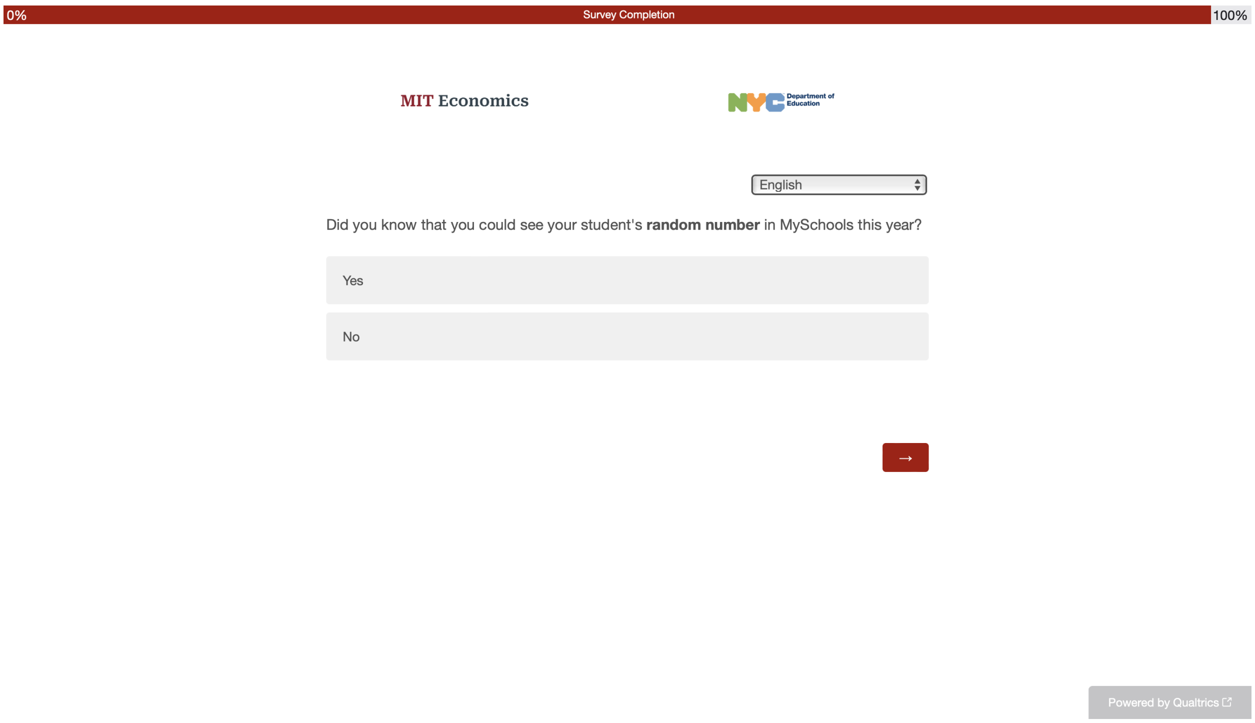


Figure B.2.40: Question 16b

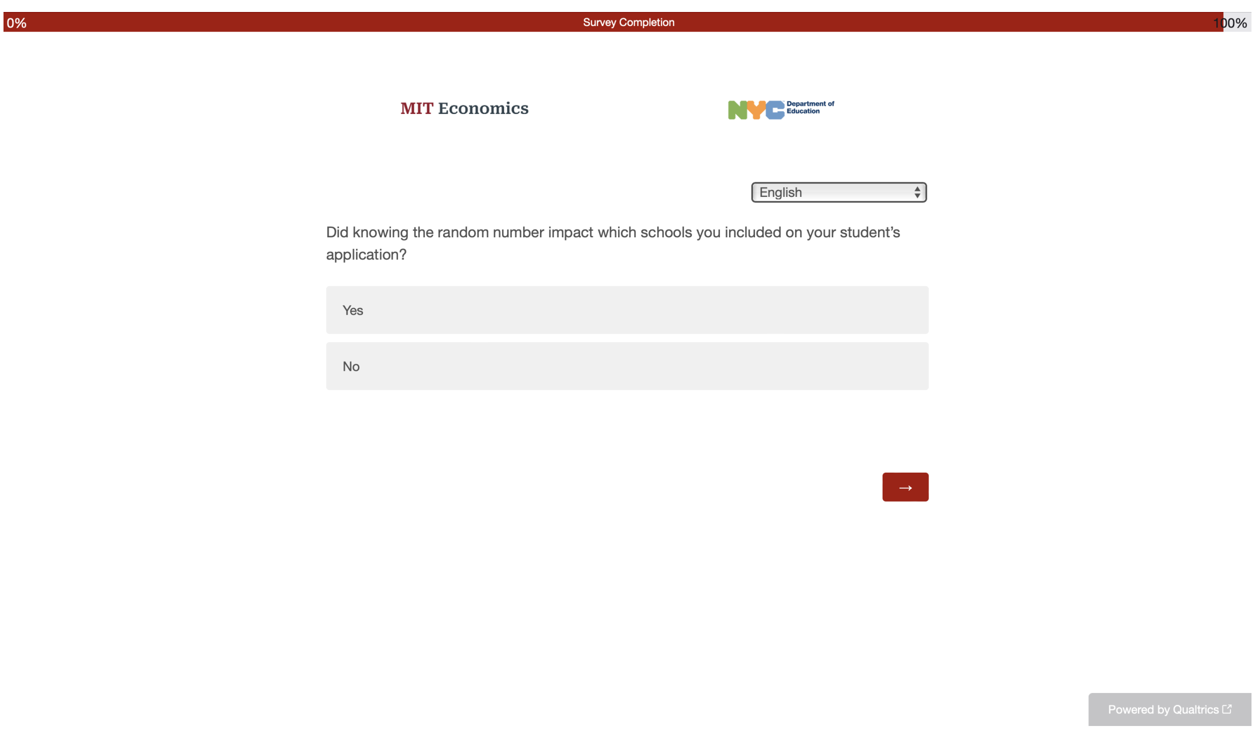
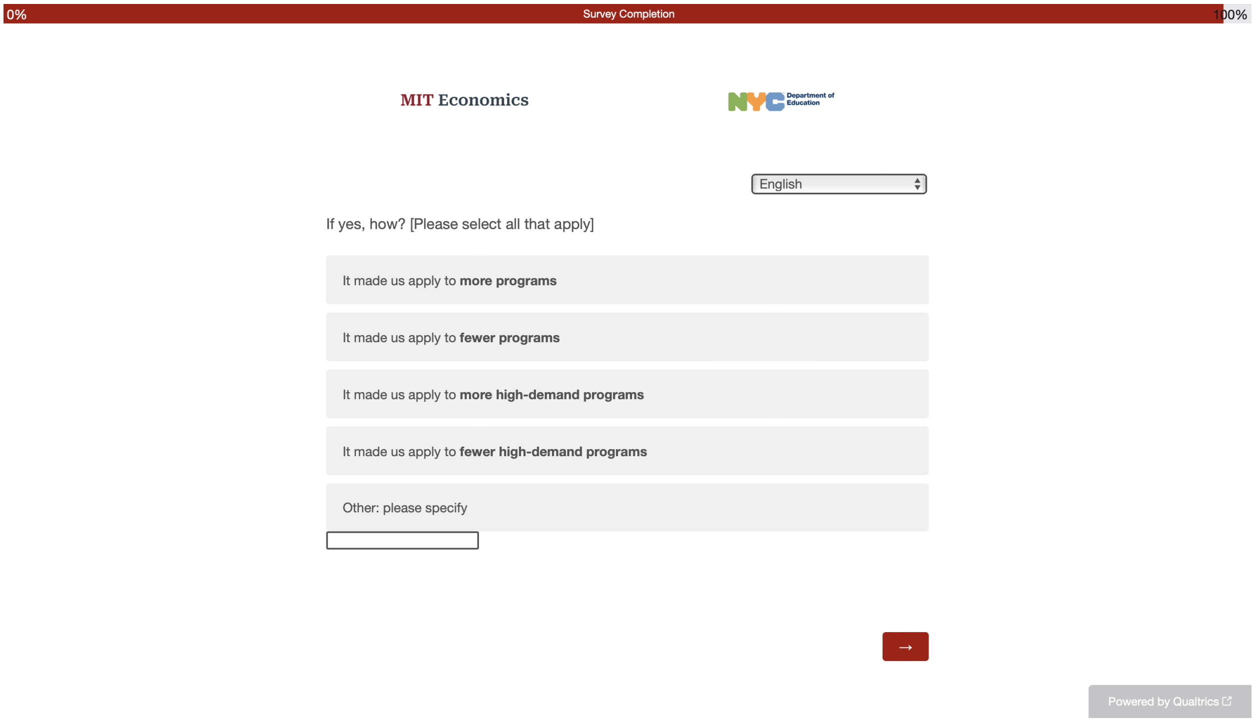


Figure B.2.41: Question 16c


0% Survey Completion 100%

MIT Economics

English

If yes, how? [Please select all that apply]

It made us apply to **more programs**

It made us apply to **fewer programs**

It made us apply to **more high-demand programs**

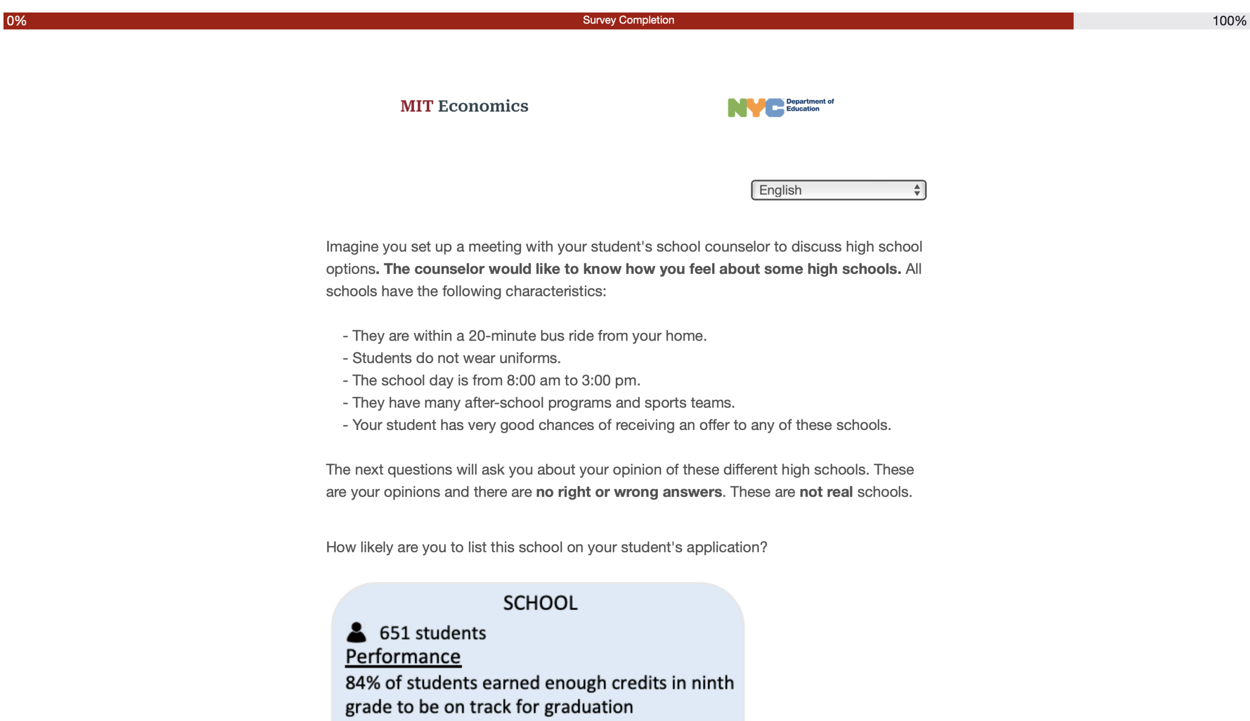
It made us apply to **fewer high-demand programs**

Other: please specify

→

Powered by Qualtrics

Figure B.2.42: Question 17


0% Survey Completion 100%

MIT Economics

English

Imagine you set up a meeting with your student's school counselor to discuss high school options**. The counselor would like to know how you feel about some high schools.** All schools have the following characteristics:

- They are within a 20-minute bus ride from your home.
- Students do not wear uniforms.
- The school day is from 8:00 am to 3:00 pm.
- They have many after-school programs and sports teams.
- Your student has very good chances of receiving an offer to any of these schools.

The next questions will ask you about your opinion of these different high schools. These are your opinions and there are **no right or wrong answers**. These are **not real** schools.

How likely are you to list this school on your student's application?

SCHOOL

651 students

Performance

84% of students earned enough credits in ninth grade to be on track for graduation

Figure B.2.43: Question 17a

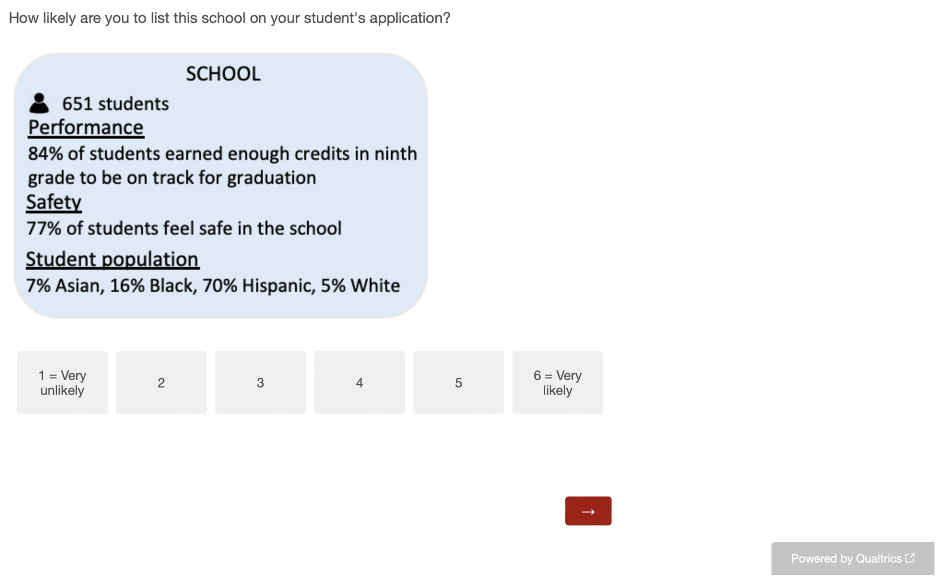


Figure B.2.44: Question 17b

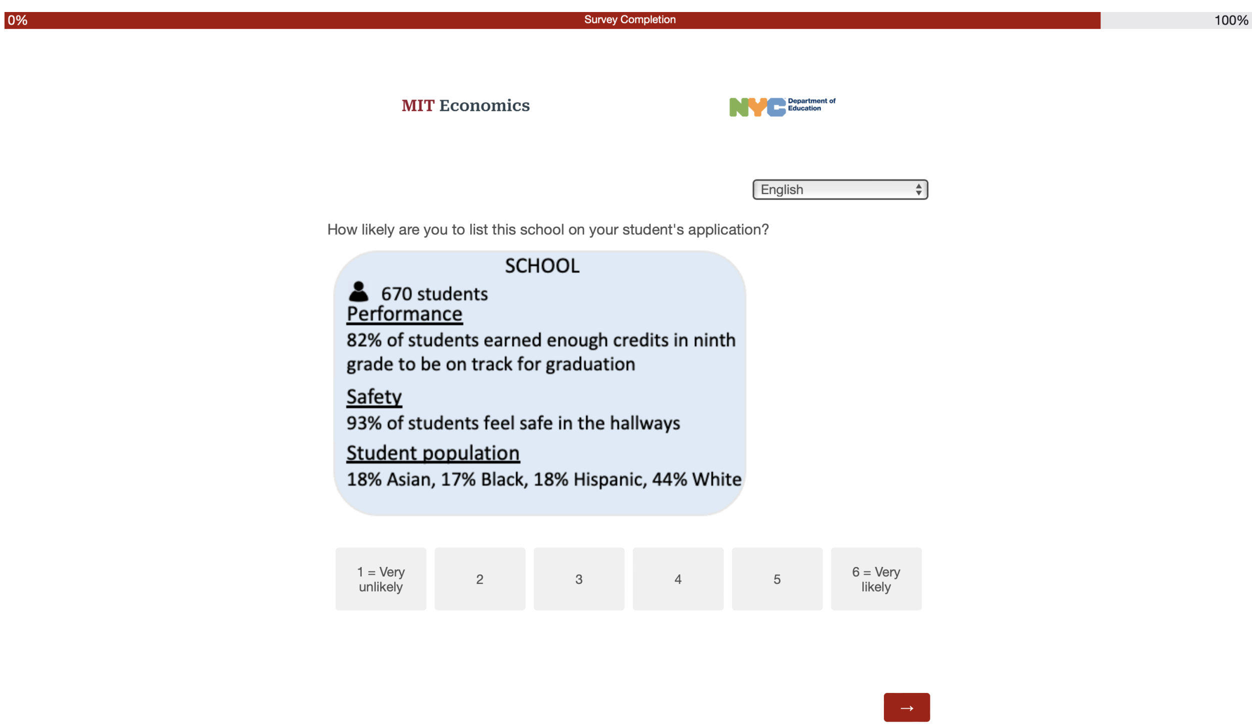


Figure B.2.45: Question 18a

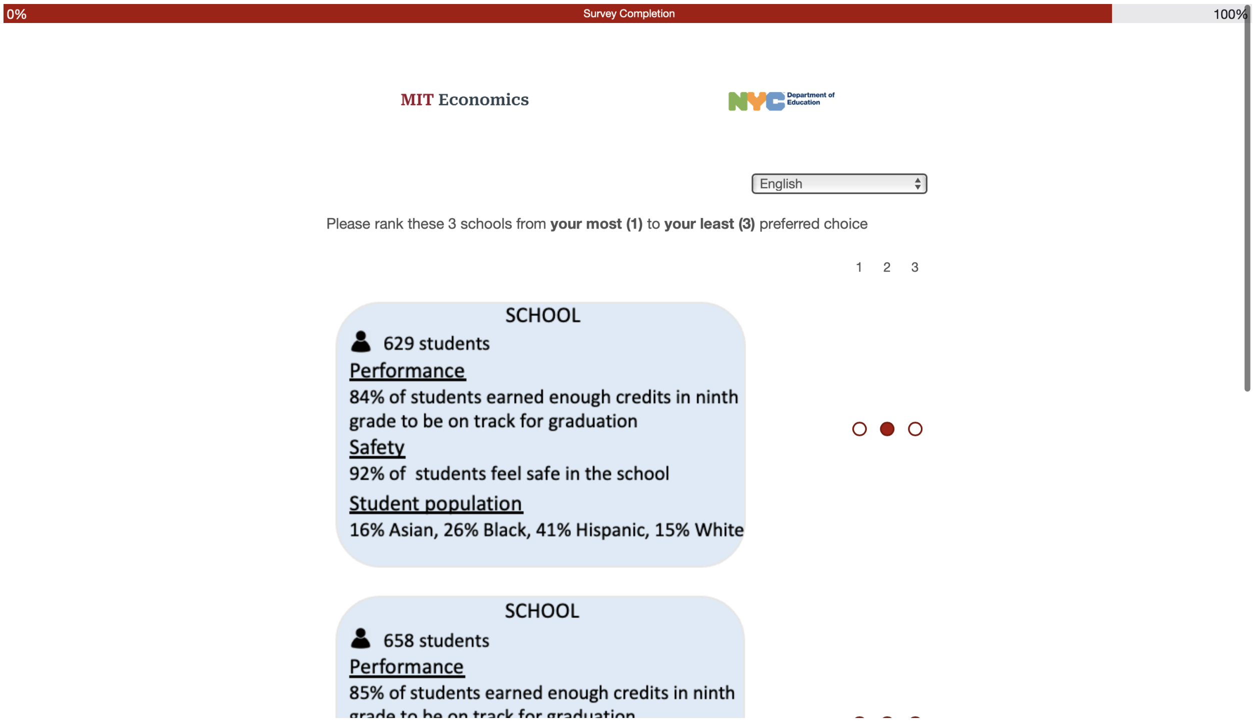


Figure B.2.46: Question 18b

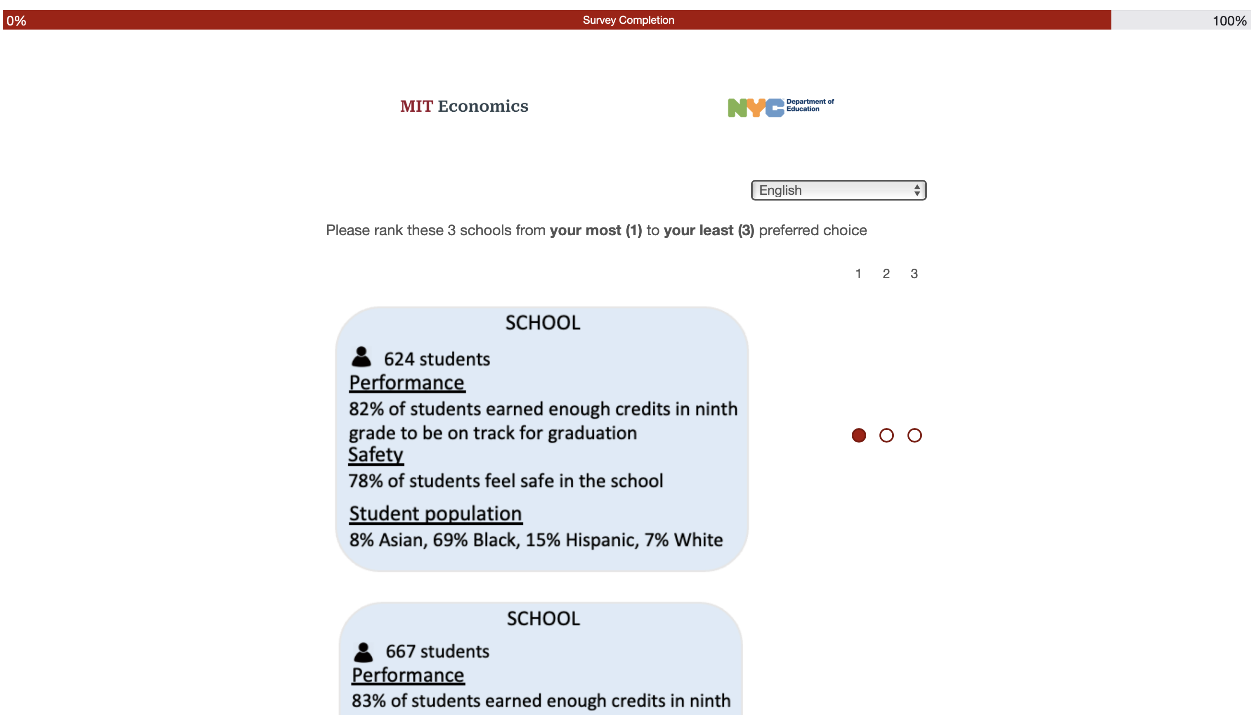
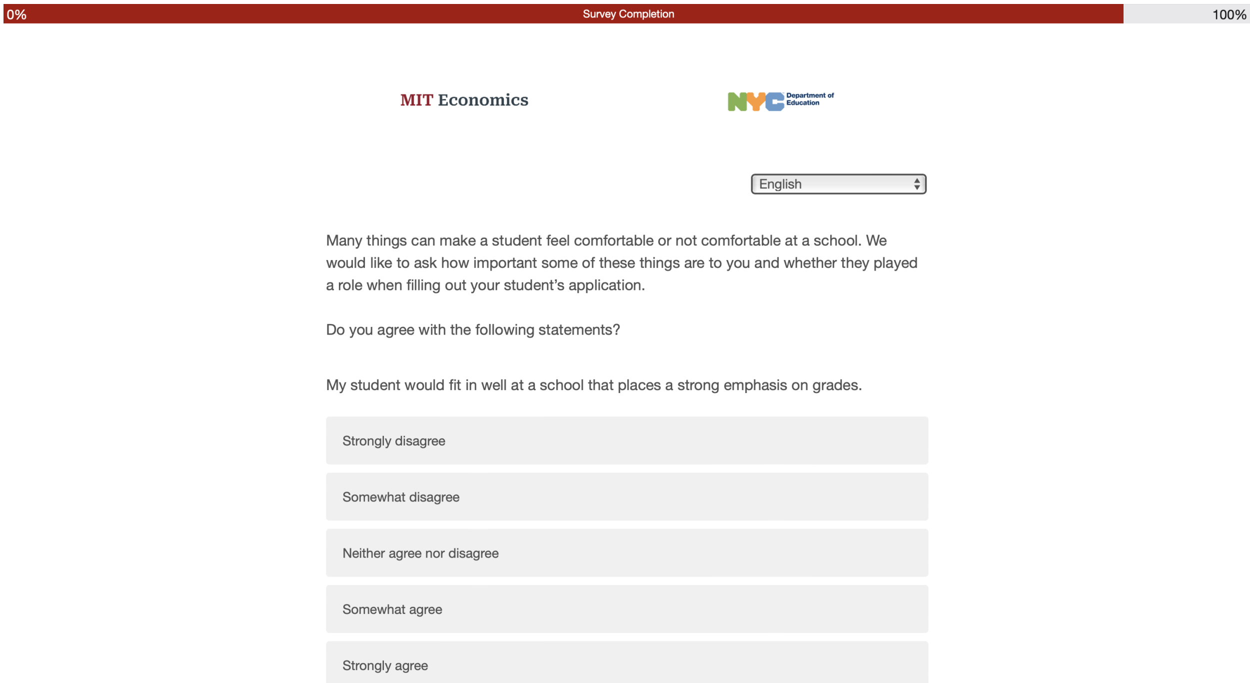


Figure B.2.47: Question 19a


0% Survey Completion 100%

MIT Economics

NYC Department of Education

English

Many things can make a student feel comfortable or not comfortable at a school. We would like to ask how important some of these things are to you and whether they played a role when filling out your student's application.

Do you agree with the following statements?

My student would fit in well at a school that places a strong emphasis on grades.

Strongly disagree

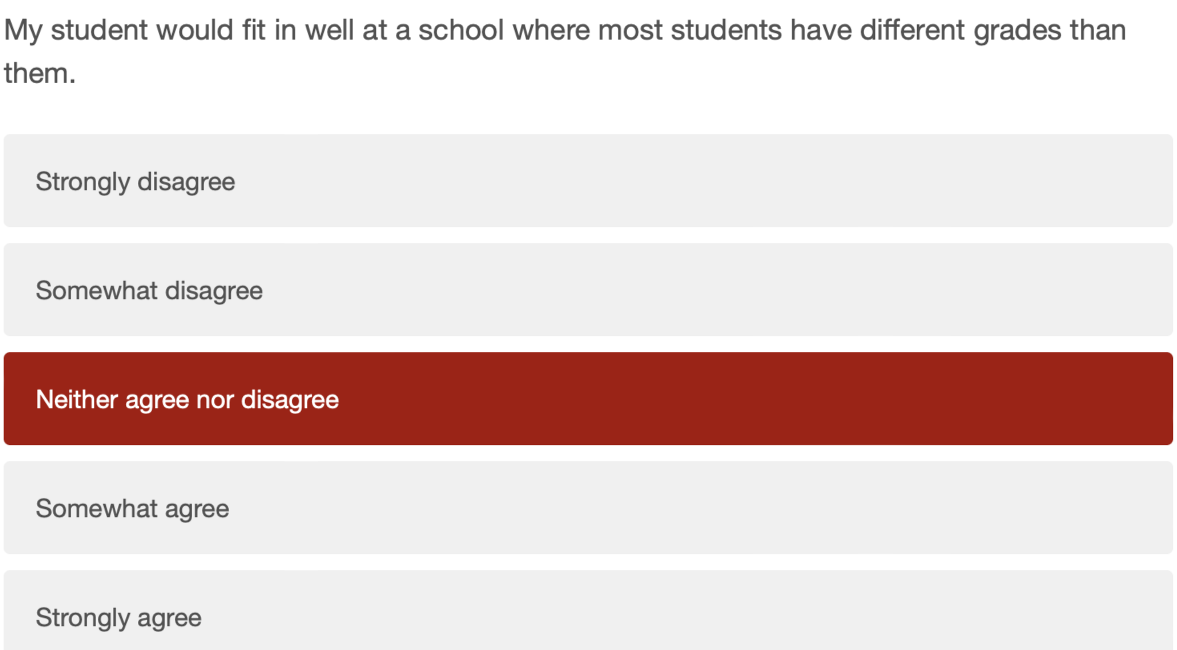
Somewhat disagree

Neither agree nor disagree

Somewhat agree

Strongly agree

Figure B.2.48: Question 19b


My student would fit in well at a school where most students have different grades than them.

Strongly disagree

Somewhat disagree

Neither agree nor disagree

Somewhat agree

Strongly agree

Figure B.2.49: Question 19c

My student would feel like they belong in a school even if most students are from a different race or ethnicity.

- Strongly disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Strongly agree

Figure B.2.50: Question 20a

My student is likely to be treated negatively by their classmates based on their race.

- Strongly disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Strongly agree

Figure B.2.51: Question 20b

My student is likely to be treated negatively by their teachers based on their race.

- Strongly disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Strongly agree

→

Powered by Qualtrics

Figure B.2.52: Question 20c

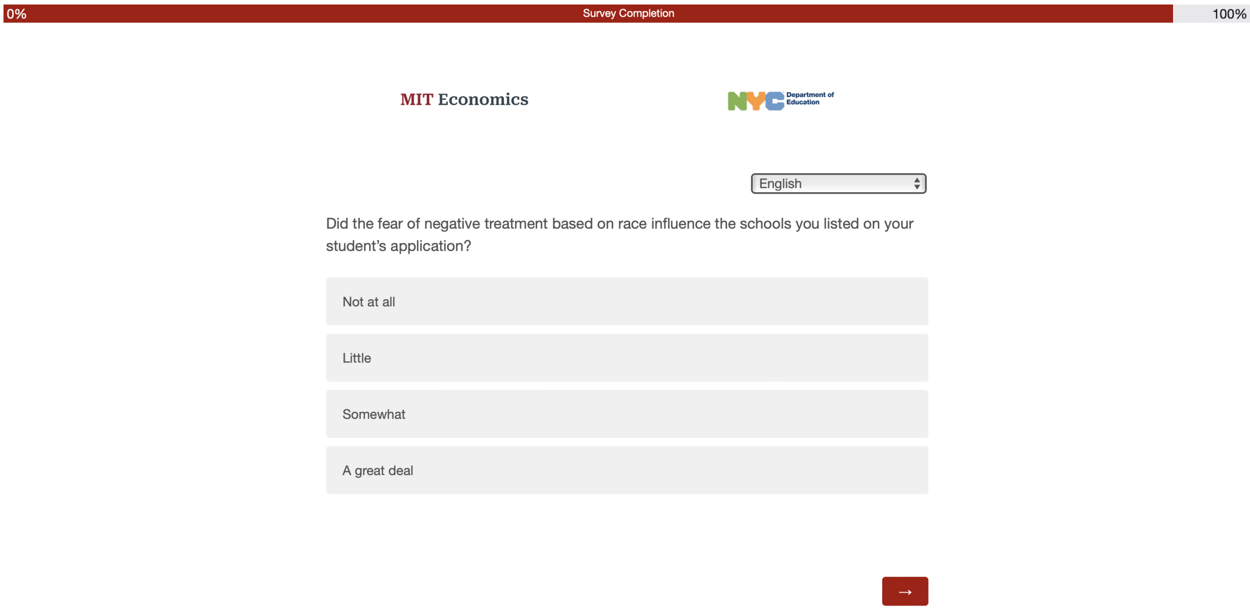
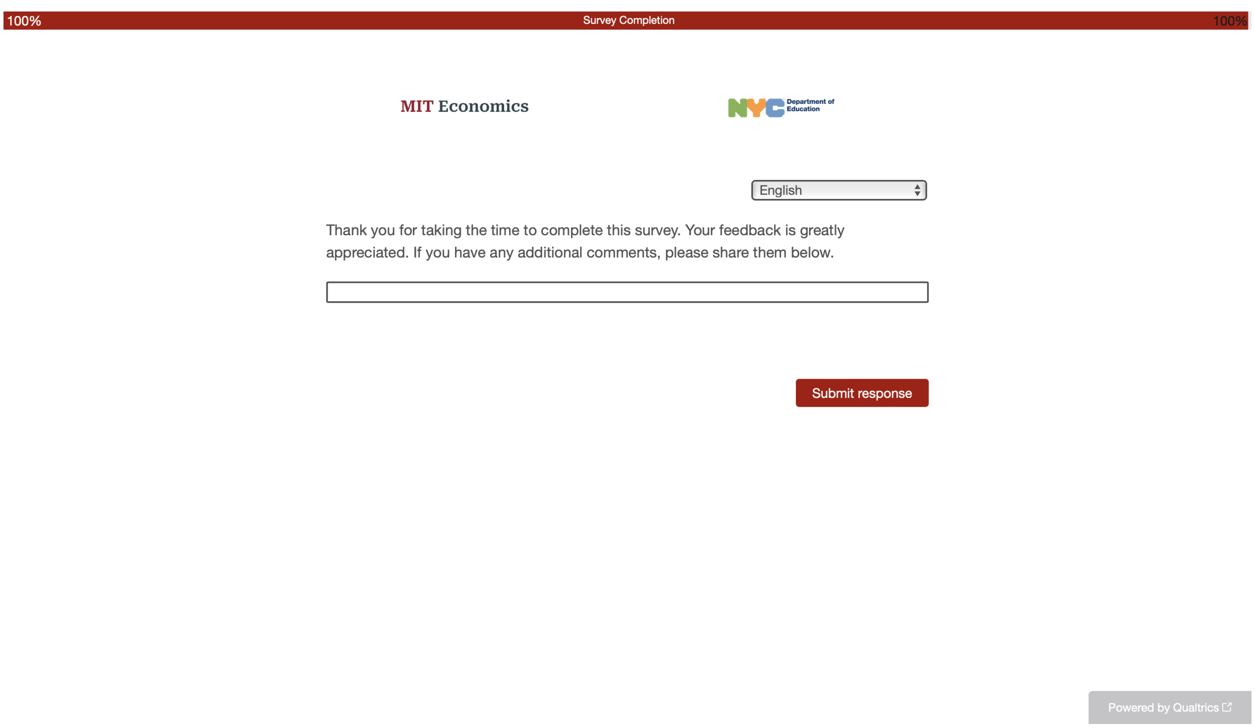


Figure B.2.53: Question 21



Thank you message

0% Survey Completion 100%

MIT Economics

NYC Department of Education

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

Powered by Qualtrics

Information questions - Version 2

Figure B.2.54: Question Q10 - Version 2

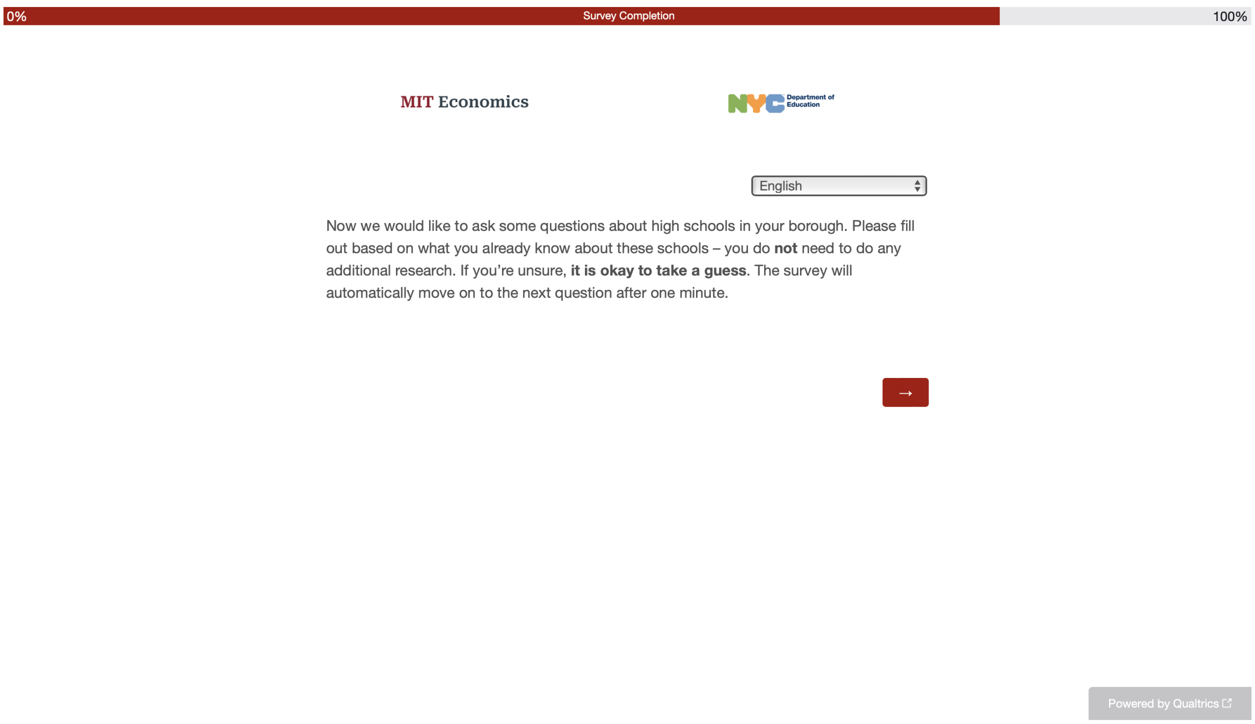


Figure B.2.55: Question Q10a - Version 2

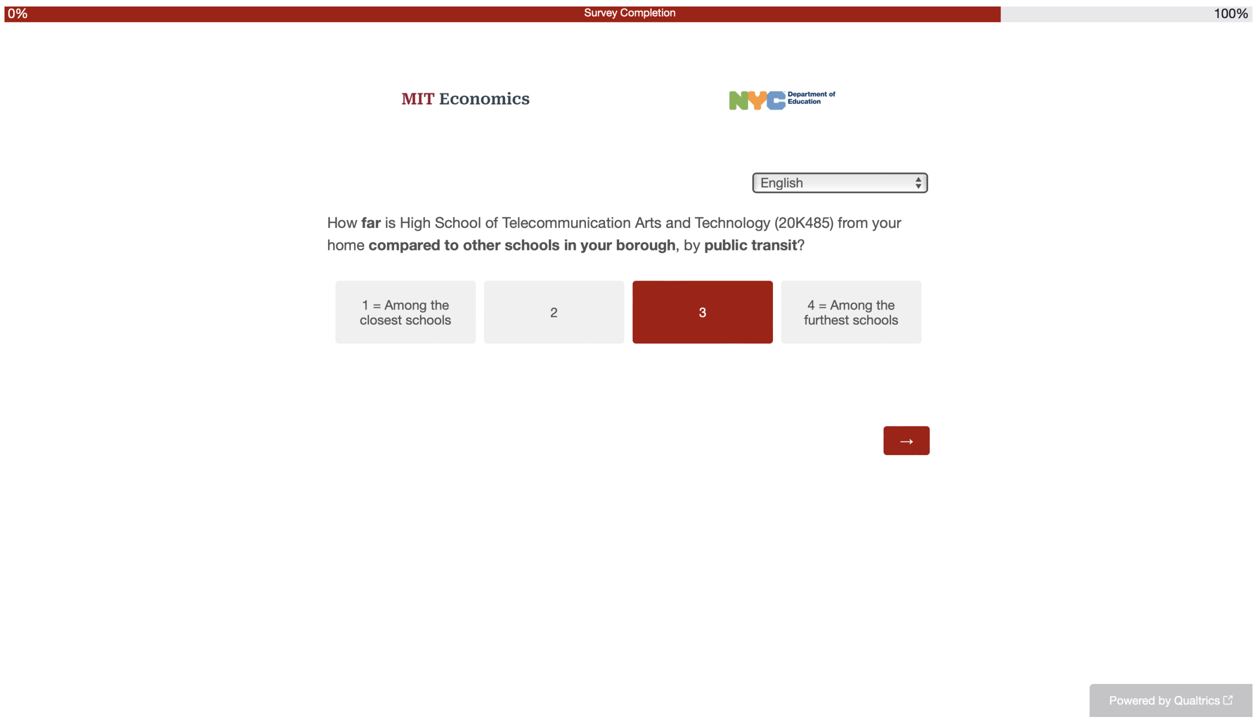


Figure B.2.56: Question Q10b - Version 2

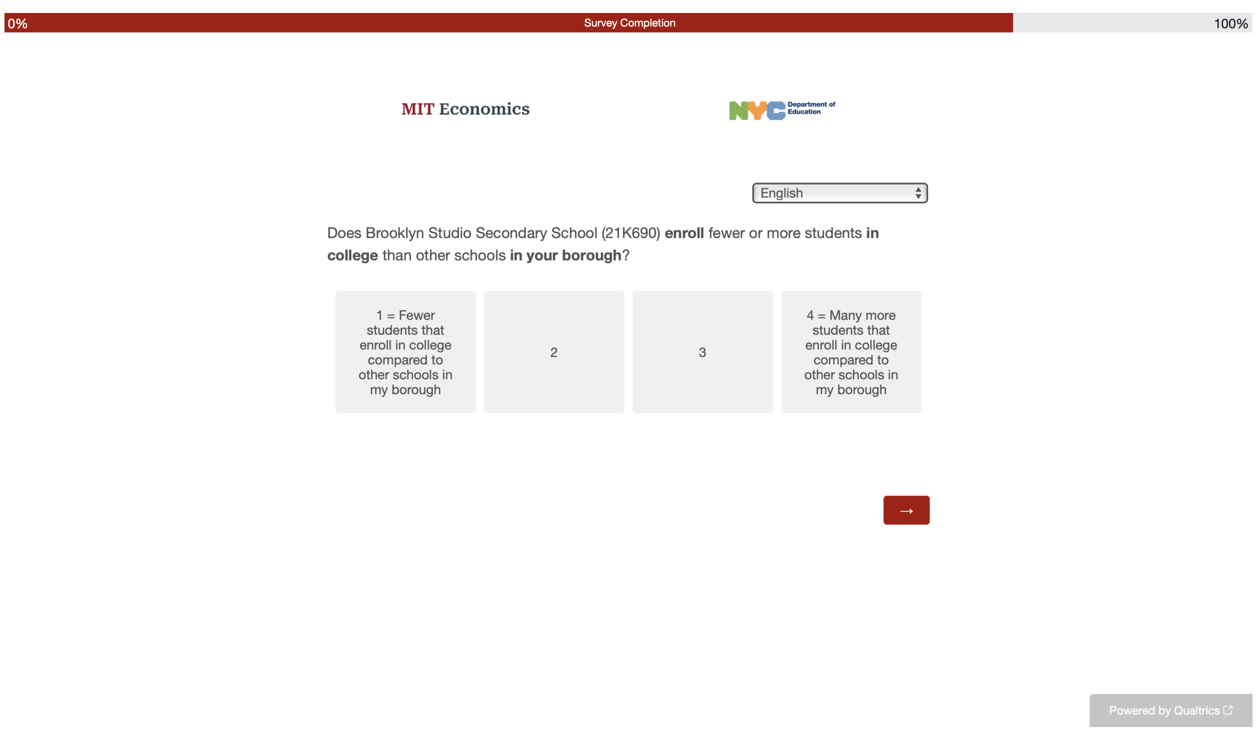


Figure B.2.57: Question Q10c - Version 2

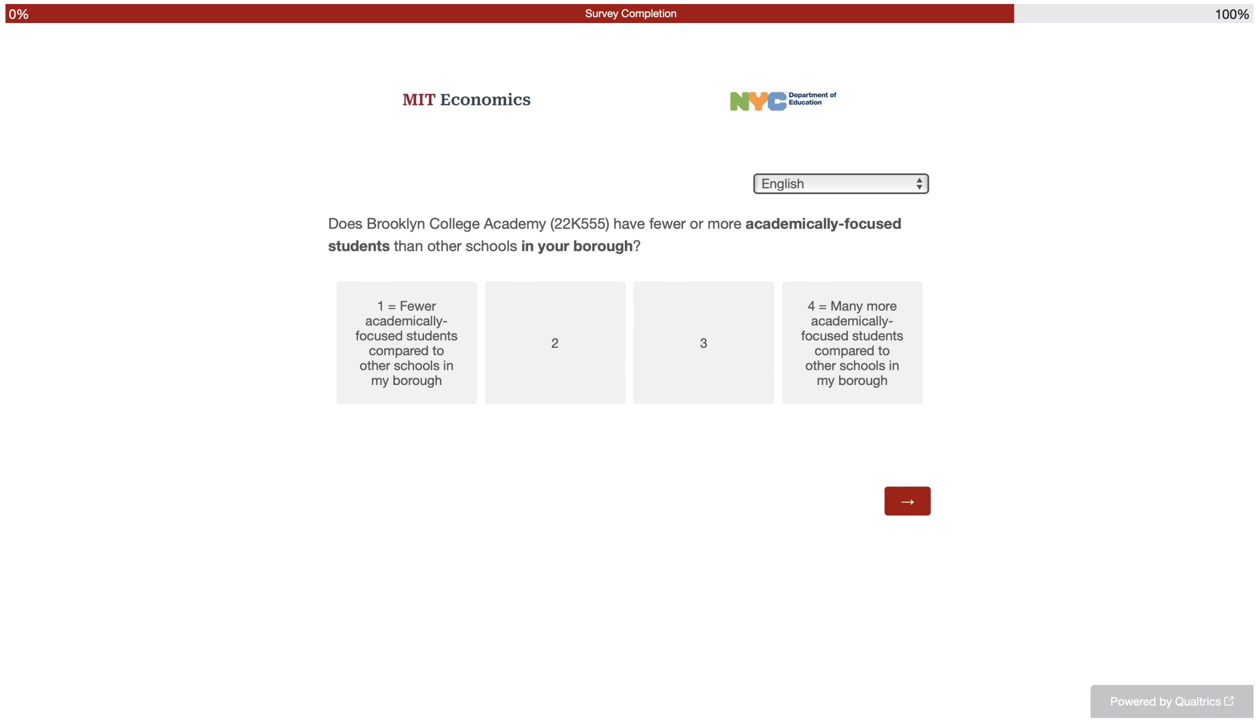


Figure B.2.58: Question Q10d - Version 2

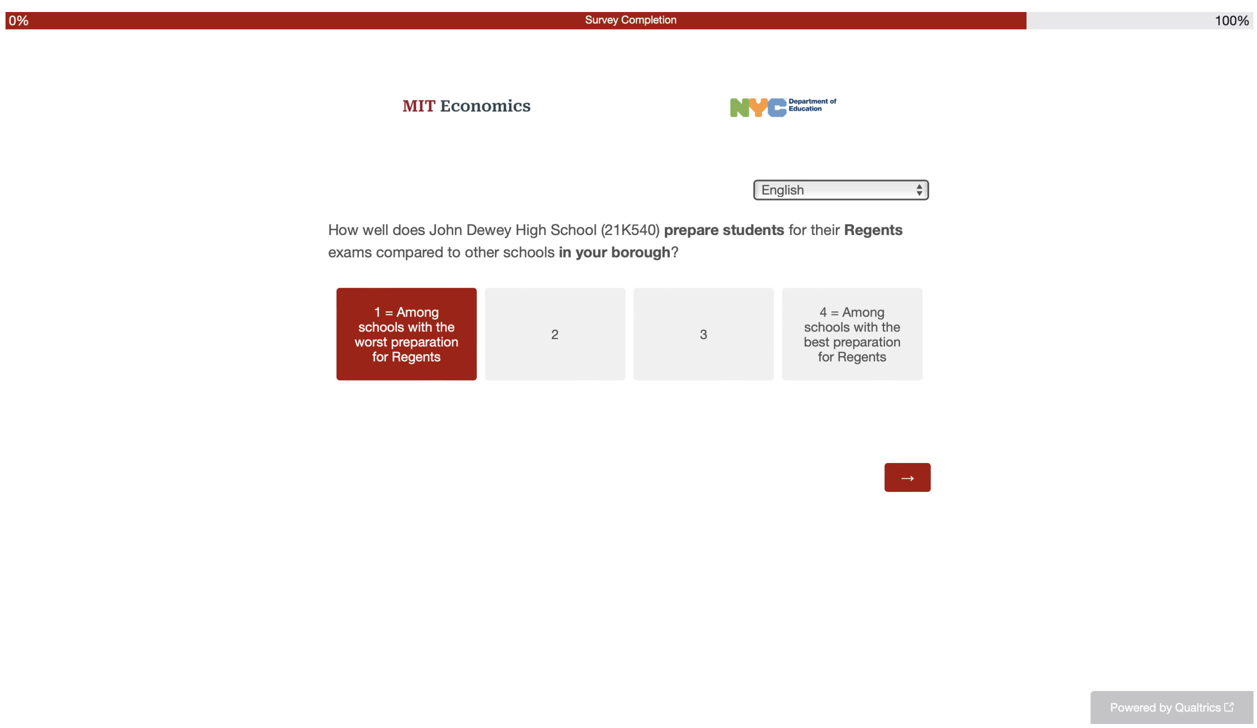


Figure B.2.59: Question Q10f - Version 2

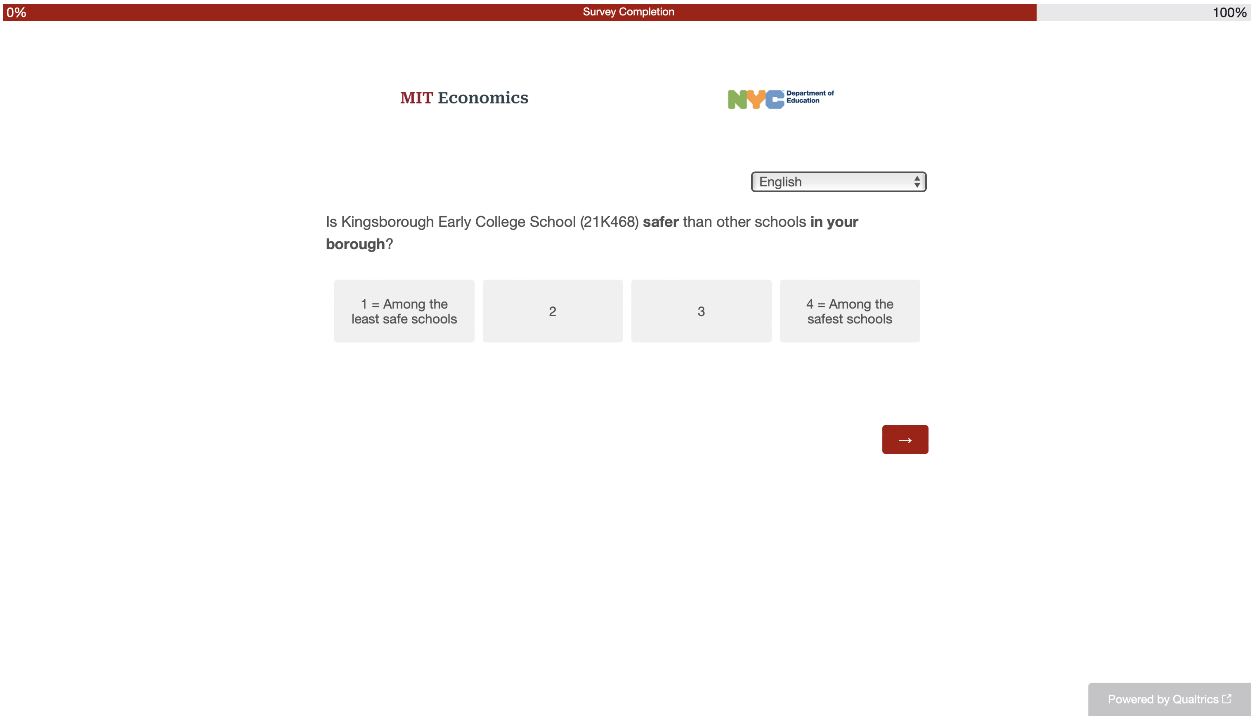
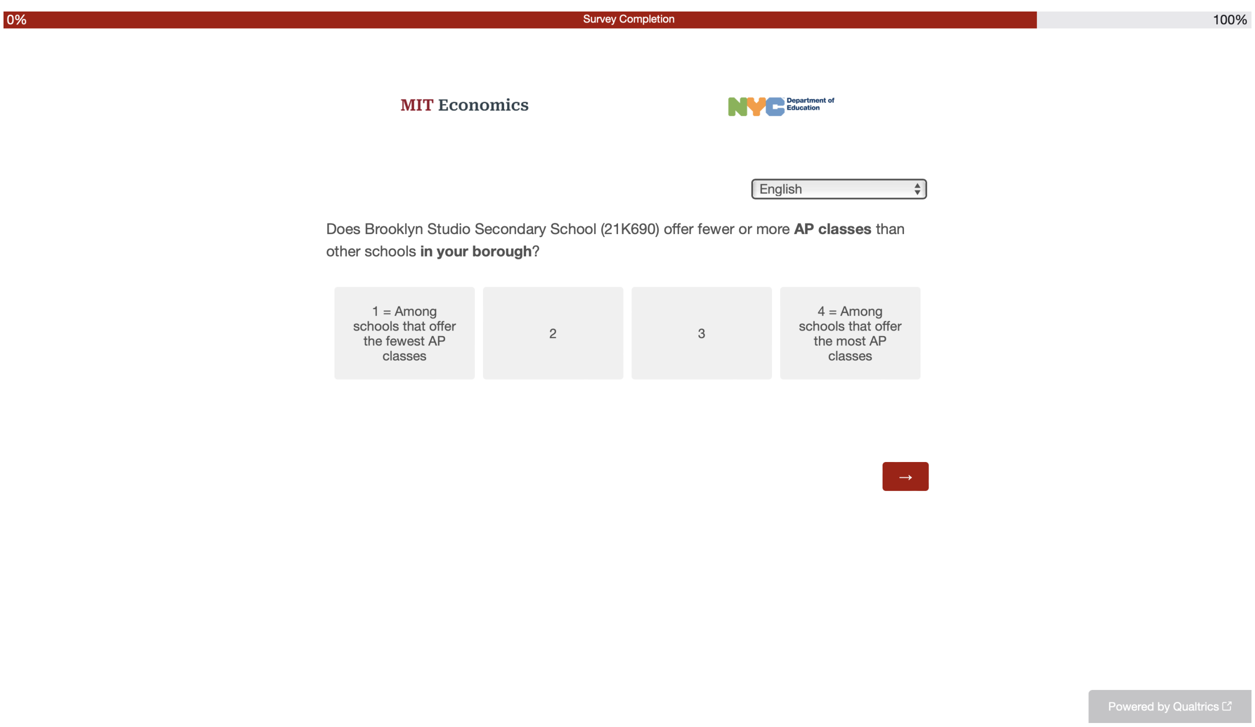


Figure B.2.60: Question Q10g - Version 2



B.2.4 Survey block design and randomization to survey version

Description of survey blocks

We grouped all the questions by type and created five different blocks (1a, 1b, 2, 3a, 3b), which are shown in Table B.2.11. Block 1 consists of the information questions. Block 2 includes questions about student aspirations, beliefs about student academic performance, and knowledge of tiebreakers and how they affect application decisions. Block 3 is the vignette experiment described further in Section 2.5.1 of the paper. It consists of questions about school preferences and perceptions of discrimination and has two versions: 3a and 3b. The version 3a includes precise academic information, while 3b presents imprecise academic information.

Randomization of participants to survey version

To reduce the time it takes to complete the survey and increase participation, we devised eight different survey versions by creating different combinations of the five question blocks. Each survey version consisted of between 31-35 questions (see Table B.2.12). All survey versions included the consent to participate, general questions, and the end-of-survey comment box. All potential participants had an equal probability of receiving any of the eight survey versions (12.5% each). The marginal probability for each block was thus 75% for block 1 (37.5% for 1a and 37.5% for 1b), 50% for block 2, and 75% for block 3 (37.5% for 3a and 37.5% for 3b). The detailed distribution of blocks to survey version is shown in Table B.2.12).

Table B.2.13 evaluates the covariate balance and attrition rates by survey version. Among all the balance regressions conducted, the majority show no statistically significant relationship between survey version assignment and the covariates. Similarly, in most of the response attrition regressions, the coefficients do not show statistical significance. These attrition findings remain consistent for all participants and when segmenting the sample by white and Asian as well as by Black and Hispanic (minority). The results confirm that the survey randomization successfully achieved the expected balance across the covariates. Regarding attrition, there is no statistically significant difference in the response rate observed by survey version among all potential participants. Two small differences are evident when dividing the sample by race. In survey version 8, white and Asian potential participants are slightly more likely to respond. Similarly, in survey version 3, Black and Hispanic potential participants show a slightly higher likelihood of responding.

Table B.2.11: Survey Questions by Block and Type

Type	Description	Number of questions and type
	General Block	
Age verification	Question to ensure participant is old enough.	**1 checkbox question** (Q0)
General questions	Questions about the relationship with the student, who played the most important role in the application, sources of information, the importance of going to school with friends, attention check, dream school, important aspects when choosing a school.	**15 possible questions:** - 9 multiple choice (Q1, Q2, Q3, Q3b, Q4, Q5, Q6, Q8a, Q9) - 2 open-ended (Q8a.2, Q3b.2) - 1 Yes/No question (Q7a): If "Yes," 1 extra multiple choice question (Q7c); if "No," 2 extra multiple choice questions (Q7b, Q7c.2)
End of survey	Question to leave any comments.	**1 open-ended question** (Q21)
	Block 1a	
Information (version 1)	Questions comparing two high schools in terms of commuting time by public transportation, academically focused students, college enrollment, Regents preparation, safe environment, and AP courses.	**6 multiple choice questions with two options each** (Q10a, Q10b, Q10c, Q10d, Q10f, Q10g)
	Block 1b	
Information (version 2)	Questions comparing a high school to the ones in the borough of residence in terms of commuting time by public transportation, academically focused students, college enrollment, Regents preparation, safe environment, and AP courses.	**6 multiple choice questions following 1-4 Likert scale** (Q10a_v2, Q10b_v2, Q10c_v2, Q10d_v2, Q10f_v2, Q10g_v2)
	Block 2	
Beliefs on academic performance and admission probability	Questions about beliefs on student 7th grade grades compared to all students in the middle school and the city, and about likelihood to admission to a school.	**3 multiple choice questions** (Q11a, Q12, Q13)
Aspirations for the student	Questions about the importance of going to college, and aspirations for the highest level of education.	**2 multiple choice questions** (Q14a, Q15)
Tiebreaker knowledge	Questions about knowledge of the tiebreaker number and how that affected the application.	**3 possible multiple choice questions:** - 1 Yes/No question (Q16a): If "Yes," 1 extra Yes/No question (Q16b); if "No," 1 extra Yes/No question (Q16c)
	Block 3a	
Preferences for attributes (experiment, version 1)	Two types of questions, the first belongs to a vignette experiment with hypothetical schools that varied by safety rating, academic performance ratings, and racial composition (read more on Section 2.5.1). The second type of question is about perceived race-based discrimination.	**10 possible questions:** - 9 multiple choice (Q17a, Q17b, Q18a, Q18b, Q19a, Q19b, Q19c, Q20a, Q20b) - 1 extra multiple choice if the response to any of the race-related questions was neutral or some degree of agreement (Q20c).
	Block 3b	
Preferences for attributes (experiment, version 2)	Two types of questions, the first belongs to a vignette experiment with hypothetical schools that varied by safety rating, academic performance ratings, and racial composition (read more on Section 2.5.1). The second type of question is about perceived race-based discrimination.	**10 possible questions:** - 9 multiple choice (Q17a, Q17b, Q18a, Q18b, Q19a, Q19b, Q19c, Q20a, Q20b) - 1 extra multiple choice if the response to any of the race-related questions was neutral or some degree of agreement (Q20c).

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

Table B.2.12: Eight Survey Versions and Their Respective Block Combinations

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

Notes: This table shows the survey blocks included in each of the eight survey versions. Specific questions within each block are detailed in Table B.2.11. The third column provides the total number of questions for each survey version.

Table B.2.13: Survey Attrition and Covariate Balance

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

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

B.2.5 School Selection for Randomized Survey Questions

Definition of school attributes

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

Attributes:

- **Demographics:** Ethnic/racial composition of students enrolled in school during the 2021-2022 school year, using all grades 9-12. In particular, we care about the share of white and Asian students (or the share of Black and Hispanic students) in the school.

- **Baseline scores:** Average (standardized) 7th grade test scores of the students enrolling in 9th grade in 2020-2021, by school. This means the test scores are typically measured in 2018-19 SY.

- **Popularity:** Popularity is the share of applicants rejected to applicants accepted for each program at each school in the 2022 admission cycle. We aggregate at the school level using a weighted average across programs at the school, with weights proportional to program capacities. The data used includes schools from any of the five NYC boroughs: Bronx (X), Brooklyn (K), Manhattan (M), Queens (Q), Staten Island (R).

- **Admission method:** We consider as screened schools those that in the 2022-2023 program crosswalk had at least one program that screened students on the basis of academics or both language and academics. These are the high school programs available for the 2023-2024 school year.

- **Language and AP stem classes:** Number of language classes and AP classes in STEM subjects offered by each school.

- **College attendance:** Share of students enrolling in college within 6 months of (on time) graduation per school for 2020-21 SY.

- **Safety:** Percent of students that felt safe in the hallways, bathrooms, locker rooms, and cafeteria by school during the 2019-20 SY.

- **Size:** Total enrollment count at school for grades 9-12 in the 2021-22 SY.

- **Applicants per seat:** Total number of applicants at the school (regardless of whether they got in a preferred school or not) per seat in 2022 admission cycle. This is a school-level measure.

- **Regents VA:** OLS VA on Algebra 1 and ELA Regents using test scores from years 2013 to 2017 cohorts (cohort = fall of 9th grade) and 7th grade baselines.

- **College VA:** OLS VA on a dummy for whether a student enrolls in any type of college using data from 2013 to 2016 cohorts and 7th grade baselines.

Districts' school choice set construction based on school characteristics

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

1. Start from schools in the 2021-2022 high school directory and keep only those in the 2022-2023 program crosswalk.

2. Drop specialized schools, special districts (75 and 79), and home schools.

3. For each district, take a subset of schools that:

   - are in the same borough, or
   - are out of borough but to which at least 1% of students in the district applies in the 2022 cycle.

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

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

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

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

- In the top 20 schools in terms of applicants per seat among students residing in the borough.

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

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

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

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

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

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

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

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

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

- **School 9:** A school that does not screen students on the basis of academics. For each district, we randomly selected two such schools from the district choice set as follows: one with a high share of white and Asian students, and one with a lower share of white and Asian students. If the restrictions returned an empty set, we selected the school with the highest (lowest) share of white students among unscreened schools. Finally, we randomized at the student level between these two unscreened schools.

- **School 10:** A large school. That is, a school with more than 622 students, which corresponds to the top 25% of schools in the city-wide distribution. For each district, we randomly selected two such schools from the district choice set as follows: one with a high share of white and Asian students, and one with a lower share of white and Asian students. If the restrictions returned an empty set, we selected the school with the largest size among high-white (low-white) schools. If empty again, we selected the school with the highest (lowest) share of white students among large-size schools. Finally, we randomized at the student level between these two large schools.

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

We measure information about schools by asking to compare two schools along the following school characteristics: baseline test scores, college enrollment rates, Regents VA, college VA, language and ap stem classes. For each district and each school characteristic, we selected four pairs of schools:

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

In each pair, the first school is the one with the highest value of the school characteristics of interest. With high-white we mean schools with a share of white and Asian students above 26%, corresponding to the 25% of schools with the highest share of white and Asian students in the city.

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

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

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

- **Baseline test scores:** Different by at least 0.33σ in the average baseline test score means of incoming students. We use an average of mean Math and mean ELA test scores for each school.

- **College enrollment rates:** Different by at least 5pp.

- **Regents VA:** Different by at least 0.3σ in the average Regents VA. We use an average of Regents Algebra VA and Regents ELA VA for each school.

- **College VA:** Different by at least 5pp.

- **Language classes:** Discrete difference (at least 1 more/less class).

- **AP stem classes:** Discrete difference (at least 1 more/less class).

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

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

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

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

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

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

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

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

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

The description of the vignette experiment is on Section 2.5.1 of the paper. The experiment includes a total of 24 possible vignettes, also referred to as school cards. First, 16 school cards show 3 school characteristics (called the "3-factor list"): academics (x2), safety (x2), and racial composition (x4). Second, we have 8 school cards that show 2 school characteristics (called the "2-factor list"): safety (x2) and racial composition (x4). Regarding academics and safety, hypothetical schools had either high-safety or low-safety ratings. In terms of student demographics, hypothetical schools had either a balanced racial composition representative of the school district, a majority of Black students, a majority of Hispanic students, or a majority of white or Asian students. The average school characteristics are in Table B.2.14. Examples of the 2- and 3-factor cards are in Figure B.2.61.

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

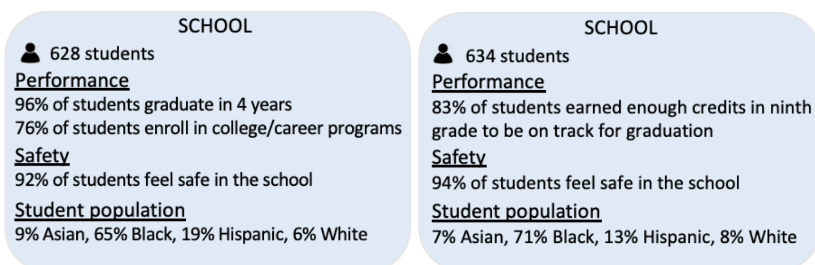
Finally, we randomized at the student level whether the student would receive the vignette with three or with two factors. We assigned 60% of students to the 3 factor version of the questions.

Table B.2.14: Different School Characteristics for the Vignette Experiment

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

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

Figure B.2.61: School Cards for Vignette Experiment



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

Appendix C

Appendix to Chapter 3

C.1 Appendix Tables and Figures

Table C.1.1: Clauses in the Intuitive Definition of Female-Centric Amenities

Group	Clause Type	Description
Leaves		
	Abortion leave	Leave in cases of miscarriage/abortion
	Adoption leave	Leave following the adoption of a child
	Maternity leave	Leave concerning the birth of a child
	Paid leave	Leave during which worker receives normal pay
	Unpaid leave	Leave during which worker does not receive normal pay
	Other: holidays and leaves	Provisions on holidays/leaves outside predefined clause types
	Female workforce	General provisions concerning female workers
Maternity and childcare		
	Childcare assistance	Payments to assist with childcare support
	Maternity assistance	Payments to assist with becoming a mother
	Abortion protections	Employment protections concerning miscarriage/abortion
	Maternity protections	Employment protections for mothers
	Paternity protections	Employment protections for fathers
	Policies for dependents	Workplace benefits that apply to dependents
Workplace harassment and discrimination		
	Sexual harassment	Rules/penalties pertaining to harassment in the workplace
	Equal opportunities	Initiatives/statements on equality of opportunity for workers
Flexibility and part-time work		
	Workday controls	Rules restricting the duration of the workday
	Special shifts	Work shifts for subgroups of workers, e.g., women, minors, students
	On-call	Rules on workers' availability outside of the normal workday
	Uninterrupted shifts	Rules concerning back-to-back shifts
	Part-time contracts	Directives on temporary/part-time employment contracts

Notes: Table lists the *Sistema Mediador* clause types used in our intuitive definition of female-centric amenities. The descriptions provided in this table are purposefully vague—clauses of a given type can vary to some degree. The clauses were chosen based on the content of CUT's fight plan and the existing literature on workplace amenities valued by women, restricting ourselves to only 20 clause types.

Table C.1.2: Examples of Female-Centric Amenities

Childcare assistance	The company will reimburse all female employees, the monthly amount of R$ 110, as a "day care allowance", per child up to 6 years old. This benefit applies to any employee with custody of the child(ren).
Absences	The employee will receive full pay for absences upon proof of the following cases: a) bereavement (5 consecutive days); b) hospitalization of direct family or legal dependents; c) medical and dental consultations; d) marriage (5 working days)
Adoption leave	The employee who adopts or obtains legal custody for adoption will be granted maternity leave as follows: a) 120 days for children up to 1 year old; b) 60 days, for children from 1 to 4 years old; c) 30 days for children from 4 to 8 years old.
Other: holidays and leaves	The start of vacations cannot coincide with Saturdays, Sundays, holidays, or days already compensated. Vacations will start on the first working day of the week, communicated to the union within 10 working days by the company.
Seniority pay	The company will pay the employee who completes 5 uninterrupted years of work an additional 5% per length of service payable monthly, calculated on the monthly fixed base salary.

Notes: Table lists examples of CBA clauses from the top 5 clause types selected as "female-centric" or "male-centric" based on our data-driven approach—refer to Section 3.3.2 for details on the data-driven approach. The clauses were selected based on the number of unique tokens appearing in the clause that are within the top 20 TF-IDF tokens of each specific clause type.

Table C.1.3: Examples of Male-Centric Amenities

On-call pay	The company will pay an additional 35% of the normal hours to employees, when scheduled to be on-call. This additional pay will not apply when the on-call becomes a service actually provided, in which case overtime will be due.
Life insurance	The company will maintain group life insurance, guaranteeing a single and total indemnity of at least R$ 10,000 in the event of death or permanent disability of the employee resulting from an accident at work.
Strike procedures	The union assumes formal commitment not to promote or encourage stoppages, except in cases of non-compliance with clauses of this agreement or current laws, and even so, only after communicating the transgressions in writing to the employers.
Other: protections for injured workers	The company will communicate to Social Security, and subsequently to the union, injuries incurred by employees at the company or while commuting to/from work.
Profit sharing	The company will maintain a Profit Sharing Program with the amount made available for payment may be up to 1 nominal salary per employee. The payment period after the calculation of the results will be the month of February.

Notes: Table lists examples of CBA clauses from the top 5 clause types selected as “female-centric" or “male-centric" based on our data-driven approach—refer to Section 3.3.2 for details on the data-driven approach. The clauses were selected based on the number of unique tokens appearing in the clause that are within the top 20 TF-IDF tokens of each specific clause type.

Table C.1.4: Robustness of Data-Driven Female-Centric Amenities

Clause type	**Times selected: data-driven** (out of 6 methods)	**Selected in baseline data-driven approach:** (no state and industry FEs)	(state and industry FEs)
Childcare assistance	6	1	1
Absences	6	1	1
Adoption leave	6	1	0
Other: holidays and leaves	6	1	1
Seniority pay	6	1	1
Maternity protections	6	1	1
Paid leave	6	1	1
Night pay	6	1	0
Abortion leave	6	1	0
Policy for dependents	6	1	0
Waiving union fees	6	1	1
Salary adjustments/corrections	6	1	0
Renewal/termination of the CBA	5	1	0
Nonwork-related injury protections	5	1	0
Extension/reduction of workday	5	1	1
Medical exams	5	1	0
Unionization campaigns	4	1	0
Abortion protections	4	1	0
Adoption protections	4	0	0
Guarantees to union officers	3	1	1
Health education campaigns	3	1	0
Military service protections	3	0	1
Separation/dismissal	2	0	1
Other employment protections	2	0	0
Awards	1	0	0
Moral harassment	1	0	1
Maternity leave	1	0	0

Notes: Table lists all of the clauses identified as female-centric in any of the 6 methods implemented based on the estimation of Equation (3.2). Methods vary in 1) the sample of establishments covered by sectoral CBAs used, i.e., a random sample or the full sample; and 2) the measure of PageRank values used to determine gender gaps, i.e., normalized, non-normalized, or rankings. The initial column simply shows the number of times the clause is picked as female-centric by one of these 6 methods (clauses in the table are sorted in descending order as per the values of this column). The next column is an indicator for whether the clauses is selected as a female-centric by the baseline method, i.e., using a random sample and normalized PageRanks. The final column is an indicator for whether the clause is selected as female-centric by the baseline method but where the lasso includes state and industry fixed effects. Note that the Spearman correlation of the coefficients on clauses using the data-driven lasso approach versus an OLS using these same clauses but adding state and industry fixed effects is 0.56 with p-value below 0.01.

Table C.1.5: Robustness of Data-Driven Male-Centric Amenities

Clause type	**Times selected: data-driven** (out of 6 methods)	**Selected in baseline data-driven approach:** (no state and industry FEs)	(state and industry FEs)
On-call pay	6	1	1
Life insurance	6	1	1
Strike procedures	6	1	1
Other: protections for injured workers	6	1	1
Female workforce	6	1	1
Machine and equipment maintenance	6	1	1
Duration and schedule	6	1	1
Working environment conditions	6	1	0
Salary payment - means and timeframes	6	1	0
Hazard pay (danger risk)	6	1	0
Workday compensation	6	1	0
Tools and equipment	6	1	0
Profit sharing	5	1	1
Transfers	5	1	0
Safety equipment	5	1	0
Other assistances	5	1	0
Death/funeral assistance	5	1	0
Salary deductions	4	1	0
Equal opportunities	4	0	0
Collective vacations	3	1	0
Union fees	3	0	0
CIPA: accident prevention committee	2	1	1
Unpaid leave	2	0	0
Part-time contracts	2	0	0
Food assistance	1	0	0
Performance evaluation	1	0	0
Employment/hiring rules	1	0	0

Notes: Table lists all of the clauses identified as male-centric in any of the 6 methods implemented based on the estimation of Equation (3.2). Methods vary in 1) the sample of establishments covered by sectoral CBAs used, i.e., a random sample or the full sample; and 2) the measure of PageRank values used to determine gender gaps, i.e., normalized, non-normalized, or rankings. The initial column simply shows the number of times the clause is picked as male-centric by one of these 6 methods (clauses in the table are sorted in descending order as per the values of this column). The next column is an indicator for whether the clauses is selected as a male-centric by the baseline method, i.e., using a random sample and normalized PageRanks. The final column is an indicator for whether the clause is selected as male-centric by the baseline method but where the lasso includes state and industry fixed effects. Note that the Spearman correlation of the coefficients on clauses using the data-driven lasso approach versus an OLS using these same clauses but adding state and industry fixed effects is 0.56 with p-value below 0.01.

Table C.1.6: Establishment Descriptives—RAIS vs. Analysis Samples

	All RAIS (1)	Amenities sample (2)	Difference p-value (3)	RAIS: employ men and women (4)	Establishment sample (5)	Difference p-value (6)
Employment and firm characteristics						
Size	16.19	143.11	0.00	31.87	150.22	0.00
Share women	0.45	0.38	0.00	0.45	0.40	0.00
Employs both men and women	0.46	0.82	0.00	1.00	1.00	1.00
Single person firm	0.27	0.04	0.00	0.00	0.00	1.00
Single establishment firm	0.77	0.65	0.00	0.77	0.63	0.00
Sector						
Agriculture & extraction	0.09	0.04	0.00	0.05	0.03	0.00
Manufacturing	0.09	0.28	0.00	0.11	0.30	0.00
Construction & utilities	0.05	0.06	0.00	0.04	0.05	0.00
Commerce	0.39	0.23	0.00	0.41	0.24	0.00
Services	0.38	0.39	0.00	0.38	0.39	0.00
Region						
North	0.04	0.05	0.00	0.05	0.05	0.00
Northeast	0.16	0.12	0.00	0.16	0.12	0.00
Central	0.10	0.07	0.00	0.10	0.08	0.00
South	0.21	0.21	0.00	0.20	0.21	0.00
Southeast	0.49	0.56	0.00	0.49	0.54	0.00
N establishments	3,798,207	80,131		1,739,255	61,752	
N workers	61,492,768	11,467,760		48,564,436	9,276,475	
% workforce	100%	19%		79%	15%	

Notes: Table compares descriptive statistics of establishments in Brazil's formal sector (Column 1) and our analysis samples, i.e., th (Column 2) and the establishment sample (Column 5). The p-values of the differences between these samples are reported in Colu bottom of the table includes the number of unique establishments and workers in each sample, as well as the percentage of the present in the corresponding sample.

Table C.1.7: Treated and Control Establishments Descriptives

	Amenities sample			Establishment sample		
	Treatment	Control	Diff. p-value	Treatment	Control	Diff. p-value
	(1)	(2)	(3)	(4)	(5)	(6)
Employment and firm characteristics						
Size	198.21	127.03	0.00	200.37	135.95	0.00
Share women	0.36	0.38	0.00	0.38	0.40	0.00
Employs both men and women	0.83	0.82	0.00	1.00	1.00	1.00
Single person firm	0.03	0.04	0.00	0.00	0.00	1.00
Single establishment firm	0.66	0.65	0.11	0.64	0.63	0.06
Sector						
Agriculture & extraction	0.03	0.04	0.00	0.02	0.03	0.00
Manufacturing	0.32	0.27	0.00	0.33	0.29	0.00
Construction & utilities	0.08	0.06	0.00	0.06	0.04	0.00
Commerce	0.21	0.24	0.00	0.19	0.25	0.00
Services	0.37	0.39	0.00	0.39	0.38	0.04
Region						
North	0.04	0.05	0.00	0.05	0.06	0.00
Northeast	0.15	0.11	0.00	0.16	0.11	0.00
Central	0.09	0.06	0.00	0.11	0.07	0.00
South	0.22	0.20	0.00	0.22	0.20	0.00
Southeast	0.50	0.58	0.00	0.46	0.56	0.00
N establishments	18,103	62,028		13,677	48,075	
N workers	3,588,153	7,879,607		2,740,517	6,535,958	

Notes: Table compares descriptive statistics of establishments between the treated (Columns 1 and 4) and comparison groups (Columns 2 and 5) in our analysis samples, i.e., the amenity sample and the establishment sample. The p-values of the differences between the treated and comparison groups are reported in Columns 3 and 6. The bottom of the table includes the number of unique establishments and workers in each group.

Table C.1.8: Effect of CUT Reform on Negotiated Amenities (Cluster at Union-Level)

	Intuitive definition (female clauses)					Data-driven		
	All	Leave	Maternity	Harassment	Flexibility	Female	Male	F/(F+M+1)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Intensive margin (number)								
$D_i \times \delta_{year \geq 2015}$	0.157*	0.078**	0.042*	0.009**	0.028	0.301**	0.130	0.032*
	(0.083)	(0.040)	(0.023)	(0.004)	(0.031)	(0.144)	(0.159)	(0.018)
Mean outcome	0.95	0.25	0.24	0.02	0.44	1.58	2.55	0.15
Panel B: Intensive margin (unique clause types)								
$D_i \times \delta_{year \geq 2015}$	0.123*	0.047	0.042*	0.008**	0.027	0.154*	0.067	
	(0.067)	(0.031)	(0.022)	(0.004)	(0.021)	(0.080)	(0.095)	
Mean outcome	0.70	0.18	0.21	0.02	0.30	1.26	1.58	
Panel C: Extensive margin								
$D_i \times \delta_{year \geq 2015}$	0.017	0.012	0.020*	0.008**	0.022	0.034*	-0.001	
	(0.015)	(0.011)	(0.012)	(0.004)	(0.015)	(0.020)	(0.015)	
Mean outcome	0.31	0.12	0.15	0.02	0.23	0.36	0.46	
Panel D: As a share of all clauses								
$D_i \times \delta_{year \geq 2015}$	0.005	0.001	0.001	0.000	0.003	0.021	-0.003	
	(0.004)	(0.001)	(0.001)	(0.003)	(0.015)	(0.006)	(0.012)	
Mean outcome	0.05	0.01	0.01	0.00	0.03	0.07	0.14	
Observations	600,960	600,960	600,960	600,960	600,960	600,960	600,960	600,960

Notes: Table reports the coefficients for DID regressions—see Equation (3.3)—estimating the effect of the CUT reform on the female-centric and male-centric amenities included in CBAs. Panel A uses the total number of clauses per pair-year as an intensive margin measure. Panel B uses the sum of the corresponding unique clause types, capturing how the space of female (male) clauses grows or shrinks. Panel C uses an indicator for pair-year observations with at least one corresponding clause as an extensive margin measure. Panel D uses the share of corresponding clauses with respect to the total contract clauses, capturing how the composition of CBAs change. Under each panel we report the mean of the dependent variable among the treated at baseline (2014). The sample is the filled panel of establishment-union pairs by year. All columns control for pair fixed effects, as well as time-varying state and industry fixed effects. Standard errors are clustered at the union level, instead of at the establishment level, which reduces the number of clusters from around 80 thousand to about 4.4 thousand.

Table C.1.9: Effect of CUT Reform on Negotiated Amenities (CBA coverage in 2014)

	Intuitive definition (female clauses)					Data-driven		
	All	Leave	Maternity	Harassment	Flexibility	Female	Male	F/(F+M+1)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Intensive margin (number)								
$D_i \times \delta_{year \geq 2015}$	0.096***	0.044***	0.020***	0.005***	0.028***	0.121***	0.111***	0.009***
	(0.015)	(0.006)	(0.004)	(0.001)	(0.010)	(0.023)	(0.031)	(0.090)
Mean outcome	1.63	0.43	0.41	0.03	0.76	2.71	4.38	0.25
Panel B: Intensive margin (unique clause types)								
$D_i \times \delta_{year \geq 2015}$	0.070***	0.023***	0.021***	0.003***	0.022***	0.076***	0.050***	
	(0.010)	(0.004)	(0.004)	(0.001)	(0.005)	(0.014)	(0.016)	
Mean outcome	1.21	0.31	0.36	0.03	0.51	2.17	2.71	
Panel C: Extensive margin								
$D_i \times \delta_{year \geq 2015}$	0.019***	0.012***	0.010***	0.004***	0.021***	0.005*	0.009**	
	(0.003)	(0.002)	(0.002)	(0.001)	(0.003)	(0.003)	(0.003)	
Mean outcome	0.53	0.21	0.25	0.03	0.40	0.62	0.79	
Panel D: As a share of all clauses								
$D_i \times \delta_{year \geq 2015}$	0.005***	0.001***	0.001***	0.000***	0.003**	0.004***	0.001	
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)	
Mean outcome	0.08	0.01	0.01	0.00	0.06	0.11	0.25	
Observations	366,468	366,468	366,468	366,468	366,468	366,468	366,468	366,468

Notes: Table reports the coefficients for DID regressions—see Equation (3.3)—estimating the effect of the CUT reform on the female-centric and male-centric amenities included in CBAs. The sample is the filled panel of establishment-union pairs by year, restricted to establishment-union pairs with CBA coverage in 2014. Panel A uses the total number of clauses per pair-year as an intensive margin measure. Panel B uses the sum of the corresponding unique clause types, capturing how the space of female (male) clauses grows or shrinks. Panel C uses an indicator for pair-year observations with at least one corresponding clause as an extensive margin measure. Panel D uses the share of corresponding clauses with respect to the total contract clauses, capturing how the composition of CBAs change. Under each panel we report the mean of the dependent variable among the treated at baseline (2014). All columns control for pair fixed effects, as well as time-varying state and industry fixed effects. Standard errors are clustered at the establishment level.

Table C.1.10: Effect of CUT Reform on Female Amenities

	Female-Centric Clauses: Intensive Margin					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Intuitive definition						
$D_i \times \delta_{year \geq 2015}$	0.157***	0.157***	0.157***	0.194***	0.297***	0.096***
	(0.013)	(0.013)	(0.013)	(0.014)	(0.019)	(0.015)
Mean outcome	0.95	0.95	0.95	0.95	0.95	1.63
Panel B: Data-driven definition						
$D_i \times \delta_{year \geq 2015}$	0.301***	0.347***	0.262***	0.332***	0.417***	0.121***
	(0.021)	(0.026)	(0.017)	(0.022)	(0.030)	(0.023)
Mean outcome	1.58	2.05	1.17	1.58	1.58	2.71
Data-driven clauses	baseline	any	all	baseline	baseline	baseline
Geography-year FEs	state	state	state	microregion	micro×ind	state
CBA coverage in 2014	no	no	no	no	no	yes
Observations	600,960	600,960	600,960	600,960	600,960	366,468

Notes: Table reports the coefficients for DID regressions—see Equation (3.3)—estimating the effect of the CUT reform on female amenities included in CBAs. The dependent variable is the total number of clauses per pair-year as an intensive margin measure, with Panel A using the intuitive definition of female-centric clauses and Panel B using the data-driven approach. Columns (1)-(3) modify the dependent variable by changing the clauses that are chosen as female-centric in the data-driven approach: a) *baseline*: top 20 clauses using a random sample and normalized PageRank values for the gender gaps; b) *any*: counts any of the clauses selected across 6 approaches as female-centric; c) *all*: counts only those clauses that are selected in all 6 approaches as female-centric. Refer to Table C.1.4 for a list of the clauses used in each of these scenarios. Columns 4 adds more granular time-varying fixed effects at the geographic level, i.e., using microregion instead of state. Columns 5 uses a microregion-industry time-varying fixed effect. Column 6 requires that pairs are covered by a CBA at baseline to test whether effects are driven by changes in the amenities among units with active CBAs rather than by gains in coverage.

Table C.1.11: Effect of CUT Reform on Female Amenities

	Female-Centric Clauses: As a Share of All Clauses					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Intuitive definition						
$D_i \times \delta_{year \geq 2015}$	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Mean outcome	0.05	0.05	0.05	0.05	0.05	0.08
Panel B: Data-driven definition						
$D_i \times \delta_{year \geq 2015}$	0.021***	0.021***	0.022***	0.017***	0.011***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Mean outcome	0.07	0.08	0.04	0.07	0.07	0.11
Data-driven clauses	baseline	any	all	baseline	baseline	baseline
Geography-year FEs	state	state	state	microregion	micro×ind	state
CBA coverage in 2014	no	no	no	no	no	yes
Observations	600,960	600,960	600,960	600,960	600,960	366,468

Notes: Table reports the coefficients for DID regressions—see Equation (3.3)—estimating the effect of the CUT reform on female amenities included in CBAs. The dependent variable is the share of female-centric clauses among all clauses per pair-year, with Panel A using the intuitive definition of female-centric clauses and Panel B using the data-driven approach. Columns (1)-(3) modify the dependent variable by changing the clauses that are chosen as female-centric in the data-driven approach: a) *baseline*: top 20 clauses using a random sample and normalized PageRank values for the gender gaps; b) *any*: counts any of the clauses selected across 6 approaches as female-centric; c) *all*: counts only those clauses that are selected in all 6 approaches as female-centric. Refer to Table C.1.4 for a list of the clauses used in each of these scenarios. Column 4 adds more granular time-varying fixed effects at the geographic level, i.e., using micro-region instead of state. Columns 5 uses a microregion-industry time-varying fixed effect. Column 6 requires that pairs are covered by a CBA at baseline to test whether effects are driven by changes in the amenities among units with active CBAs rather than by gains in coverage. Standard errors are clustered at the establishment level.

Table C.1.12: Differential Effects by Gender for Incumbent Workers

	Stay at baseline employer (1)	Employed in formal sector (2)	Log wages (3)
$D_i \times \delta_{year \geq 2015}$	0.010***	0.002	-0.000
	(0.002)	(0.002)	(0.001)
$D_i \times \delta_{year \geq 2015} \times Female_i$	0.008***	0.005**	0.002
	(0.003)	(0.002)	(0.002)
Observations	55,658,796	55,658,796	46,668,757
R^2	0.63	0.44	0.90

Notes: Table reports the coefficients for the gender-pooled DID regression estimating the effect of the CUT reform on retention, formal sector employment, and wages of incumbent workers. Treatment status of incumbent workers is based on the CUT-affiliation of the union negotiating with their baseline (2014) employer. These workers are tracked wherever they go. The regression interacts treatment status with dummy variables for the post period (after 2014) and gender. Regressions include worker fixed effects, industry-year-gender fixed effects, microregion-year-gender fixed effects, and tenure-year-gender fixed effects. To make treatment effects in worker-level regressions interpretable as establishment-level averages, we weight each incumbent worker by the inverse of employment at their baseline employer. Standard errors are clustered by establishment and reported in parentheses.

Table C.1.13: Impact of CUT Reform on Worker Composition (Female)

	Share poached in (1)	Mean years of age (2)	Mean months of tenure (3)	Mean hours in contract (4)	Mean years of schooling (5)
$D_i \times \delta_{year \geq 2015}$	-0.001	-0.012	0.172	-0.033	-0.001
	(0.002)	(0.041)	(0.215)	(0.025)	(0.010)
Mean outcome	0.209	33.5	43.1	42.0	11.3
Observations	342,207	342,207	342,207	342,207	342,207

Notes: Table reports the coefficients for the establishment-level DID regression from Equation (3.3), comparing treated to comparison establishments on characteristics of their female workforce. An establishment is treated if the union with which it negotiates is affiliated to CUT in 2012. Each regression includes establishment fixed effects, industry-year fixed effects, and microregion-year fixed effects. Standard errors are clustered by establishment and reported in parentheses.

Table C.1.14: Welfare Estimation

		Women 20-35	All women	Men 20-35	All men
		(1)	(2)	(3)	(4)
$ln\phi_{t-1,t}$		0.044	0.059	-0.005	0.013
		(0.0062)	(0.0066)	(0.0048)	(0.0045)
Components breakdown:					
	$ln(\lambda_{t,t-1}) - ln(\lambda_{t-1,t})$	-0.012	-0.018	-0.005	-0.006
	$ln(\bar{w}_t^*) - ln(\bar{w}_{t-1}^*)$	0.015	0.022	-0.001	0.011
	$ln(\bar{S}_t^*) - ln(\bar{S}_{t-1}^*)$	-0.046	-0.058	0.013	0.001
η (calibrated)	1.015				
N establishments		58,417	60,651	59,438	60,651
N establishments in $\Omega_{t,t-1}$		45,331	47,195	46,182	47,195

Notes: Table reports the estimated welfare change for different groups of workers: women between 20 and 35 years old, all women, men between 20 and 35 years old, all men. It also reports estimates of the three components that make the welfare index, namely the Feenstra "new varieties" term $ln(\lambda_{t,t-1}) - ln(\lambda_{t-1,t})$, the change in the geometric average of the wages of non-CUT firms $ln(\bar{w_t^*}) - ln(\bar{w_{t-1}^*})$, and the change in the geometric average of the labor income shares of non-CUT firms. Standard errors in parenthesis come from the bootstrap procedure described in Appendix C.4.

C.2 Data Appendix

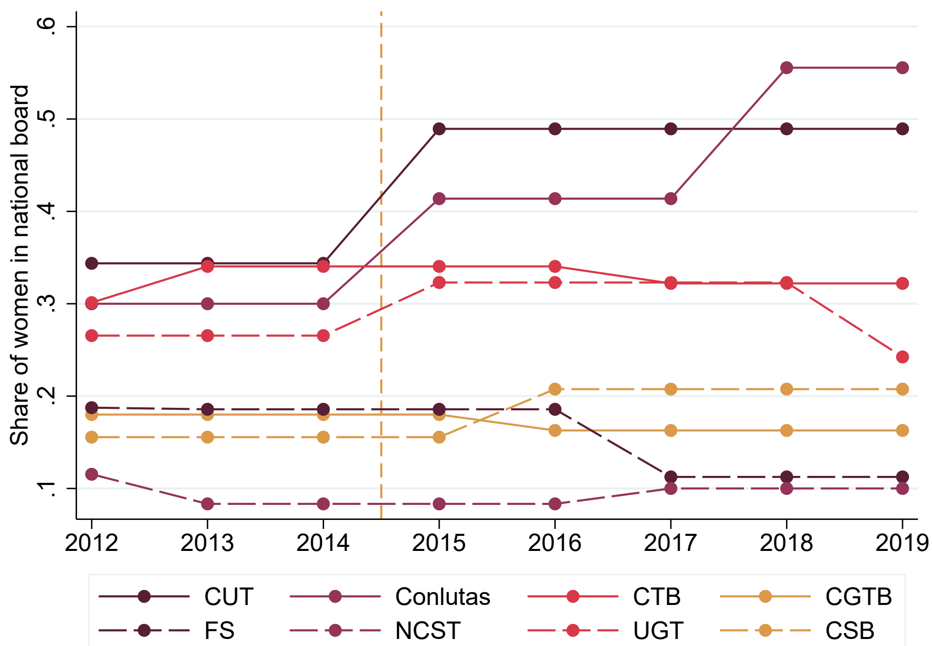
C.2.1 Sample construction

To analyze the CUT reform's impact on various outcomes, we construct three main analysis samples. The first is a sample to study changes in CBA clauses at the establishment-union pair level (henceforth, simply *pair* level). The second is a sample at the establishment level to study changes in the workplace. The third is a sample at the worker level used to track the labor market outcomes of incumbent workers. In addition to these three main samples, we also construct two panel datasets at the local union level and at the union central level to study the gender composition of their boards.

Amenities sample Amenities (on paper) are captured by CBA clauses signed by establishment-union pairs. We first construct a yearly panel of the new CBAs signed by a pair in a given year, i.e., new contracts. We then use this sample to construct a balanced panel containing the active clauses applying to a pair over time, i.e., filled panel.

1. New contracts: We construct this sample using the set of CBAs registered on *Sistema Mediador*. We restrict to valid, non-amendment, firm-level CBAs signed between 2012 and 2017 (inclusive). Each CBA contains information on who signs the agreement—the CNPJ identifiers

Figure C.1.1: Gender Parity in National Leadership by Union Central

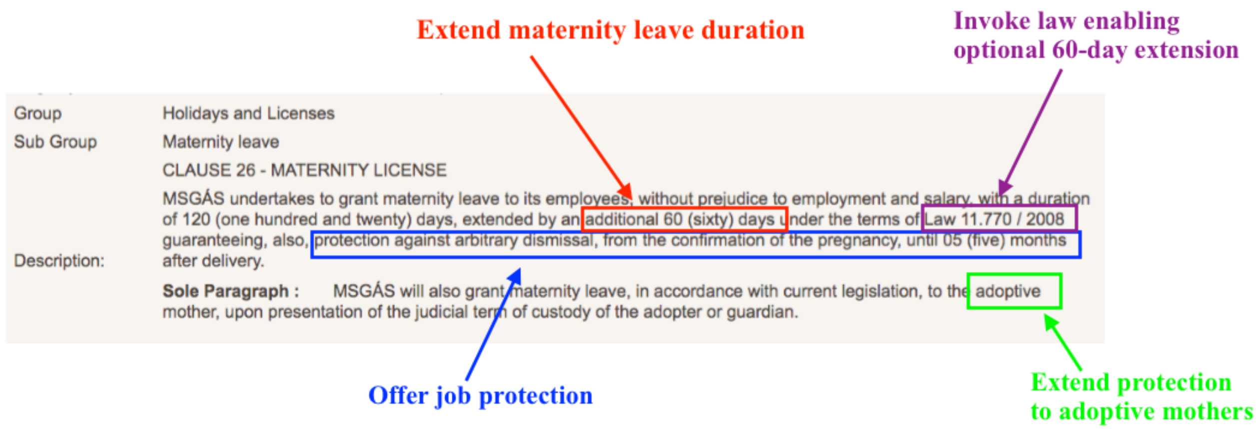


Notes: Figure plots the annual share of women on each union central's national executive committee (*Intersindical* is dropped due to missing information on its board). The line for CUT is the same as in Figure 3-2a, while the unweighted average of all other union centrals make up the other line reported in Figure 3-2a. Solid lines refer to "combative" union centrals, while dashed lines represent "cooperative" union centrals. The second largest union central and main competitor to CUT is *Força Sindical* (FS).

Figure C.1.2: Cover of CUT's Female-Centric "Fight Plan"

Notes: The 2015 CUT reform consisted of two parts. The first is a 50% quota for women in CUT's state and national executive bodies. The second is the adoption of a bargaining agenda more attentive to the needs of female workers. Figure C.1.2 is the cover page of the book of resolutions (or "fight plan") developed at the 2015 meeting of CUT Women to detail concrete strategies for achieving parity in practice at all levels of unions within CUT. It recommends steps for giving women more actual voice in all levels of the union—like representation on committees and a say in union's list of demands (or *pautas*). It also specifies amenities like maternity leave extensions and subsidized childcare to highlight during collective bargaining. This book of resolutions was subsequently adopted by delegates at the 2015 CUT National Congress (full text here). The word count for *mulheres* (women) in the National Congress book of resolutions increased from 46 in 2012 to 203 in 2015.

Figure C.1.3: Example of a Maternity Leave Clause

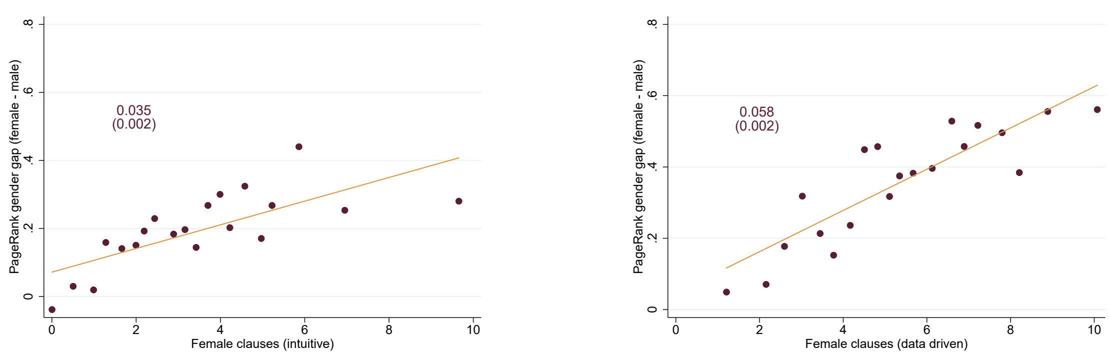


Notes: Figure shows an example of a maternity leave clause in a CBA. The clause is classified under the "Holidays and Licenses" broad group (9 in total) and the "Maternity Leave" clause types (137 in total). This particular clause extends maternity leave duration from the state-mandated 120 days to 180 days—inclusive to adopting mothers. It also extends post-maternity job protection by 6 months. The paper relies on the clause type classification of the different clauses, ignoring the variation in the text that may exist within each individual clause belonging to a specific type.

Figure C.1.4: Additional Sense Checks for Female- and Male-Centric Amenities

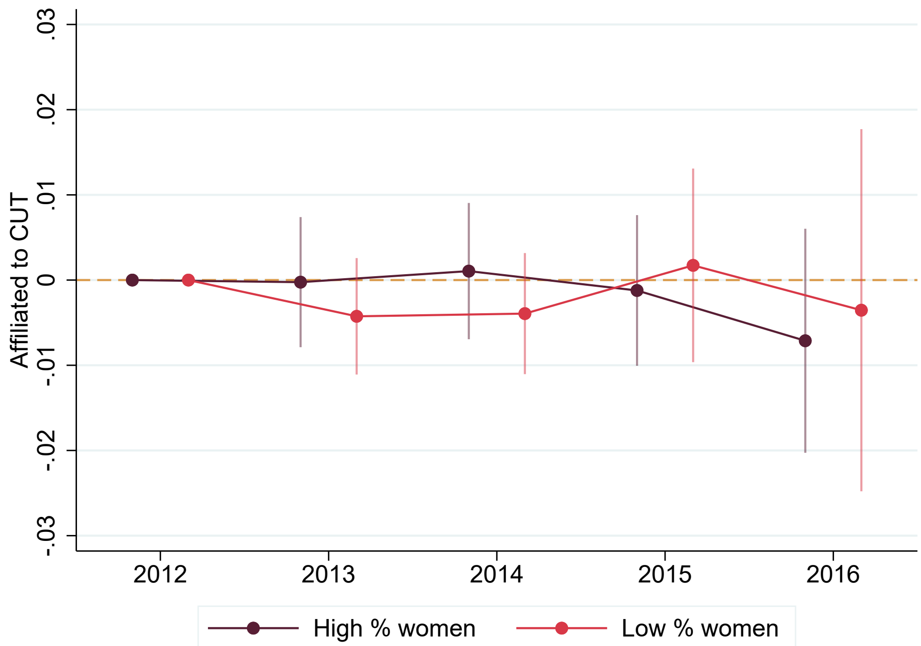
(a) Value gaps and intuitive female clauses

(b) Value gaps and data-driven female clauses



Notes: Figures depict binned scatterplots of the establishment-level gender gaps in PageRank values by the average female-centric clauses from sectoral CBAs applying to the establishment. Figure C.1.4a uses the intuitive definition of female-centric amenities, while Figure C.1.4b uses the data-driven approach. The sample used is the one used to estimate Equation (3.2), i.e., establishments in the intersection of the gender-specific super-connected sets covered by sectoral CBAs in at least 4 different years between 2009-2016.

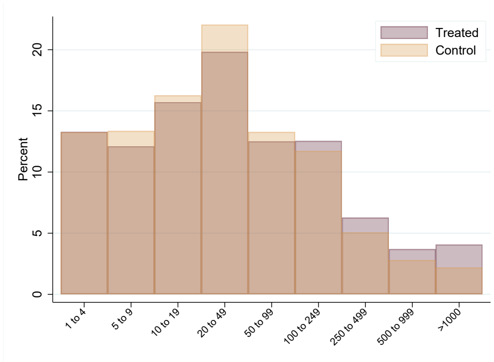
Figure C.1.5: Union Affiliation to CUT Over Time



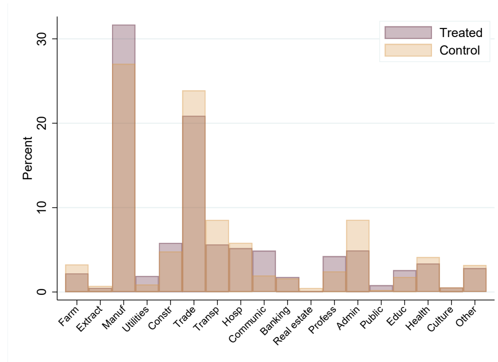
Notes: Figure plots changes in the probability of being affiliated to CUT between 2012 and 2016 separately for unions having either a high or a low share of women among the workers they represent (above or below the mean, i.e., 33% women). Coefficients represent the change with respect to 2012, in which the probability of being a CUT-affiliate is normalized to zero. Unions are weighted by the size of the workforce that they represent, computed by summing the 2012 worker count across establishments negotiating firm-level CBAs with the union. That is, if an establishment negotiates with n unions, we split the workforce count evenly to those n unions (results are robust to removing these weights). The sample is restricted to the unions in the filled panel, where only 3% of unions ever switch affiliation to or from CUT. Standard errors are clustered at the union level.

Figure C.1.6: Baseline Characteristics of Treated and Control Establishments

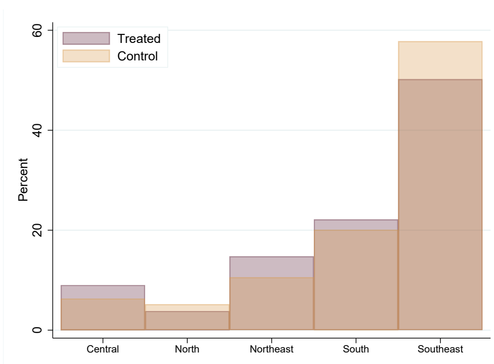
(a) Establishment size



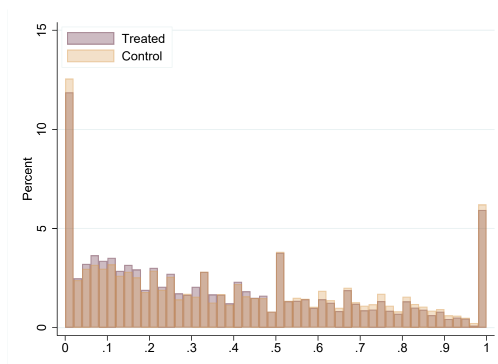
(b) Industry



(c) Region



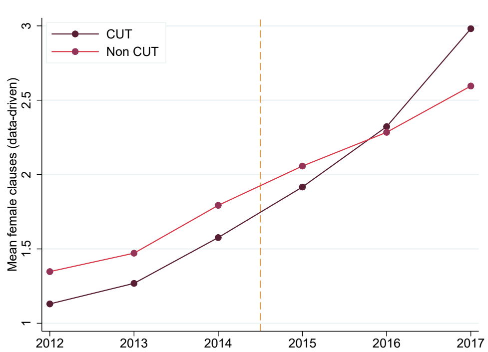
(d) Share of women employees



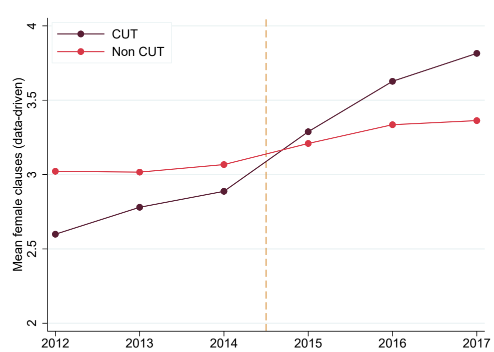
Notes: Figures show the treated and control establishments distributions of size, industry, regional location, and female share of employment at baseline. The establishments come from the starting sample detailed in Table 3.1.

Figure C.1.7: Trends in Female-Centric Clauses (Data-Driven Approach)

(a) Filled panel



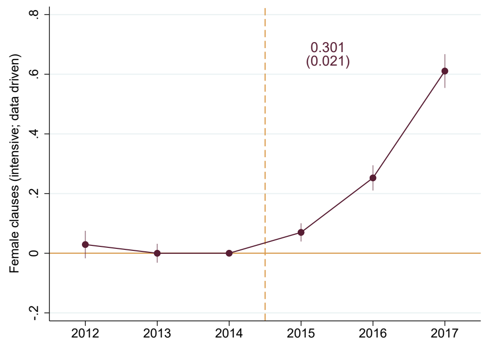
(b) New contracts



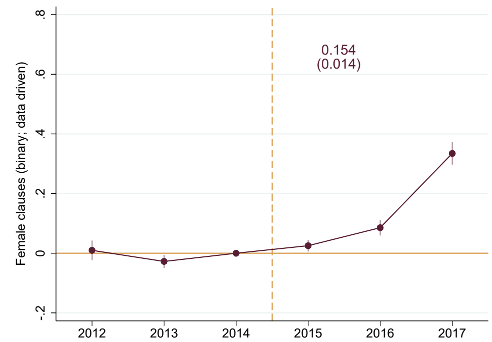
Notes: Figures plot the raw average number of female-centric clauses for treated (CUT) and control (Non CUT) establishment-union pairs over the years. Female-centric clauses are based on the data-driven classification. Figure C.1.7a plots the average number of female-centric clauses for the filled panel, while Figure C.1.7b plots the average number of female-centric clauses in newly signed contracts of the given year. Mean female clauses are lower in the filled panel and react slowly to changes in new contracts because of pairs that do not have CBA coverage in a given year.

Figure C.1.8: Effect of the CUT Reform on Female-Centric Amenities

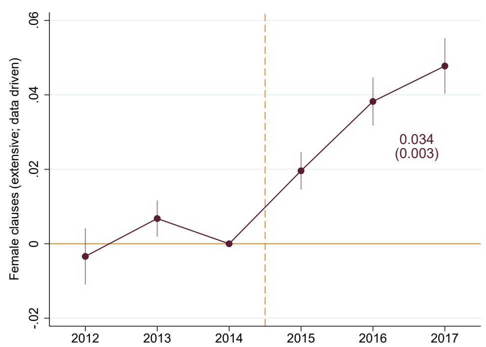
(a) Intensive margin (number)



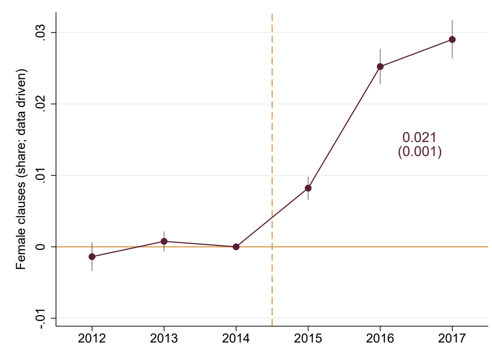
(b) Intensive margin (unique clause types)



(c) Extensive margin



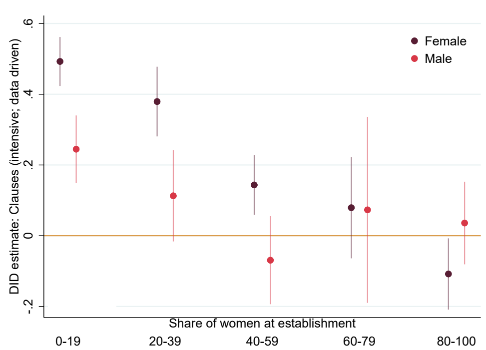
(d) As a share of clauses



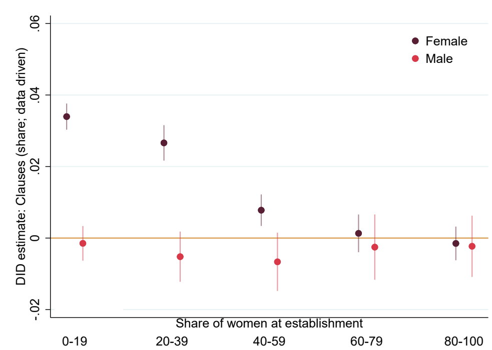
Notes: Figures show estimates of the δ_t coefficients for $t \in [2012, 2017]$ (with 2014 omitted) from the DID specification in Equation (3.3) on all margins considered for female-centric clauses, defined using the data-driven method. Confidence intervals at a 95% level are reported. Standard errors are clustered at the establishment level. All figures use the filled panel.

Figure C.1.9: Effect on Amenities by Share of Female Workers at Establishment

(a) Intensive margin



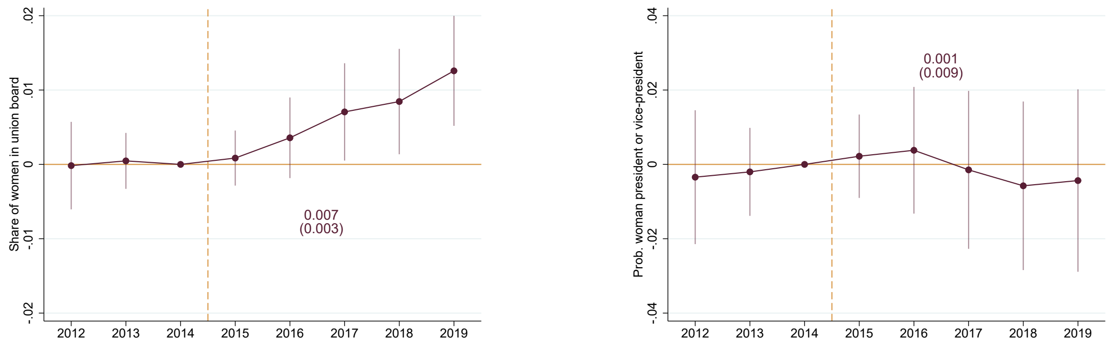
(b) As a share of all clauses



Notes: Figures show estimates of the treatment effect ($\delta_{year \geq 2015}$) from the DID specification in Equation (3.3) on the number of female- and male-centric clauses (data-driven approach) computed on subsamples of establishments divided according to the 2014 share of female workers. We use the filled panel. Confidence intervals at a 95% level are shown. Standard errors are clustered at the establishment level.

Figure C.1.10: Impact on Gender Representation in Local Union Boards

(a) Share of women in union board (b) Woman president or vice-president



Notes: Figures show estimates of the δ_t coefficients for $t \in [2012, 2019]$ (with 2014 omitted) from an event-study specification similar to the one in Equation (3.3) on measures of women representation within local union boards. The sample is restricted to unions in our analysis sample (unlike Figure 3-2b). The equation we estimate is slightly different from Equation (3.3) as the unit of observation here is the union-year so we include union fixed effects instead of establishment-union pair fixed effects. Figure C.1.10a uses the share of women in the union board as a dependent variable, while Figure C.1.10b uses a dummy indicating whether the union's president (or vicepresident) is a woman. Confidence intervals at a 95% level are reported. Standard errors are clustered by union.

of the employer(s) and union(s) signing it—and, importantly for our analysis, how many clauses it contains classified into clause types.[1]

The union identifier allows us to merge these data with data on union affiliation to union centrals coming from CNES. The employer identifier allows us to merge these data with information in RAIS, e.g., industry, microregion, and employment. We drop CBAs signed by unions with missing information about their 2012 union central affiliation (around 1.5% of contacts).[2] We additionally drop contracts signed by multiple unions with different union central affiliations: this is fewer than 0.33% of CBAs.[3]

Almost all pairs negotiate at most one contract per year: 96% of CBAs are the only agreement signed by a pair that year and 85% of pairs always negotiate at most one CBA per year during our study period. As for the remaining 15%, we take the maximum count of a given clause type across the CBAs negotiated by the pair in a given year.[4] In this way we obtain a sample of newly

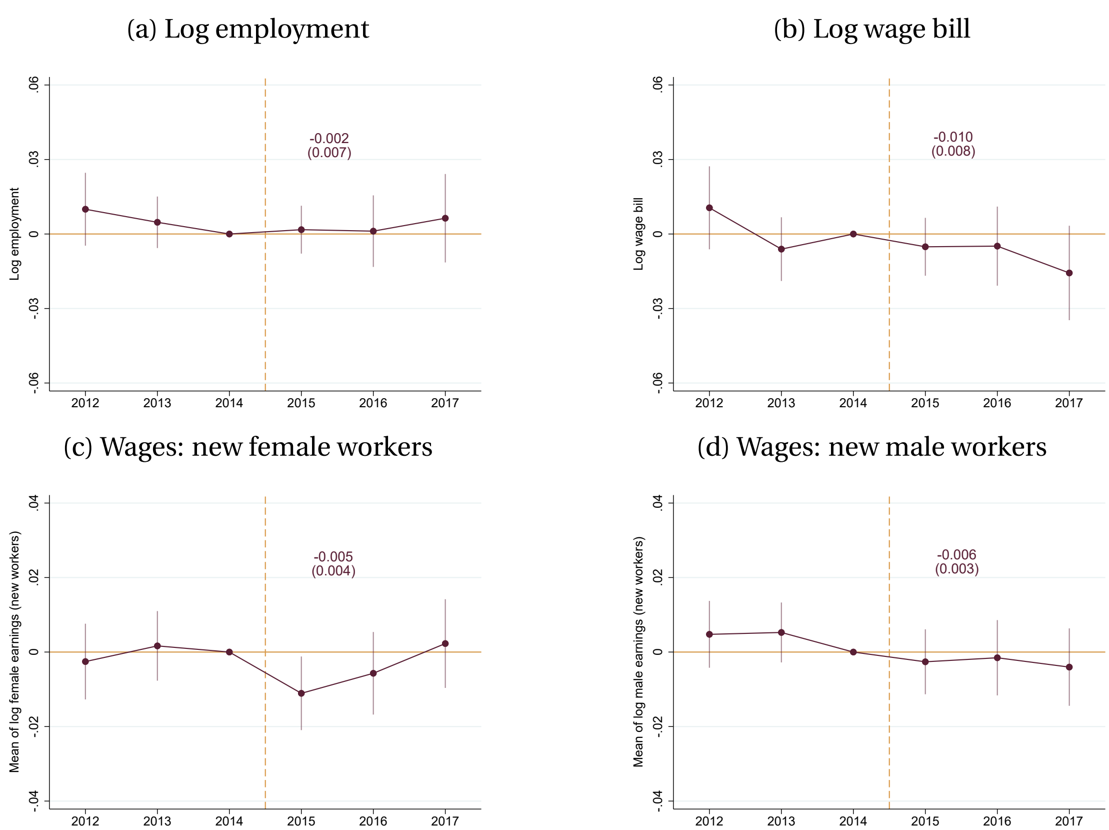
[1] *Sistema Mediador* classifies clauses into 137 categories, e.g., maternity assistance, overtime pay, life insurance, procedures in relation to strikes and strikers, etc.

[2] Unions that decide not to affiliate with any union central—which are registered in CNES as "Not-Affiliated"—are not dropped. The CBAs signed by these unions are part of the control group.

[3] Of the remaining agreements, 89.8% are negotiated between a single establishment and a single union, 7.3% are negotiated by a single union with two or more establishments, 2.5% are signed by one establishment and two or more unions with the same CUT or non-CUT affiliation, and only 0.5% by multiple unions and multiple establishments.

[4] We do this to avoid double-counting clauses as the multiple agreements per pair-year often result from misclassified CBA amendments or single-issue CBAs that are renegotiated more frequently than a year.

Figure C.1.11: Effects on Employment and Wages



(a) Log employment

(b) Log wage bill

(c) Wages: new female workers

(d) Wages: new male workers

Notes: Figures report the results of the establishment-level DID regression in Equation (3.3) with outcome variables: log of total employment, log of the wage bill, mean log wages for new female hires, and mean log wages for new male hires. Each regression includes establishment fixed effects, industry-year fixed effects, and microregion-year fixed effects. The figure plots estimates of the δ_t coefficients for $t \in [2012, 2017]$ with 2014 omitted. Confidence intervals at a 95% level are reported. Standard errors are clustered by establishment.

negotiated CBAs at the pair-year level, reporting the number of clauses for each clause type.

On the signing establishment's side, we restrict to pairs that have non-missing industry and microregion information, and that employs workers at baseline (2014). These restriction drop 8.5% of observations. This comprises the starting sample with observations at the pair-year level reported in the descriptive statistics of Table 3.1.

2. Filled panel: This sample fills in the amenities information for pairs in the *new contracts* sample for years when a new firm-level CBA was not signed. In filling the panel, we consider the institutional context regarding the automatic extension of CBAs into the future. That is, for a given pair, contracts expiring after September 25 of 2012 are automatically extended into the future until a new CBA is signed (Lagos, 2021). Although CBAs expiring before that date were not extended, we observe contracts starting 3 years prior to our study period, i.e., starting in 2009. Since the maximum duration of a CBA is 24 months, by the start of our study period (i.e., 2012) we can already be certain whether any CBA applies to a given pair-year. As such, these institutional features allow us to generate a balanced panel at the pair-year level.

To aggregate amenities at the pair-year level, for each year we only consider the contract(s) covering at least 6 months of the year.[5] If more than one contract per pair-year remains, we take the maximum count of a given clause type across CBAs—similarly to what done for the *new contracts* sample. If a pair is not covered by a firm-level CBA in a given year (even after filling the panel), we set the clause count for each clause type to zero. As such, this procedure produces a yearly balanced panel at the establishment-union pair level.

Establishment sample To study changes in the workplace, we match the contracts in our *amenities sample* to the signing establishments in RAIS. Establishments covered by contracts negotiated by unions affiliated to CUT in 2012 form our treatment group, while establishments covered by CBAs signed by unions not affiliated to CUT in 2012 make up our comparison group.

We start with the list of establishments that are part of the pairs in our *new contracts* sample. We restrict to establishments employing both men and women at baseline, dropping 15,550 establishments. We further restrict this list to establishments in the geographic coverage of their "baseline CBA", defined as the firm-level agreement closest to the 2015 CUT reform among those signed by the establishment. The reason for this restriction is that, for multiple-establishment firms, the CNPJ listed as the employer counterpart in the CBA need not be covered.[6] Restricting

[5] All other restrictions used in the *new contracts* sample apply.

[6] Firm-level CBAs apply to workers at all establishments of the signing firm that are in the geographic coverage specified in the contract. In case of multi-establishment firms, the establishment signing a CBA could be the firm headquarter but the contract might cover only subsidiaries located in other municipalities.

to signing establishments in the geographic coverage of their baseline CBA further drops 8,684 establishments, leaving us with 61,752 establishments.

For each establishment in this list we compute outcomes at the establishment-year level, such as mean log wages or total female employment, either using all job spells registered at that establishment in the year or using workers' "main job spell" in each year. We define the "main job spell" as the employment spell at which the worker worked the longest during the year. In case all job spells have the same duration, we break ties by keeping only one spell at random.

Because the same establishment can negotiate CBAs with more than one union, the final step to construct the *establishment sample* involves determining treatment status at the establishment level. We assign establishments to the treatment group as long as they are part of at least one treated pair. In practice, this decision is innocuous. Because the great majority (93.5%) of establishments always bargain with the same union, treatment assignment is trivially defined for most establishments. There are 4.4% of establishments that sign CBAs with more than one union over the time frame we consider, and all the unions they negotiate with have the same treatment status, e.g. they are all affiliated to CUT (or they are not) in 2012.The remaining 2.1% of establishments negotiate with more than one union over time and these unions have different treatment status. We conservatively assign this last group of establishment to the treatment group, which should run counter to finding effects if some of these establishments are not affected by the CUT reform.

Incumbent workers sample Incumbent workers are defined as those employed at a treated or comparison establishment as of 2014 (based on the *establishment sample*). Their treatment status depends on the treatment status of their baseline (2014) employer, as explained above in the description of the *establishment sample* construction. Leveraging the linked employer-employee feature of RAIS, incumbent workers are tracked across jobs from 2012 to 2017—that is, we are not restricting to job spells at employers in the *establishment sample*. In constructing this sample, we only consider the "main job spell" for each worker in each year.

Union and union central boards For each Brazilian union central, we construct a yearly panel with information on the gender composition of their national board between 2012 and 2019. The raw data contains the full name of all the board members, which allows us to infer their gender. We do so using the R package *genderBR*, which codes a name as female if most people with that name are women in the Brazilian census (and similarly for men).[7] We use this data to check that the introduction of the CUT gender quota had bite.

[7]Developed by Fernando Meireles and posted on GitHub.

We similarly construct a yearly panel with information on the gender composition of local union boards, the gender of their presidents and vice-presidents, and their affiliation to union centrals between 2012 and 2019. We use these data 1) to assign treatment status to unions; 2) to understand whether the reform had spillovers on local union boards; and 3) conduct heterogeneity analyses concerning women's representation in unions.

C.2.2 Construction of variables

Amenities In the analysis we adopt two different ways of classifying clauses as female-centric amenities. The first is guided by intuition to select clause types that are of plausibly of greater value to women than men (intuitive definition). The second definition is data-driven, where we use lasso to pick clauses that are most predictive of women's value of employment (relative to men) at an establishment in the cross-section. An important advantage of the data-driven approach—compared to the intuitive definition—is that it also identifies clauses that are valued relatively more by men, i.e., male-centric amenities.

We also generate four different outcome margins for clauses at the pair-year level. First, the *intensive margin (count)* measures the sum of the clause counts from the clause types categorized as either female- or male-centric in the corresponding contract. Second, the *intensive margin (sum of indicators)* measures the sum of clause type indicators for those categorized as either female- or male-centric in the corresponding contract. Third, the *extensive margin* simply indicates whether any female (or male) clause exists in the CBA of interest. Finally, we calculate the *share* of the intensive margin (count) relative to the total clause count in the CBA.

1. Intuitive definition: Guided by CUT's "fight plan" and previous work documenting the value women place on flexibility (Goldin and Katz, 2011; Mas and Pallais, 2017; Maestas et al., 2018), we identified 4 themes as female-centric: 1) leaves; 2) maternity and childcare; 3) workplace harassment and discrimination; and 4) flexibility and part-time work. From these themes we restricted ourselves to select 20 clause types. These clauses are listed in Table C.1.1—which includes clauses on maternity leave, childcare assistance, prevention of sexual harassment—all of which are conceivably of greater value to women than men.

2. Data-driven definition: The data-driven definition of amenities selects clauses that are most predictive of gender differences in the value of employment at an establishment, controlling for gender-specific wage premiums.[8] In practice, we estimate the following cross-sectional

[8] Section 3.3.2 provides a detailed justification for this approach.

specification using lasso:

$$V_j^F - V_j^M = \beta_w^F \psi_j^F - \beta_w^M \psi_j^M + \sum_{z \in Z} \beta_z a(z)_j + \epsilon_j$$

where V_j^G is the PageRank value of establishment j for workers of gender G, ψ_j^G is the establishment fixed-effect for workers of gender G at employer j from an AKM regression on wages, and $a(z)_j$ is the average clause count of amenity z (one among the 137 clause types) offered in the CBAs covering workers. We select the 20 clause types with the highest β_z and label them as "female-centric" amenities. Conversely, the 20 clause types with the lowest β_z comprise our "male-centric" amenities. Results are shown in Table 3.2.

PageRank values. To estimate PageRank values we take job spells of full-time workers, ages 18-54, on open-ended contracts, and earning monthly wages in private sector establishments from RAIS (2009-2016). For each gender, we find the largest strongly connected set of establishments based on worker flows, i.e., a link between two establishments is defined as having at least one inflow and one outflow. We restrict to establishments that have at least 10 hires overall, with at least one of these coming from non-employment. To solve for the vector of PageRank values (see Appendix C.3), we follow Morchio and Moser (2020) and only consider employment-to-employment flows to be month-to-month job transitions. In addition, we set the damping factor used in finding the fixed point in the linear system of normalized flows to 0.8—one of the standard values in computer science. That is, the "random surfer" moving through the labor market restarts his search at a new establishment with 80% probability. As shown in Sorkin (2018), PageRank values are unique up to an unknown multiplicative factor. Below we discuss robustness to assumptions about the multiplicative factor applying to women versus men to obtain $V_j^F - V_j^M$.

Wage premiums. To estimate the establishment fixed effect from AKM we take job spells of full-time workers, ages 18-54, on open-ended contracts, and earning monthly wages in private sector establishments from RAIS (2009-2016). For each gender, we find the largest strongly connected set of establishments based on worker flows, i.e., a link between two establishments is defined as having at least one inflow and one outflow. We restrict to establishments that have at least 10 workers (on average across years) and are observed at least 4 years in RAIS. Following Gerard et al. (2021), the model includes dummies for individual workers (α_i) and individual establishments (ψ_j), year dummies interacted with five education dummies, and quadratic and cubic terms in age interacted with the education dummies (X_{it})—see Appendix C.3. For the baseline year, the worker effects are measured as of age 40 to correspond to the approximate peaks of experience profiles. The establishment fixed effects for each gender—i.e., ψ_j^F and ψ_j^M—are normalized relative to the restaurant industry, where rents are assumed to be negligible.

Clause counts. To get a measure of $a(z)_j$ for each establishment, we take a yearly average of the number of clauses in each of the 137 clause groups found in sectoral CBAs negotiated between 2009 and 2016. To assign coverage from sectoral CBAs to establishments, we first need to map the signing employer association to the firms being represented. Using the equivalent of a FOIA request, we obtained the universe of establishments paying dues to employer associations. We then take sectoral CBAs and match them to all establishments paying dues to the signing employer association. The next step is to assign coverage only to establishments located in the geographic region specified in the CBA. Finally, to reduce overlap in CBA coverage, we exploit information on negotiated wage floors to assign a "main CBA" to each establishment-year.[9]

Robustness. We check the robustness of our data-driven method on two dimensions: 1) two different ways of selecting the establishment sample used in the regressions: either a 50% random split-sample (used in our baseline approach) or the full estimation sample of establishments; and 2) three definitions of the gender gap in PageRank values, i.e., $V_j^F - V_j^M$. The first definition (used in our baseline approach) chooses the establishment with the smallest wage premium gap as the normalizing establishment, and then adjusts female values relative to the male values by multiplying the former by the ratio of the female-to-male PageRank values of the normalizing establishment. The second definition simply assumes the multiplicative factor is the same for both genders, i.e., no normalization is needed. The third definition uses a (within-gender) normalized index from 0 to 100 of V_j^F and of V_j^M.

Tables C.1.4 and C.1.5 show all the clause types selected by any of the combinations above. These tables also show how many of these 6 different combinations choose a given clause type as either female- or male-centric, as well as those selected under the baseline approach but adding state and industry fixed effects.

Labor market outcomes We briefly describe how we define the outcomes used for the establishment-level and incumbent worker-level analyses. While for all worker-level outcomes we use the main job spell, some establishment-level outcomes are constructed with all job spells. We first describe establishment-level outcomes derived with all job spells and then those derived using

[9]Specifically, we first define an establishment's "core union" to be the modal union involved in negotiating wage floors that have bite on the wage distribution. Among the CBAs negotiated by the "core union" in a given year, the "main CBA" is the one with the wage floor that has the largest mass of workers.

main job spells. Finally, we describe worker-level outcomes.

Establishment level outcomes - all job spells:

- Total employment. The total number of workers employed at an establishment in a given year.
- Share of women in the workforce. Share of women employed in a given establishment-year among all workers.
- Share of women in the probationary workforce. Share of women employed in a given establishment-year with less than 3 months of tenure among all workers with fewer than 3 months of tenure. Brazil's federal labor code allows for at most 3 months of probation, after which employment terminations imply severance payments.
- New hires. Number of workers recently hired by the establishment, defined as the number of workers employed in a given establishment-year with less than 12 months of tenure.
- Share of women among new hires. Share of women employed in a given establishment-year with less than 12 months of tenure among all workers with fewer than 12 months of tenure.
- Share of women among separating workers. Share of women among workers who separate from the establishment in that year. Separating workers are defined as those who are no longer employed at the establishment by the end of the year.
- Establishment exit. A dummy variable indicating whether the establishment does not appear in RAIS in 2017.

Establishment level outcomes - main job spell:

- Mean log wage. For any given worker subgroup, we take the mean of the wage outcome (defined below) in logs across all workers in the subgroup employed at the establishment in that year. This variable is defined for the following worker subgroups: women and men with more than 12 months of tenure, women and men with less than 12 months of tenure.
- Mean gender wage gap. The difference between the mean log wage for women and the mean log wage for men for a given establishment-year.
- Wage bill. The monthly wage bill for the establishment. That is, we sum the wage outcome (defined below) for all workers employed by the establishment in that year.

- Share of women poached in. Share of new female hires that are poached from another firm among all female workers. New hires are defined as workers with less than 12 months of tenure at that establishment in a given year. Poached hires are defined as workers who in the preceding year worked at another firm in RAIS, as opposed to being unemployed or out of the (formal) workforce.
- Age of female workforce. Mean age of female workers employed at an establishment in a given year.
- Tenure of female workforce. Average months of tenure of female workers employed at an establishment in a given year.
- Hours of female workforce. Average contracted hours of work per week of female workers employed at an establishment in a given year. Weekly contracted hours are those agreed upon hiring, and do not include overtime work.
- Education of female workforce. Average years of schooling of female workers employed at an establishment in a given year.
- Share of women among managers. The share of women among workers with an occupation code corresponding to a managerial role. Occupation codes corresponding to manager positions are those starting with 12, 13 or 14 (as per CBO: *Classificação Brasileira de Ocupações*).
- Maternity leave benefits. The share of women taking maternity leaves longer than 120 days among women employed at an establishment that start their maternity leave in a given year. We are able to identify women taking maternity leave thanks to detailed information on both the length and the reason of the three longest leave spells per job spell. We think that it is very unlikely that maternity leaves are not among the three longest leave spells in a year for a woman on maternity leave. For this reason we are confident that we are observing the near universe of maternity leave spells.
- Job protection after maternity. The share of women working at the same employer where they were working at the start of maternity leave by end-of-year for the year when their maternity leave ends, among women employed at said establishment who start their maternity leave in the same year.
- Injury leave. The share of workers taking leave due to a workplace injury among all workers employed at an establishment during a given year.

Establishment level outcomes - not in RAIS:

- CBA wage adjustments. The largest percentage wage adjustment negotiated among the firm-level CBAs covering an establishment. For years without a wage adjustment clause or without a negotiated CBA, the assigned wage adjustment is zero.

- Profit margin. The mean profit margin (in percentage terms) over 2012-2014 and 2015-2017. The sample is restricted to establishments reporting profit margin information to Orbis in both the pre- and post-reform periods.

Worker level outcomes - main job spell of incumbent workers

- Wages. The average monthly earnings that a worker makes during a job spell in a given year. We always use earnings in real terms by using the December CPI (i.e., the *Índice Nacional de Preços ao Consumidor* reported by IBGE) with 2015 as the base year.

- Retention. A dummy that indicates whether the worker is observed working at the baseline employer in any given year, where the baseline employer is defined as the (main) establishment of employment in 2014.

- Employed in formal sector. A dummy that indicates whether the worker is observed working in the formal sector in that year, i.e., they have a job spell registered in RAIS in that year.

C.3 AKM and PageRank Model

Our data-driven approach to identify female- and male-centric amenities requires establishment-level estimates of gender-specific PageRank values and AKM wage premiums. This appendix presents the model underlying these estimates. For simplicity, we present the model without any reference to gender specificity. We also use establishment and firm interchangeably.

Denote $\tilde{V}_j$ as the common value of employment for any worker i at firm j. Common value means that all workers agree on $\tilde{V}_j$ such that a single job ladder exists ranking firms according to this value. All else equal, workers value higher compensation bundles so that one can write $\tilde{V}_j = h\left(w_j, a_j\right)$, where $h(\cdot)$ in strictly increasing in both the wage w_j and the amenity a_j arguments. The utility of workers from employment at the establishment, however, is heterogeneous and given by $u_{ij} = h(w_j, a_j) + \varepsilon_{ij}$, where ε_{ij} captures an individual's idiosyncratic preferences for working at j.

PageRank values

The starting point here is $u_{ij} = \tilde{V}_j + \varepsilon_{ij}$. In a market with only two firms and independently distributed type I Extreme Value ε_{ij} across workers, the probability that a worker prefers firm j over k is given by $\frac{\exp(\tilde{V}_j)}{\exp(\tilde{V}_j)+\exp(\tilde{V}_k)}$. With N workers and letting M_{jk} denote the number of workers choosing firm j over k, the following relation between employment decisions and valuations of firm-specific employment is simply $M_{kj}/M_{jk} = \exp(\tilde{V}_k)/\exp(\tilde{V}_j)$.

In a labor market with multiple firms $j \in \mathscr{J}$, the above condition imposes a restriction on each pair of firms, i.e.,

$$M_{kj}\exp(\tilde{V}_j) = M_{jk}\exp(\tilde{V}_k), \forall j \in \mathscr{J}. \tag{C.1}$$

Following Sorkin (2018), one can relax this condition by imposing a single restriction per firm that guarantees a consistent valuation of employers (e.g., no Condorcet cycles), as well as a unique set of firm-level values that best explains worker flows across establishments. Summing equation (C.1) across all employers and rearranging terms gives

$$\frac{\overbrace{\sum_{j\in\mathscr{J}} M_{kj}\exp(\tilde{V}_j)}^{\text{value-weighted entry}}}{\underbrace{\sum_{j\in\mathscr{J}} M_{jk}}_{\text{exits}}} = \underbrace{\exp(\tilde{V}_k)}_{\text{value}}, \tag{C.2}$$

which implies a single linear restriction per establishment.

The intuition behind equation (C.2) is that a valuable firm tends to be chosen over other valuable firms and has fewer workers leave it. This recursive definition of $\exp(\tilde{V}_j)$ is closely linked to Google's PageRank algorithm for ranking web-pages in a search. Along these lines, one can solve for $\exp(\tilde{V}_j)$ as a fixed point in a linear system. Moreover, a unique solution exists if the set of employers are strongly connected, i.e., an establishment has to both hire a worker from and have a worker hired by another establishment in the set.

AKM premiums

The starting point again is $u_{ij} = \tilde{V}_j + \varepsilon_{ij}$ but with the assumption that $\tilde{V}_j = \beta \log(w_j - b) + \eta \log(a_j - q)$. The parameters b and q are the workers' reference wage and amenity levels, and $\epsilon_{i,j}$ refers to the idiosyncratic preferences from working at establishment j. Assuming that the $\{\epsilon_{i,j}\}$ are independent draws from a Type I Extreme Value distribution and the number of establishments $\mathscr{J}$ is very large, workers' choice probabilities are closely approximated by exponential probabilities.[10] Hence, the establishment-specific labor supply functions are approximated by:

$$\log(L_j) = \log(\lambda) + \beta \log(w_j - b) + \eta \log(a_j - q). \tag{C.3}$$

The employer's problem is to post the wages and amenities that minimize production costs given labor supply in (C.3). The posted wages and amenities are common to all workers since employers cannot discriminate on the basis of their idiosyncratic preferences $\{\epsilon_{i,j}\}$.[11] The optimal choice is the solution to the following cost-minimization problem:

$$\min_{w,a} (w_j + \xi_j a_j) L(w_j, a_j) \quad \text{s.t. } T_j f(L(w_j, a_j)) \geq \bar{Y}, \tag{C.4}$$

where ξ_j captures heterogeneity in the marginal cost of amenity provision across employers.

The first order conditions imply that the optimal compensation package is given by:

$$w_j = T_j f'(L_j) \mu_j \left(\frac{e^L_{wj}}{1 + e^L_{wj} + e^L_{aj}} \right) \tag{C.5}$$

$$a_j = T_j f'(L_j) \mu_j \left(\frac{e^L_{aj}}{\xi_j (1 + e^L_{wj} + e^L_{aj})} \right). \tag{C.6}$$

[10] The exponential probabilities are $p_j \approx \lambda \exp(\beta \log(w_j - b) + \eta \log(a_j - q))$, where λ is a constant common across all establishments in the market.

[11] This asymmetry in information, rather than labor market concentration, is the source of monopsony power. Recall that $\mathscr{J}$ is large so as to ignore strategic interactions in posting.

Rearranging equations (C.5) and (C.6), one can write wages and amenities as weighted averages of the marginal revenue product of labor and their respective reference values, i.e.,

$$w_j = \left(\frac{\beta}{1+\beta+e^L_{aj}}\right) T_j f'(L_j)\mu_j + \left(\frac{1+e^L_{aj}}{1+\beta+e^L_{aj}}\right) b \tag{C.7}$$

$$a_j = \left(\frac{\eta}{\xi_j(1+\eta+e^L_{wj})}\right) T_j f'(L_j)\mu_j + \left(\frac{1+e^L_{wj}}{1+\eta+e^L_{wj}}\right) q. \tag{C.8}$$

Assume a linear technology $f(L_j) = \theta L_j$ and price-taking employers in the output market to specify the marginal revenue product of labor: $T_j f'(L_j)\mu_j = T_j P_j \theta$. To simplify further, assume that reference wages and amenities are proportional to productivity ($b = \bar{b}\theta$ and $q = \bar{q}\theta$). Rearranging terms and taking logs results in

$$\log(w_j) = \log\left(\frac{\theta\bar{b}(1+e^L_{aj})}{1+\beta+e^L_{aj}}\right) + \log\left(1+\beta R^w_j\right) \tag{C.9}$$

$$\log(a_j) = \log\left(\frac{\theta\bar{q}(1+e^L_{wj})}{1+\eta+e^L_{wj}}\right) + \log\left(1+\eta R^a_j\right), \tag{C.10}$$

where $R^w_j = T_j P_j / [(1+e^L_{aj})\bar{b}]$ and $R^a_j = T_j P_j / [\xi_j(1+e^L_{wj})\bar{q}]$. With relatively small values of βR^w_j and ηR^a_j, log wages and log amenities are functions of a fixed worker component and a fixed establishment component as in Abowd et al. (1999)—henceforth AKM. Specifically,

$$\log(w_j) = \log\left(\frac{\bar{b}(1+e^L_{aj})}{1+\beta+e^L_{aj}}\theta\right) + \beta R^w_j \tag{C.11}$$

$$\log(a_j) = \log\left(\frac{\bar{q}(1+e^L_{wj})}{1+\eta+e^L_{wj}}\theta\right) + \eta R^a_j. \tag{C.12}$$

In short, equations (C.11) and (C.12) imply that the wages and amenities of workers can be written in the form $\log(w_j) = \alpha^w + \psi^w_j$ and $\log(a_j) = \alpha^a + \psi^a_j$, where $\psi^w = \beta R^w_j$ is an establishment-specific wage premium and $\psi^a = \eta R^a_j$ is an establishment-specific amenity premium. To separately identify these premiums from the worker fixed effects, one must focus on a set of firms that are connected through worker flows.

C.4 Welfare Model

Following the CUT reform that increased female-centric amenities at CUT-affiliated establishments, we found that women separate from treated establishments less and queue for jobs there. These are revealed preference measures of firm value, suggesting that the reform causes treated establishments to become disproportionately valuable to women.

By how much did women's welfare increase? To answer this question we adapt an approach measuring changes in welfare from the introduction of new or improved varieties in a consumer setting to our labor market setting. We model workers as choosing firms, just like consumers choose products. Because of the reform, the quality of CUT-affiliated firms is changing, differently by gender. From a modeling perspective, this is analogous to a situation in which the quality of certain goods is improving or when new, improved, good varieties are introduced in the market. This appendix describes the model used to estimate welfare change and the estimation strategy. It also discusses how the model underlies our data-driven classification of amenities.

Model

The model assumes that workers have CES preferences over firms. One advantage of the CES demand structure is that it can be microfounded using a continuum of workers making discrete choices over where to work—as shown in Anderson et al. (1992)—and derived below. This is a common way to model the labor market (Berger et al., 2022; Card et al., 2018; Lamadon et al., 2022; Sorkin, 2018).

Worker's problem and solution A representative worker with CES preferences over $\mathcal{J}$ firms chooses the number of hours to supply to each firm to maximize total income subject to a total hours constraint:

$$\max_{\{n_j\}} \sum_{j \in \mathcal{J}} w_j n_j \quad s.t. \quad \left[\sum_j (b_j n_j)^{\frac{1+\eta}{\eta}} \right]^{\frac{\eta}{\eta+1}} = N \tag{C.13}$$

where n_j is the number of hours worked at firm j, w_j is the wage at firm j, b_j is a taste-shifter governing the disutility of working at j, and η is the (constant) elasticity of substitution across firms. The parameter b_j captures workers' valuation of firm attributes other than wages. The constraint is not hours but disutility-weighted hours. Because the representative worker solution is the same as aggregating discrete choices of a continuum of workers deciding where to work, n_j can also be seen as the measure of workers working at firm j.

Optimal labor supply to each firm is given by:

$$n_j^* = \left(\frac{w_j}{\tilde{W}}\right)^{\eta} \frac{1}{b_j^{1+\eta}} N \tag{C.14}$$

where $\tilde{W}$ is a book-keeping term called the wage index, closely related to welfare (as seen below) and defined as follows:

$$\tilde{W} = \left[\sum_{j\in\mathscr{J}} \left(\frac{w_j}{b_j}\right)^{1+\eta}\right]^{\frac{1}{1+\eta}} \tag{C.15}$$

In addition, the share of "expenditure" (i.e., labor income) at any given firm is:

$$S_j = \frac{w_j n_j}{\sum_k w_k n_k} = \frac{(\frac{w_j}{b_j})^{1+\eta}}{\sum_k (\frac{w_k}{b_k})^{1+\eta}} \tag{C.16}$$

Wage index interpretation and welfare The wage index represents how much workers are paid to work one more disutility-adjusted hour and is thus a measure of worker welfare. This can be seen by taking the envelope condition around the optimal solution to the worker's problem: $\sum_j w_j n_j = \tilde{W} N$. Formally

$$\frac{\partial}{\partial N} \sum_{j\in\mathscr{J}} w_j n_j^*(w_j, w_{-j}) = \tilde{W}$$

so that when $\tilde{W}$ rises it means workers are now paid more for providing one additional unit of disutility-weighted labor supply, thereby increasing their welfare.[12]

How welfare changes when firm attributes change When firms change attributes or amenities this changes the disutility of working there (b_{jt}). The change in welfare is measured by the ratio of the new and old wage indices:

$$\frac{\tilde{W}_t}{\tilde{W}_{t-1}} = \frac{\left[\sum_{j\in\mathscr{J}_t} \left(\frac{w_{jt}}{b_{jt}}\right)^{1+\eta}\right]^{\frac{1}{1+\eta}}}{\left[\sum_{j\in\mathscr{J}_{t-1}} \left(\frac{w_{j,t-1}}{b_{j,t-1}}\right)^{1+\eta}\right]^{\frac{1}{1+\eta}}} \tag{C.17}$$

where $\mathscr{J}_t$ are the firms observed in period t.

[12]In this way, the wage index is to welfare in the labor setting like the price index is to welfare in consumer theory. In consumer theory, the price index captures the cost of purchasing one util of utility. Welfare rises as it gets cheaper to purchase one more util.

The key challenge to estimating this change in welfare is that firm quality b_{jt} is unobserved or, in our case, is difficult to model because it would require specifying exactly how 137 clause types enter the worker's utility function. However, as first shown in (Feenstra, 1994), assuming CES demand circumvents this problem because the welfare change depends on two terms that are observed in the data: 1) the wage index of firms whose quality (b_{jt}) remains unchanged and are "common" across periods; and 2) a variety-adjustment term to account for changes at firms that do change b_{jt}. That is, the welfare change is given by

$$\phi_{t-1,t} = \left[\frac{\lambda_t}{\lambda_{t-1}}\right]^{-\frac{1}{1+\eta}} \frac{\sum_{j\in\Omega_{t,t-1}}(\frac{w_{jt}}{b_{jt}})^{1+\eta}}{\sum_{j\in\Omega_{t,t-1}}(\frac{w_{jt-1}}{b_{jt-1}})^{1+\eta}} = \left[\frac{\lambda_t}{\lambda_{t-1}}\right]^{-\frac{1}{1+\eta}} \frac{\tilde{W}_t^*}{\tilde{W}_{t-1}^*} \tag{C.18}$$

Here $\Omega_{t,t-1} = \mathscr{J}_t \cap \mathscr{J}_{t-1}$ are firms common to both periods—in our case, non-CUT firms. The asterisk * in W_t^* and W_{t-1}^* denotes that these are wage indices over the common set of firms. Finally, λ_t is the share of the wage bill at common firms in t (using wages at t).

To get an expression for $\tilde{W}_t^*/\tilde{W}_{t-1}^*$, we use Equations (C.15) and (C.16) to obtain

$$[\tilde{W}_t^*]^{1+\eta} = \frac{1}{S_{jt}^*}\left(\frac{w_{jt}}{b_{jt}}\right)^{1+\eta} \quad \forall j \in \Omega_{t,t-1} \tag{C.19}$$

Following Redding and Weinstein (2016), we take logs of both sides, difference over time, and sum over all $j \in \Omega_{t,t-1}$ to get

$$\log\left(\frac{\tilde{W}_t^*}{\tilde{W}_{t-1}^*}\right) = \log\left(\frac{\bar{w}_t^*}{\bar{w}_{t-1}^*}\right) - \frac{1}{1+\eta}\log\left(\frac{\bar{S}_t^*}{\bar{S}_{t-1}^*}\right) - \log\left(\frac{\bar{b}_t^*}{\bar{b}_{t-1}^*}\right) \tag{C.20}$$

where the bars indicate a geometric average and the last term is zero because we assume quality remains the same for these common firms. Thus, a change in welfare depends only on three terms that are observed in the data and η

$$\log\phi_{t-1,t} = -\frac{1}{1+\eta}\log\left(\frac{\lambda_t}{\lambda_{t-1}}\right) - \frac{1}{1+\eta}\log\left(\frac{\bar{S}_t^*}{\bar{S}_{t-1}^*}\right) + \log\left(\frac{\bar{w}_t^*}{\bar{w}_{t-1}^*}\right) \tag{C.21}$$

Microfoundation of CES demand using discrete choices

Following the CES demand in (Berger et al., 2022), workers' utility for working at a firm has a component that is common across workers, encompassing wages and a common taste for the firm amenities, and an idiosyncratic shock that follows a logit distribution. Firms post utility offers—we don't model the source of firm heterogeneity and assume that they exogenously differ. There is a unit measure of workers indexed by $i \in [0,1]$. Each worker has a disutility for working at firm j:

$$v_{ij} = \exp^{-\xi_{ij}} h_{ij} b_j$$

with ξ_{ij} iid across workers and drawn from a multivariate Gumbel distribution with parameter η. Each worker must earn $y \sim F(y)$, where earnings $y_i = w_j h_{ij}$. The worker chooses firm j to minimize disutility:

$$\min_j \{\log h_{ij} + \log b_j - \xi_{ij}\} = \max_j \{\log w_j - \log y_i - \log b_j + \xi_{ij}\}$$

Following McFadden (1973) on logit, the probability that worker i chooses to work at firm j is:

$$p_i(\tilde{w}) = \frac{\tilde{w}_j^{1+\eta}}{\sum_k \tilde{w}_k{}^{1+\eta}}$$

where $\tilde{w}_j := \frac{w_j}{b_j}$. The aggregate labor supply to firm j is then found by integrating the probability that a worker works at that firm times the hours supplied by that worker, over the mass of all workers:

$$\begin{aligned}
n_j &= \int p_i(\tilde{w}) \cdot h_{ij} \cdot dF(y) \quad \text{where} \quad h_{ij} = y_i / w_j \\
n_j &= \frac{\tilde{w}_j^{1+\eta}}{\sum_k \tilde{w}_k{}^{1+\eta}} \frac{1}{w_j} \int y_i dF(y) \\
&= \left(\frac{w_j}{\tilde{W}}\right)^{\eta} \frac{1}{b_j^{1+\eta}} N
\end{aligned}$$

This is exactly the aggregate labor supply to firm j as in the representative worker's problem with CES demand. The last line follows from the fact that in equilibrium:

$$Y = \int y_i dF(y) = \sum_{j \in \mathscr{J}} w_j n_j^* = \tilde{W} N$$

Estimation

To get at welfare changes by gender, we estimate equation (C.21) separately for men and women. Starting from the same establishment-year panel that we use to study labor market outcomes, we compute the average earnings and total employment for each group of workers employed at an establishment during two periods: the pre-reform period (2012-2014), denoted by $t-1$, and the post-reform period (2015-2017), denoted by t. To do that, we take averages of establishment level quantities across years.

We separately estimate each one of the terms in the right hand side of equation (C.21), that is, $\log\left(\frac{\lambda_t}{\lambda_{t-1}}\right)$, $\log\left(\frac{\bar{w}_t^*}{\bar{w}_{t-1}^*}\right)$ and $\log\left(\frac{\bar{S}_t^*}{\bar{S}_{t-1}^*}\right)$ and we combine them with an estimate of η that we calibrate from Felix (2022).

The ideal experiment to estimate the welfare change due to the CUT reform would be to randomly shock some labor markets with the reform while leaving other markets unaffected. Lacking this ideal setting, we estimate the welfare components from pre-post comparisons within establishments. As any pre-post strategy, we recognize that it might also pick up the effect of other things changing during the period under study that might affect wages or employment within establishments over time.

Changes in $\bar{w}_t^*$ and in $\bar{S}_t^*$ can be directly estimated through an establishment-level regression. Note that the difference in the log of the geometric mean of a variable x is equivalent to the average change in $\log(x)$ between t and $t-1$ across units. In our case

$$\log\left(\frac{\bar{w}_t^*}{\bar{w}_{t-1}^*}\right) = \log \bar{w}_t^* - \log \bar{w}_{t-1}^* = \frac{1}{N_\Omega}\left(\sum_{j\in\Omega_{t,t-1}} \log w_{jt} - \sum_{j\in\Omega_{t,t-1}} \log w_{jt-1}\right) = \mathbb{E}[\Delta \log w_{jt} | j \in \Omega_{t,t-1}]$$

where N_Ω denotes the number of firms in $\Omega_{t,t-1}$, that is, the number of comparison (non-CUT affiliated) firms. We estimate the component of welfare due to changes in $\bar{w}_t^*$ as the average change in log wages across non-CUT establishments, captured by the coefficient β in the following regression:

$$\log w_{jt} = \alpha + \beta Post_t + \mu_j + \epsilon_{jt}, \quad j \in \Omega_{t,t-1}$$

where μ_j denotes establishment fixed effects and $Post_t$ is a dummy for the post-reform period (2015-2017). We estimate $\log\left(\frac{\bar{S}_t^*}{\bar{S}_{t-1}^*}\right)$ with a similar regression, using $\log(s_{jt})$ as dependent variable, where $s_{jt} = \frac{w_{jt}n_{jt}}{\sum_{k\in\Omega_{t,t-1}} w_{kt}n_{kt}}$.

To estimate the change in λ, we instead take a first order approximation around λ_{t-1}

$$\begin{aligned}
\Delta\lambda_t = \lambda_t - \lambda_{t-1} &= \sum_{j\in\mathscr{J}} \frac{\partial}{\partial w_j}\lambda \cdot dw_j + \sum_{j\in\mathscr{J}} \frac{\partial}{\partial n_i}\lambda \cdot dn_j \Big|_{w_{t-1},n_{t-1}} \\
&= \frac{\sum_{j\in(\mathscr{J}\setminus\Omega)} w_{jt-1}n_{jt-1}}{(\sum_{j\in\mathscr{J}} w_{jt-1}n_{jt-1})^2}\left(\sum_{j\in\Omega} n_{jt-1}\cdot dw_j + \sum_{j\in\Omega} w_{jt-1}\cdot dn_j\right) \\
&\quad - \frac{\sum_{j\in\Omega} w_{jt-1}n_{jt-1}}{(\sum_{j\in\mathscr{J}} w_{jt-1}n_{jt-1})^2}\cdot\left(\sum_{i\in\mathscr{J}\setminus\Omega} n_{jt-1}\cdot dw_j + \sum_{j\in\mathscr{J}\setminus\Omega} w_{jt-1}\cdot dn_j\right)
\end{aligned}$$

where to simplify notation we use Ω in place of $\Omega_{t,t-1}$ to denote the set of non-CUT firms (of measure N_Ω) and $\mathscr{J}\setminus\Omega$ to denote the set of CUT-affiliated firms (of measure $N_{\mathscr{J}\setminus\Omega}$).

We define $\tilde{s}_{jt} = \frac{w_{jt-1}n_{jt}}{\sum_{k\in\mathscr{J}} w_{kt-1}n_{kt-1}}$ and $\hat{s}_{jt} = \frac{w_{jt}n_{jt-1}}{\sum_{k\in\mathscr{J}} w_{kt-1}n_{kt-1}}$ and re-write the expression above as

$$\begin{aligned}\Delta\lambda_t &= N_\Omega(1-\lambda_{t-1})\left(\mathbb{E}[\Delta\tilde{s}_{jt}|j\in\Omega] + \mathbb{E}[\Delta\hat{s}_{jt}|j\in\Omega]\right)\\ &\quad - N_{\mathscr{J}\setminus\Omega}\lambda_{t-1}\left(\mathbb{E}[\Delta\tilde{s}_{jt}|j\in\mathscr{J}\setminus\Omega] + \mathbb{E}[\Delta\hat{s}_{jt}|j\in\mathscr{J}\setminus\Omega]\right)\end{aligned}$$

where $\mathbb{E}[.]$ denotes an average across firms. Finally, because $\log\left(\frac{\lambda_t}{\lambda_{t-1}}\right) = \log\left(\frac{\Delta\lambda_t}{\lambda_{t-1}}+1\right) \approx \frac{\Delta\lambda_t}{\lambda_{t-1}}$, we can write:

$$\begin{aligned}&\log\left(\frac{\lambda_t}{\lambda_{t-1}}\right) \approx\\ &\quad N_\Omega\frac{(1-\lambda_{t-1})}{\lambda_{t-1}}\left(\mathbb{E}[\Delta\tilde{s}_{jt}|j\in\Omega]\mathbb{E}[\Delta\hat{s}_{jt}|j\in\Omega]\right) - N_{\mathscr{J}\setminus\Omega}\left(\mathbb{E}[\Delta\tilde{s}_{jt}|j\in\mathscr{J}\setminus\Omega] + \mathbb{E}[\Delta\hat{s}_{jt}|j\in\mathscr{J}\setminus\Omega]\right)\end{aligned}$$

.

We estimate the average change in $\hat{s}_{jt}$ and $\tilde{s}_{jt}$ across establishments with a within-establishment pre-post comparison. That is, we run four regressions of the form

$$y_{jt} = \alpha + \beta Post_t + \mu_j + \epsilon_{jt}$$

using as dependent variable $\hat{s}_{jt}$ and $\tilde{s}_{jt}$—separately for CUT and non-CUT firms—and we combine these estimates with λ_{t-1}, N_Ω, and $N_{I\setminus\Omega}$ which are directly computed from the data.

To obtain standard errors around total welfare and each one of the three welfare components, we bootstrap the entire estimation exercise 1000 times. In each bootstrap we re-draw with replacement a new sample of establishments from our initial sample and re-run the establishment-level regressions on the new sample.

Bibliography

Abdulkadiroglu, A., Angrist, J., Narita, Y., and Pathak, P. A. (2022). Breaking Ties: Regression Discontinuity Design Meets Market Design. *Econometrica*, 90(1):117–151. 86, 106, 107, 206

Abdulkadiroğlu, A., Agarwal, N., and Pathak, P. A. (2017a). The welfare effects of coordinated assignment: Evidence from the new york city high school match. *American Economic Review*, 107(12):3635–89. 21, 52, 89

Abdulkadiroğlu, A., Angrist, J. D., Narita, Y., and Pathak, P. A. (2017b). Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation. *Econometrica*, 85(5):1373–1432. 86

Abdulkadiroğlu, A., Pathak, P. A., and Roth, A. E. (2005). The new york city high school match. *American Economic Review*, 95(2):364–367. 23

Abdulkadiroğlu, A., Pathak, P. A., and Roth, A. E. (2009). Strategy-proofness versus efficiency in matching with indifferences: Redesigning the nyc high school match. *The American Economic Review*, 99(5):1954–1978. 23

Abdulkadiroğlu, A., Pathak, P. A., Schellenberg, J., and Walters, C. R. (2020). Do Parents Value School Effectiveness? *American Economic Review*, 110(5):1502–1539. 21, 22, 32, 46

Abdulkadiroğlu, A., Pathak, P. A., and Walters, C. R. (2018). Free to choose: Can school choice reduce student achievement? *American Economic Journal: Applied Economics*, 10(1):175–206. 22

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High Wage Workers and High Wage Firms. *Econometrica*, 67(2):251–333. 143, 309

Agarwal, N. and Somaini, P. (2018). Demand Analysis Using Strategic Reports: An Application to a School Choice Mechanism. *Econometrica*, 86(2):391–444. 52

Agarwal, N. and Somaini, P. (2020). Revealed Preference Analysis of School Choice Models. *Annual Review of Economics*, 12:471–501. 45

Ainsworth, R., Dehejia, R., Pop-Eleches, C., and Urquiola, M. (2023). Why do households leave school value added on the table? the roles of information and preferences. *American Economic Review*, 113(4):1049–82. 22, 32, 46, 86, 91

Allende, C. (2020). Competition under social interactions and the design of information policies. Technical report. 21

Allende, C., Gallego, F., and Neilson, C. (2019). Approximating the equilibrium effects of information interventions. Technical report. 22, 55, 86, 89, 91

Amin, R. (2019). The iconic nyc high school directory is going largely virtual. will that improve the daunting process or disadvantage some families? 24

Anderson, S. P., De Palma, A., and Thisse, J.-F. (1992). *Discrete Choice Theory of Product Differentiation*. MIT press. 158, 310

Andrabi, T., Das, J., and Khwaja, A. I. (2017). Report cards: The impact of providing school and child test scores on educational markets. *American Economic Review*, 107(6):1535–63. 22, 55, 91

Angrist, J. (2013). The perils of peer effects. Working Paper 19774. 111

Angrist, J., Gray-Lobe, G., Idoux, C. M., and Pathak, P. A. (2022a). Still Worth the Trip? School Busing Effects in Boston and New York. Working Paper 30308. Series: Working Paper Series. 84, 86, 105, 106

Angrist, J., Hull, P., Pathak, P., and Walters, C. (2016). Interpreting tests of school vam validity. *American Economic Review*, 106(5):388–92. 27, 205, 206

Angrist, J., Hull, P., Pathak, P. A., and Walters, C. (2021). Credible school value-added with undersubscribed school lotteries. *The Review of Economics and Statistics*, pages 1–46. 27, 89, 93, 121, 205, 208

Angrist, J., Hull, P., and Walters, C. R. (2022b). Methods for measuring school effectiveness. Working Paper 30803. 27, 205

Angrist, J. D., Hull, P. D., Pathak, P. A., and Walters, C. R. (2017). Leveraging lotteries for school value-added: Testing and estimation. *The Quarterly Journal of Economics*, 132(2):871–919. 206

Arteaga, F., Kapor, A., Neilson, C., and Zimmerman, S. (2021). Smart Matching Platforms and Heterogeneous Beliefs in Centralized School Choice. Technical Report w28946, Cambridge, MA. 86, 92

Ashenfelter, O. (1972). Racial Discrimination and Trade Unionism. *Journal of Political Economy*, 80(3, Part 1):435–464. 135

Atkin, D., Chaudhry, A., Chaudhry, S., Khandelwal, A. K., and Verhoogen, E. (2015). Markup and Cost Dispersion across Firms: Direct Evidence from Producer Surveys in Pakistan. *American Economic Review*, 105(5):537–44. 158

Barahona, N., Dobbin, C., and Otero, S. (2023a). Affirmative action in centralized college admission systems. Technical report. 23, 52

Barahona, N., Otero, C., and Otero, S. (2023b). Equilibrium effects of food labeling policies. *Econometrica*, 91(3):839–868. 46

Barseghyan, L., Clark, D., and Coate, S. (2019). Peer preferences, school competition, and the effects of public school choice. *American Economic Journal: Economic Policy*, 11(4):124–58. 17, 21

Berger, D., Herkenhoff, K., and Mongey, S. (2022). Labor market power. *American Economic Review*, 112(4):1147–93. 134, 158, 161, 310, 312

Bergman, P. (2020). Nudging Technology Use: Descriptive and Experimental Evidence from School Information Systems. *Education Finance and Policy*, 15(4):623–647. 37

Bergman, P., Chan, E. W., and Kapor, A. (2020). Housing search frictions: Evidence from detailed search data and a field experiment. Working Paper 27209. 22, 91

Bertrand, M., Black, S. E., Jensen, S., and Lleras-Muney, A. (2018). Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway. *The Review of Economic Studies*, 86(1):191–239. 135, 162

Beuermann, D. W. and Jackson, C. (2018). Do parents know best? the short and long-run effects of attending the schools that parents prefer. IDB Working Paper Series IDB-WP-914, Washington, DC. 21

Bhargava, S. and Manoli, D. (2015). Psychological frictions and the incomplete take-up of social benefits: Evidence from an irs field experiment. *American Economic Review*, 105(11):3489–3529. 37

Biasi, B. and Sarsons, H. (2022). Flexible Wages, Bargaining, and the Gender Gap. *The Quarterly Journal of Economics*, 137(1):215–266. 135

Bjerre-Nielsen, A. and Gandil, M. H. (2020). Attendance Boundary Policies and the Limits to Combating School Segregation. 87

Black, S. E., Denning, J. T., and Rothstein, J. (2023). Winners and losers? the effect of gaining and losing access to selective colleges on education and labor market outcomes. *American Economic Journal: Applied Economics*, 15(1):26–67. 23

Blandhol, C., Mogstad, M., Nilsson, P., and Vestad, O. L. (2020). Do Employees Benefit from Worker Representation on Corporate Boards? *NBER Working Paper No. w28269*. 136

Blau, F. D. and Kahn, L. M. (2006). The U.S. Gender Pay Gap in the 1990S: Slowing Convergence. *ILR Review*, 60(1):45–66. 162

Blau, F. D. and Kahn, L. M. (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature*, 55(3):789–865. 162

Bleemer, Z. (2021). Affirmative Action, Mismatch, and Economic Mobility after California's Proposition 209*. *The Quarterly Journal of Economics*, 137(1):115–160. 23

Boisjoly, J., Duncan, G. J., Kremer, M., Levy, D. M., and Eccles, J. (2006). Empathy or Antipathy? The Impact of Diversity. *American Economic Review*, 96(5):1890–1905. 86, 111

Bolotnyy, V. and Emanuel, N. (2022). Why do women earn less than men? evidence from bus and train operators. *Journal of Labor Economics*, 40(2):283–323. 135

Boudreau, L. (2023). Multinational enforcement of labor law: Experimental evidence from bangladesh's apparel sector. *Mimeo*. 136

Cai, J., De Janvry, A., and Sadoulet, E. (2015). Social networks and the decision to insure. *American Economic Journal: Applied Economics*, 7(2):81–108. 83

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, 82(6):2295–2326. 107

Calsamiglia, C., Fu, C., and Guell, M. (2020). Structural Estimation of a Model of School Choices: The Boston Mechanism versus Its Alternatives. *Journal of Political Economy*, 128(2):642–680. 52

Campos, C. (2023a). The impact of public school choice: Evidence from los angeles' zones of choice. Technical report. 22

Campos, C. (2023b). Social interactions and preferences for schools:experimental evidence from los angeles. Technical report. 22, 83, 91

Card, D. (1996). The Effect of Unions on the Structure of Wages: A Longitudinal Analysis. *Econometrica*, pages 957–979. 135

Card, D., Cardoso, A. R., Heining, J., and Kline, P. (2018). Firms and Labor Market Inequality: Evidence and Some Theory. *Journal of Labor Economics*, 36(S1):S13–S70. 134, 158, 310

Card, D., Lemieux, T., and Riddell, W. C. (2004). Unions and Wage Inequality. *Journal of Labor Research*, 25(4):519–559. 135

Card, D., Lemieux, T., and Riddell, W. C. (2020). Unions and Wage Inequality: The Roles of Gender, Skill and Public Sector Employment. *Canadian Journal of Economics*, 53(1):140–173. 135

Carlana, M., La Ferrara, E., and Pinotti, P. (2022). Goals and gaps: Educational careers of immigrant children. *Econometrica*, 90(1):1–29. 18, 84, 89

Carrell, S. E., Hoekstra, M., and West, J. E. (2019). The Impact of College Diversity on Behavior toward Minorities. *American Economic Journal: Economic Policy*, 11(4):159–182. 85, 86, 111

Chade, H. and Smith, L. (2006). Simultaneous Search. *Econometrica*, 74(5):1293–1307. 90

Chattopadhyay, R. and Duflo, E. (2004). Women as Policy Makers: Evidence from a Randomized Policy Experiment in India. *Econometrica*, 72(5):1409–1443. 135, 157

Chernew, M., Gowrisankaran, G., and Scanlon, D. P. (2008). Learning and the value of information: Evidence from health plan report cards. *Journal of Econometrics*, 144(1):156–174. 46

Chetty, R., Deming, D. J., and Friedman, J. N. (2023). Diversifying Society's Leaders? The Causal Effects of Admission to Highly Selective Private Colleges. Working Paper 31492. Series: Working Paper Series. 84

Chetty, R., Friedman, J. N., Saez, E., Turner, N., and Yagan, D. (2020). Income AScergorsesgCatoilolengaesndinInthteergUennietreadtiSotnaatlesM obility. *Quaterly Journal of Economics*. 84

Chetty, R. and Hendren, N. (2018). The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects*. *The Quarterly Journal of Economics*, 133(3):1107–1162. 83

Chetty, R., Hendren, N., and Katz, L. F. (2016a). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, 106(4):855–902. 83

Chetty, R., Hendren, N., Lin, F., Majerovitz, J., and Scuderi, B. (2016b). Childhood environment and gender gaps in adulthood. *American Economic Review*, 106(5):282–88. 83

Cohen, D. (2021). NYC School Segregation Report Card: Still Last, Action Needed Now. 21, 56, 84

Cohodes, S., Corcoran, S., Jennings, J., and Sattin-Bajaj, C. (2022). When do informational interventions work? experimental evidence from new york city high school choice. Working Paper 29690. 22, 31, 55, 61

Conley, T. G. and Udry, C. R. (2010). Learning about a new technology: Pineapple in ghana. *American Economic Review*, 100(1):35–69. 83

Corcoran, S. P., Jennings, J. L., Cohodes, S. R., and Sattin-Bajaj, C. (2022). Leveling the playing field for high school choice: results from a field experiment of informational interventions. Working paper. 22, 37, 55

Corno, L., Ferrara, E. L., and Burns, J. (2019). Interaction, stereotypes and performance. Evidence from South Africa. IFS Working Papers W19/03. 86, 92, 111

Corradini, V. and Idoux, C. (2023). Untangling the paradox of choice: Why minority families choose lower-quality schools (and what can be done about it). Technical report. 18, 22, 23, 24, 26, 40, 46, 219

Cullen, J. B., Jacob, B. A., and Levitt, S. (2006). The effect of school choice on participants: Evidence from randomized lotteries. *Econometrica*, 74(5):1191–1230. 17, 21

DiNardo, J., Fortin, N., and Lemieux, T. (1996). Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica*, 64(5):1001–44. 135

Dittrich, M., Knabe, A., and Leipold, K. (2014). Gender differences in experimental wage negotiations. *Economic Inquiry*, 52(2):862–873. 135

Dranove, D. and Sfekas, A. (2008). Start spreading the news: A structural estimate of the effects of new york hospital report cards. *Journal of Health Economics*, 27(5):1201–1207. 46

Dur, U., Kominers, S. D., Pathak, P. A., and Sönmez, T. (2018). Reserve design: Unintended consequences and the demise of boston's walk zones. *Journal of Political Economy*, 126(6):2457–2479. 23

Ellison, G. and Pathak, P. A. (2021). The efficiency of race-neutral alternatives to race-based affirmative action: Evidence from chicago's exam schools. *American Economic Review*, 111(3):943–75. 23

Epple, D. and Romano, R. E. (2011). Chapter 20 - peer effects in education: A survey of the theory and evidence. volume 1 of *Handbook of Social Economics*, pages 1053–1163. North-Holland. 83

Fack, G., Grenet, J., and He, Y. (2019). Beyond Truth-Telling: Preference Estimation with Centralized School Choice. page 79. 52

Farber, H. S. (1986). The Analysis of Union Behavior. *Handbook of Labor Economics*, 2:1039–1089. 135, 162

Farber, H. S., Herbst, D., Kuziemko, I., and Naidu, S. (2021). Unions and Inequality over the Twentieth Century: New Evidence from Survey Data*. *The Quarterly Journal of Economics*, 136(3):1325–1385. 135

Feenstra, R. C. (1994). New Product Varieties and the Measurement of International Prices. *American Economic Review*, pages 157–177. 135, 158, 312

Felix, M. (2022). Trade, labor market concentration, and wages. 160, 314

Flabbi, L., Macis, M., Moro, A., and Schivardi, F. (2019). Do Female Executives Make a Difference? The Impact of Female Leadership on Gender Gaps and Firm Performance. *Economic Journal*, 129(622):2390–2423. 135

Friedman, M. (1955). The role of government in education. In *Economics and the Public Interest.* 17

Gerard, F., Lagos, L., Severnini, E., and Card, D. (2021). Assortative Matching or Exclusionary Hiring? The Impact of Employment and Pay Policies on Racial Wage Differences in Brazil. *American Economic Review*, 111(10):3418–57. 302

Godinho Delgado, D. (2017). Trade Unions in Transformation. Building Trade Union Power with Gender Equality: The Case of the Unified Workers' Central of Brazil. Technical report. 139

Goldin, C. (2014). A Grand Gender Convergence: Its Last Chapter. *American Economic Review*, 104(4):1091–1119. 131, 162

Goldin, C. and Katz, L. F. (2011). The Cost of Workplace Flexibility for High-Powered Professionals. *The Annals of the American Academy of Political and Social Science*, 638(1):45–67. 142, 301

Golub, B. and Sadler, E. (2016). 504Learning in Social Networks. In *The Oxford Handbook of the Economics of Networks*. Oxford University Press. 83

Goolsbee, A. and Petrin, A. (2004). The consumer gains from direct broadcast satellites and the competition with cable tv. *Econometrica*, 72(2):351–381. 48

Gruber, J. (1994). The Incidence of Mandated Maternity Benefits. *American Economic Review*, 84(3):622–641. 136

Hahm, D. W. and Park, M. (2022). A dynamic framework of school choice:effects of middle schools on high school choice. Technical report. 87

Hailey, C. A. (2022). Racial Preferences for Schools: Evidence from an Experiment with White, Black, Latinx, and Asian Parents and Students. *Sociology of Education*, 95(2):110–132. _eprint: https://doi.org/10.1177/00380407211065179. 46, 91

Harju, J., Jäger, S., and Schoefer, B. (2021). Voice at work. *Working Paper No. 28522*. 136

Hastings, J., Neilson, C. A., and Zimmerman, S. D. (2015). The effects of earnings disclosure on college enrollment decisions. Working Paper 21300. 22

Hastings, J. S., Kane, T. J., and Staiger, D. O. (2009). Heterogeneous preferences and the efficacy of public school choice. Working paper. 18, 21

Hastings, J. S. and Weinstein, J. M. (2008). Information, School Choice, and Academic Achievement: Evidence from Two Experiments*. *The Quarterly Journal of Economics*, 123(4):1373–1414. 22, 55, 86

Hill, H. (1996). The Problem of Race in American Labor History. *Reviews in American History*, 24(2):189–208. 135

Holzer, H. J., Katz, L. F., and Krueger, A. B. (1991). Job Queues and Wages. *The Quarterly Journal of Economics*, 106(3):739–768. 133, 154

Horn, C. H. (2009). *Collective Bargaining in Brazil: A Study on Joint Regulation of the Employment Relationship in Manufacturing*. VDM Verlag. 137, 146

Hoxby, C. and Turner, S. (2013). Expanding College Opportunities for High-Achieving, Low Income Students. *Education Next,*, 13(4):66–73. 86

Hoxby, C. M. (2000). Does competition among public schools benefit students and taxpayers? *American Economic Review*, 90(5):1209–1238. 17, 21

Hoxby, C. M. (2003). School choice and school productivity (or could school choice be a tide that lifts all boats?). In *The Economics of School Choice*, number 8873 in Working Paper Series, pages 287–342. University of Chicago Press. 17, 21

Hoxby, C. M. and Avery, C. (2012). The missing : The hidden supply of high-achieving, low income students. Working Paper 18586. 18

Hoxby, C. M. and Turner, S. (2015). What high-achieving low-income students know about college. *American Economic Review*, 105(5):514–17. 84, 86

Idoux, C. (2021). Integrating new york city schools: The role of admission criteria and family preferences. 23, 52, 56, 57, 84, 87, 89, 102

ILO and Gallup Inc. (2017). Towards a Better Future for Women and Work: Voices of Women and Men. Technical report. 131

Jäger, S., Schoefer, B., and Heining, J. (2021). Labor in the Boardroom. *The Quarterly Journal of Economics*, 136(2):669–725. 150

Kamenica, E. (2019). Bayesian persuasion and information design. *Annual Review of Economics*, 11(1):249–272. 23

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6):2590–2615. 58

Kapor, A. (2020). Distributional effects of race-blind affirmative action. Technical report. 23, 52

Kapor, A. J., Neilson, C. A., and Zimmerman, S. D. (2020). Heterogeneous beliefs and school choice mechanisms. *American Economic Review*, 110(5):1274–1315. 22, 52, 86, 92

Kleven, H., Landais, C., and Sogaard, J. E. (2019). Children and Gender Inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, 11(4):181–209. 131

Krueger, A. and Summers, L. (1988). Efficiency Wages and the Inter-industry Wage Structure. *Econometrica*, 56(2):259–93. 133, 136, 153

Ladd, H. (2002). Market-based reforms in urban education. Technical report. 17, 21

Lagos, L. (2021). Labor Market Institutions and the Composition of Firm Compensation: Evidence from Brazilian Collective Bargaining. *Working Paper*. 136, 144, 147, 155, 299

Lamadon, T., Mogstad, M., and Setzler, B. (2022). Imperfect Competition, Compensating Differentials, and Rent Sharing in the US Labor Market. *American Economic Review*, 112(1):169–212. 136, 158, 310

Lang, K. and Gagnon, M.-J. (2009). Brazilian Trade Unions: In (In)Voluntary Confinement of the Corporatist Past. *Relations Industrielles / Industrial Relations*, 64(2):250–269. 139

Larroucau, T. and Rios, I. (2020). Do "Short-List" Students Report Truthfully? Strategic Behavior in the Chilean College Admissions Problem. page 65. 52

Laverde, M. (2020). Unequal Assignments to Public Schools and the Limits of School Choice. *job market paper*, page 48. 18, 84, 87, 89

Leibbrandt, A. and List, J. A. (2015). Do Women Avoid Salary Negotiations? Evidence from a Large-Scale Natural Field Experiment. *Management Science*, 61(9):2016–2024. 135

Liukkunen, U. (2019). *Collective Bargaining in Labour Law Regimes: A Global Perpective*, chapter The Role of Collective Bargaining in Labour Law Regimes: A Global Approach, pages 1–64. Springer Nature. 138, 146

Lowe, M. (2021). Types of contact: A field experiment on collaborative and adversarial caste integration. *American Economic Review*, 111(6):1807–44. 85, 111

Maestas, N., Mullen, K. J., Powell, D., von Wachter, T., and Wenger, J. B. (2018). The value of working conditions in the united states and implications for the structure of wages. *Working Paper No. 25204*. 136, 142, 301

Maida, A. and Weber, A. (2020). Female Leadership and Gender Gap within Firms: Evidence from an Italian Board Reform. *ILR Review*, 0(0):0019793920961995. 135, 162

Mas, A. and Pallais, A. (2017). Valuing Alternative Work Arrangements. *American Economic Review*, 107(12):3722–59. 136, 142, 143, 301

Matsa, D. A. and Miller, A. R. (2011). Chipping Away at the Glass Ceiling: Gender Spillovers in Corporate Leadership. *American Economic Review*, 101(3):635–39. 135

Mizala, A. and Urquiola, M. (2013). School markets: The impact of information approximating schools' effectiveness. *Journal of Development Economics*, 103:313–335. 22

Morchio, I. and Moser, C. (2020). The Gender Gap: Micro Sources and Macro Consequences. In *CEPR Discussion Paper No. 16383*. 133, 136, 143, 302

Neilson, C. (2019). The rise of centralized choice and assignment mechanisms in education markets around the world. Technical report. 17

Pande, R. and Ford, D. (2012). Gender Quotas and Female Leadership: A Review. *Background Paper for the World Development Report on Gender. April, 2011*. 135, 157

Park, M. and Hahm, D. (2023). Location choice, commuting, and school choice. Technical report. 23, 56

Pathak, P. A. and Sönmez, T. (2008). Leveling the playing field: Sincere and sophisticated players in the boston mechanism. *American Economic Review*, 98(4):1636–52. 86

Rao, G. (2019). Familiarity does not breed contempt: Generosity, discrimination, and diversity in delhi schools. *American Economic Review*, 109(3):774–809. 85, 86, 111

Ravitch, D. (2010). *The Death and Life of the Great American School System: How Testing and Choice are Undermining Education*. ReadHowYouWant. 17, 24

Redding, S. J. and Weinstein, D. E. (2016). A Unified Approach to Estimating Demand and Welfare. *CEPR Discussion Paper No. DP11421*. 135, 158, 312

Rockoff, J. and Turner, L. J. (2010). Short-run impacts of accountability on school quality. *American Economic Journal: Economic Policy*, 2(4):119–47. 22, 25

Rosen, S. (1986). The Theory of Equalizing Differences. In *Handbook of Labor Economics*, volume 1, pages 641–692. Elsevier. 134, 154

Ross, A. M. (1950). *Trade Union Wage Policy*. Univ of California Press. 162

Rothstein, J. M. (2006). Good principals or good peers? parental valuation of school characteristics, tiebout equilibrium, and the incentive effects of competition among jurisdictions. *American Economic Review*, 96(4):1333–1350. 17, 21, 32, 46

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly Journal of Economics*, 116(2):681–704. 83

Sacerdote, B. (2011). Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far? In *Handbook of the Economics of Education*, volume 3, pages 249–277. Elsevier. 83

Sattin-Bajaj, C., Jennings, J. L., Corcoran, S. P., Baker-Smith, E. C., and Hailey, C. (2018). Surviving at the street level: How counselors' implementation of school choice policy shapes students' high school destinations. *Sociology of Education*, 91(1):46–71. 24

Skorge, Ø. S. and Rasmussen, M. B. (2022). Volte-face on the welfare state: Social partners, knowledge economies, and the expansion of work-family policies. *Politics & Society*, 50(2):222–254. 132

Son, S. J. (2020). Distributional Impacts of Centralized School Choice. page 62. 23

Sorkin, I. (2018). Ranking Firms Using Revealed Preference. *The Quarterly Journal of Economics*, 133(3):1331–1393. 133, 136, 143, 158, 302, 307, 310

Staiger, D. O., Spetz, J., and Phibbs, C. S. (2010). Is There Monopsony in the Labor Market? Evidence from a Natural Experiment. *Journal of Labor Economics*, 28(2):211–236. 161

Summers, L. H. (1989). Some Simple Economics of Mandated Benefits. *American Economic Review*, 79(2):177–183. 136, 154

Taber, C. and Vejlin, R. (2020). Estimation of a Roy/Search/Compensating Differential Model of the Labor Market. *Econometrica*, 88(3):1031–1069. 136

Tincani, M., Kosse, F., and Miglino, E. (2021). Subjective beliefs and inclusion policies: Evidence from college admissions. 23

Vatter, B. (2022). Quality disclosure and regulation: Scoring design in medicare advantage. Technical report. 19, 23, 44, 46, 49, 210

Visser, J. (2019). Institutional Characteristics of Trade Unions, Wage Setting, State Intervention and Social Pacts (ICTWSS) Database, version 6.0. *Amsterdam Institute for Advanced Labour Studies (AIAS), University of Amsterdam*. 132

Walters, C. R. (2018). The demand for effective charter schools. *Journal of Political Economy*, 126(6):2179–2223. 22

Wiswall, M. and Zafar, B. (2017). Preference for the Workplace, Investment in Human Capital, and Gender. *The Quarterly Journal of Economics*, 133(1):457–507. 136, 143