Designing Optimal School Choice Admissions Policies: Insights from a Lottery-Based Reform*

Sherrie Cheng
University of Pennsylvania

Zach Weingarten
University of Pennsylvania
Job Market Paper

December 26, 2025

Click here for the latest version.

Abstract

Many public school districts changed admissions criteria for magnet and other selective schools to expand access to underrepresented groups. Such reforms can also change students' application and enrollment choices, with downstream effects on school-level test score value-added, peer composition, and student-school match quality. This paper considers how to best design admissions rules to balance equity and efficiency trade-offs when students respond strategically and peer effects are endogenous. We study a natural experiment in Toronto, where achievement-based admissions to high school magnet programs were replaced with a centralized lottery that included race and gender quotas. We use an instrumental variables strategy based on randomized offers to identify the causal effects of magnet admission and exploit the policy's rollout to estimate its impact on district-wide student sorting, peer composition, and school-level value-added. We then develop and estimate an equilibrium school choice model with strategic application choices, peer effects that affect both the demand for schools and their value-added, and admissions decisions under both policies. In counterfactual policy experiments, we find that the reform raised average student welfare but generated large losses for students who would have been admitted under the original policy. In addition, the reform reduced segregation at the expense of match quality. This equity-efficiency trade-off persists even when either of the two studied Toronto policies are re-optimized to maximize efficiency. We show that strategy-proof, two-sided stable mechanisms based on Deferred Acceptance deliver the highest match quality among the policies considered and improve equity relative to the baseline.

JEL Classification: I20, I24, I31, D47, D63

Keywords: school choice; equilibrium effects; equity-efficiency trade-offs; magnet programs

---

*We are thankful to Petra Todd, Francesco Agostinelli, and Margaux Luflade for invaluable feedback and support. We also thank Juan Pablo Atal, Micah Baum, Matteo Bobba, Juan Camilo Castillo, Hanming Fang, Chao Fu, Shresth Garg, Oriol González Casasús, Olivier De Groote, Ioana Marinescu, Viviana Rodriguez, and Holger Sieg for helpful comments and suggestions. We benefited from additional feedback from participants at the Association for Education Finance and Policy (2025), the CESifo/ifo Junior Workshop on the Economics of Education 2025, the Toulouse School of Economics 2nd Economics of Education Workshop, and the 3rd Structural Microeconomics Conference at Duke University. We finally thank the Toronto District School Board for providing us with data. This paper was supported by the National Science Foundation Graduate Research Fellowship Program through grant number DGE-1845298.
Contact information: Cheng: scheng1@upenn.edu. Weingarten: zachwein@upenn.edu.

1 Introduction

Many public school districts offer selective educational options—such as gifted programs in elementary school or magnet programs in secondary school—that provide high-achieving students with an advanced curricula beyond that available in their default public school.[1] Recent concerns about equitable access to these programs have led districts worldwide to replace achievement-based admissions, which screen applicants on the basis of test scores and course grades, with achievement-blind alternatives such as randomized lotteries. These reforms aim to reduce historical and structural barriers facing disadvantaged students.[2] Despite their increasing prevalence, these policies remain controversial. Proponents argue that eliminating achievement-based admissions expands fair access to selective schools (Shapiro, 2021). Critics counter that such reforms can generate inefficiencies: high-achieving students may lose access to rigorous programs, while lower-achieving students may enter programs for which they are unprepared (Elsen-Rooney, 2024).

Resolving this debate is challenging for several reasons. First, changing admissions criteria can shift application and enrollment behavior. Students may update beliefs about their admission probabilities and the peers they expect to encounter, then revise where they apply and, conditional on receiving offers, where they enroll. Second, because school-level value-added depends on peer composition (Allende, 2019; Crema, 2024), large shifts in student sorting may reshape the district-wide distribution of value-added and feed back into program demand. Finally, although critics argue that achievement-blind initiatives are inherently inefficient, who benefits most from high-quality schools depends on student-school match quality, which is ambiguous. If magnet programs provide the greatest gains to high ability students, then achievement-based screening may maximize efficiency. However, if low ability students benefit most, then expanding access for them may be optimal.

In this paper, we analyze how admissions rules should best be designed to promote equity and efficiency in the presence of strategic student responses and peer effects. To do so, we study a natural experiment in the Toronto public school district, where achievement-based admissions to high school magnet programs were abruptly replaced with a centralized, achievement-blind lottery beginning in the 2023–2024 academic year. To further promote representation, the assignment mechanism incorporated race and gender quotas: 20% of all seats at magnet programs became reserved for Black, Hispanic, Middle Eastern, and Indigenous

[1] Magnet programs represent a large fraction of the secondary education market in the United States. Approximately 1 in 15 high schoolers attends a magnet program and 25% of the top 100 high schools in the U.S. are magnets (Magnet Schools of America, 2019).

[2] For example, Shanghai abolished a policy in 2018 that favored high socioeconomic status (SES) students (Zhu, Zhang and Wang, 2023), and Chile replaced selective admissions with a centralized assignment mechanism in 2015 (Kutscher, Nath and Urzúa, 2023). In the United States, large districts in Boston (Barry, 2021), Chicago (Pathak and Sönmez, 2013), New Orleans (Abdulkadiroğlu et al., 2016), and Philadelphia (Mezzacappa, 2021) introduced similar reforms to increase access and equity for marginalized students.

students (underrepresented minorities, URM) and 50% of seats in STEM programs were reserved for female students. In this paper, we show how this policy change provides exogenous sources of variation to identify the causal effects of admissions design on student sorting, peer composition, value-added, and equity-efficiency trade-offs.

We implement a two-part analysis to address our research question. First, we exploit the lottery-based admissions offers to estimate the causal effect of magnet attendance on school-level value-added (VA). In the spirit of Bau (2022), we allow VA to vary by student ability to capture heterogeneous match effects between students and schools. We then decompose these estimates into peer and non-peer components, which allows us to trace how changes in enrollment patterns feed back into short-run VA measures. Our focus is on short-run changes arising solely from changes in peer composition, holding programs' curricula and pedagogy fixed. With additional cohorts, a richer notion of VA and overall school quality can incorporate programs' endogenous adjustments to such policy reforms.

Second, we develop and estimate a static equilibrium model of high school admissions to capture the consequences of changing admissions rules. The model recovers two-sided preferences, that is, students' preferences over program characteristics and programs' preferences over student characteristics, and the model incorporates strategic application behavior, endogenous peer composition, and policy-dependent admission probabilities. Using the estimated model, we quantify equilibrium changes in student utility, achievement, segregation, school-level VA, and match quality that occur when the lottery-based admissions policy is introduced. We then conduct counterfactual analyses that optimize equity- and efficiency-oriented social objectives and map the trade-offs between them. We benchmark these policies against centralized, strategy-proof mechanisms and show that variants of Deferred Acceptance can raise match quality and reduce segregation beyond even optimized versions of the existing Toronto policies. Taken together, the results show how admissions design can meaningfully relax equity-efficiency trade-offs.

Our empirical analysis is based on newly collected student-level administrative data built in collaboration with the Toronto District School Board (TDSB). The data cover three cohorts of 9th grade students and include demographics (race, gender), school attendance (magnet program and zoned schools), and standardized test scores in math (grades 3 and 9) and reading (grade 3). For the first cohort subject to the achievement-blind admissions policy, we also observe detailed application records, including rank-ordered lists, initial offers, and lottery waitlist positions. We supplement these administrative records with publicly available information on school locations and amenities, and we combine school locations with students' imputed home addresses to incorporate a spatial component into the demand model.

Descriptive evidence shows that the policy change led to major shifts in student composition across

Toronto. URM representation in magnet programs rose by 88% (14 percentage points), and female representation in STEM programs increased by 5% (2.2 percentage points). These gains in representation, however, came with declines in academic preparedness: incoming achievement fell by 0.10 standard deviations (σ) on average, as measured by lagged reading and math scores. The resulting peer composition changes translated into lower 9th grade outcomes as well. Across all magnet programs, 9th grade math scores declined by 0.15σ, with declines nearly twice as large in STEM programs.

Motivated by the descriptive evidence of substantial changes in test scores and in student sorting across schools, we first analyze how the reform impacted the test-score value-added distribution across the district. We develop an instrumental variables framework that leverages randomized lottery offers and default neighborhood school assignments as instruments for enrollment, while controlling for students' *ex ante* likelihood of admission to schools based on their submitted rank-ordered lists and the rules of the Toronto lottery mechanism.[3] This approach provides credible estimates of each school's value-added effect on 9th grade achievement. We allow for heterogeneity in student ability by separately estimating the model for low and high ability students, as measured by their third grade standardized test score performance. This approach provides us with a measure of student-school match quality. We then decompose the estimated VA into peer and non-peer components to understand how enrollment shifts feed back into the overall VA distribution (Allende, 2019). We find that peers play a central role: having high ability peers raises returns for both high and low ability students, and having underrepresented minority peers lowers returns, especially for high ability students. As a result of the sorting changes induced by the reform, VA at traditional high schools rose by 0.1σ on average while VA at magnet programs fell by 0.2σ. In this sense, the policy acted as a VA "equalizer," reducing the discrepancy in VA access across school types by 72%.

Although the previous reduced-form analysis sheds light on the policy's short-run impact on VA, it does not capture how eliminating admissions criteria affects student welfare, nor does it identify which students should be prioritized in magnet program admissions or how to design admissions rules to meet particular social objectives. We therefore develop a static equilibrium model of school choice under different admissions regimes. On the demand side, students make strategic application and enrollment decisions, factoring in their policy-relevant likelihood of admission and the peers they expect to encounter. We show that these features are essential for capturing the equilibrium effects of counterfactual policies.

On the supply side, magnet programs admit up to capacity using program-specific rankings of applicants. Under the achievement-based regime, each program orders its applicants by a quality index that

[3] Kirkebøen, Leuven and Mogstad (2016) show that with unordered assignments, standard IV can yield coefficients that are difficult to interpret. We mitigate this by including low-dimensional controls that summarize students' partial preference orderings inferred from their rank-ordered lists, in the style of Angrist et al. (2024).

places weight on predicted 9th grade achievement and allows differential weights by URM and female status. Under the achievement-blind regime, school seats are allocated by the district's lottery subject to group quotas (URM for all magnets and female quotas for STEM programs).

We estimate the model in two parts. First, we recover students' preferences using simulated method of moments (SMM) and using the data gathered during the achievement-blind policy regime, when programs had no discretion over which students to admit. This setting isolates the demand side and simplifies identification of student preferences. With the demand side estimated, we then estimate programs' preferences for applicants through a separate SMM procedure applied to data gathered during the original regime.

We find substantial heterogeneity in preferences across student demographics. Students positively value a school's ability to raise their 9th grade test scores, but their choices are strongly influenced by peer composition. In particular, they exhibit a pronounced homophily bias, gaining significant utility from increases in the share of observably similar peers. On the supply side, STEM programs place the greatest weight on predicted 9th grade math achievement. All programs exhibit marginal preferences for female applicants, with art and IB programs demonstrating a marginal distaste for URM applicants.

The estimated model is first used to conduct a policy experiment to assess the welfare effects of the Toronto reform. Specifically, we simulate the effects of the achievement-based policy on the achievement-blind cohort and compare model-predicted utilities. Average student utility rises by about 2% under the achievement-blind policy, but the effects are heterogeneous. Female students see increases of 2.5% on average while male students see increases of 1%. High ability URM students experience modest gains, whereas low ability URM students lose on average. The largest winners are students who enroll in magnet programs post-reform and benefit from lottery quotas—female and URM admits gain about 30%. The largest losses are concentrated among students who would have been admitted under the achievement-based system but either lose an offer under the reform or shift to a lower-quality program. Utility for this group declines by more than 30% on average.

We next use the estimated model to evaluate a sequence of counterfactual policy experiments with the goal of designing admissions rules that balance equity and efficiency. We evaluate six potential objectives a social planner might prioritize—three related to equity (school-level segregation in magnet programs, representation gaps between magnet and traditional schools, and achievement gaps across student groups) and three related to efficiency (student-school match quality, student utility, and magnet program value-added). We trace out the equity-efficiency trade-off frontiers by considering all combinations of these objectives, using either the achievement-based or achievement-blind regime as the baseline. In the achievement-based setting, we recover the admission weights on student achievement and demographics that best achieve each objective. In the achievement-blind setting, we identify the optimal configuration of lottery quotas.

Our counterfactual analyses show that both of the admissions regimes that were implemented involve meaningful trade-offs and their design could be improved. However, maximizing match quality under the achievement-based (achievement-blind) regime exacerbates segregation by 45% (3%). This result raises the question of whether a mechanism can improve both outcomes relative to Toronto's original policy. We therefore study a class of allocation mechanisms that are strategy-proof (students truthfully report preferences over magnet programs) and yield two-sided stable matches (no student-program pair would prefer to deviate). Our focus is the Deferred Acceptance (DA) algorithm, which satisfies both properties.[4]

We study three variants of DA. The first, an achievement-based DA, uses the program priority structure recovered in our estimation, effectively applying Toronto's pre-reform policy through a DA mechanism. The second, an achievement-blind DA, replaces priorities with random lotteries, mirroring the post-reform policy applied to the DA structure. The third, which we refer to as the constrained-optimization DA, features programs ranking students by their marginal contribution to value-added, subject to the district's affirmative action quota for underrepresented minority students. The achievement-based and optimization-constrained variants produce more efficient outcomes than either Toronto policy, and all three versions minimize segregation relative to the pre-reform policy. These findings suggest that districts can sustain equitable initiatives without necessarily sacrificing efficiency under a well-designed, yet easy-to-implement allocation mechanism.

Related literature. This paper relates to several strands of the literature. First, there is a growing body of work on school choice mechanisms in large urban school districts and the subsequent demand for public education. Recent work in this strand focuses on settings with centralized assignment that utilize either the Deferred Acceptance (DA) algorithm or the Boston mechanism (Abdulkadiroğlu and Sönmez, 2003; Neilson, 2013; Allende et al., 2019; Agarwal and Somaini, 2020; Kapor, Neilson and Zimmerman, 2020; Abdulkadiroğlu et al., 2020; Calsamiglia, Fu and Güell, 2020; Campos and Kearns, 2022; Arteaga et al., 2022). By contrast, our setting features an abrupt shift from decentralized assignment to a fully centralized process that relies on a non-strategy-proof mechanism. We show that this reform increased student welfare on average but harmed students who would have attended magnet programs under the prior policy regime. We rigorously compare this mechanism to iterations of the DA algorithm to assess its efficiency and provide

[4] These two properties hold with two additional assumptions: students' rank-ordered lists are unconstrained, meaning students are free to rank all 28 magnet programs in the district if they so choose, and there are no marginal costs associated with adding another school to the list.

new insights on gains from DA adoption.

We also contribute methodologically by bridging two strands of the school choice literature that are typically treated separately. Structural models of school choice and student behavior often omit peers due to concerns about multiple equilibria (Agarwal and Somaini, 2018). Those that do incorporate endogenous peer effects generally rely on data from districts with strategy-proof centralized assignment or assume that students do not behave strategically (Epple and Romano, 1998; Rothstein, 2006; Fruehwirth, 2013; Allende, 2019; Barseghyan, Clark and Coate, 2019; Abdulkadiroğlu et al., 2020; Bau, 2022; Crema, 2024). Conversely, recent work that allows for strategic behavior in student decision-making typically excludes peer effects (Agarwal and Somaini, 2018; Luflade, 2018; Fack, Grenet and He, 2019; Calsamiglia, Fu and Güell, 2020; Arteaga et al., 2022). We develop and estimate an equilibrium model that incorporates both endogenous peer effects and strategic behavior. This distinction matters, as students' expectations about admissions depend not only on their likelihood of acceptance but also on the peers they anticipate to enroll with, both of which vary with policy. In our setting, a model that omits these components would overstate average utility gains. Although multiplicity is a concern, we impose a fixed equilibrium-selection rule across all counterfactuals so that estimated effects reflect policy variation rather than equilibrium switching.

Our paper also relates to work on affirmative action policies in education markets. Prior research on policies that grant admissions advantages to underrepresented groups has focused largely on higher education (Epple, Romano and Sieg, 2008; Arcidiacono et al., 2011; Belasco, Rosinger and Hearn, 2015; Arcidiacono, Lovenheim and Zhu, 2015; Arcidiacono and Lovenheim, 2016; Kapor, 2020; Otero, Barahona and Dobbin, 2021; Bennett, 2022; Borghesan, 2022; Bleemer, 2022, 2023). Relatively little attention has been given to primary and secondary education, partly because such policies are rarely feasible in U.S. public school districts. Existing K-12 studies in the United States, therefore, tend to focus on policies that broaden access without explicitly addressing race-based affirmative action (Abdulkadiroğlu, Pathak and Walters, 2018; Barrow, Sartain and De La Torre, 2020; Dur, Pathak and Sönmez, 2020; Ellison and Pathak, 2021; Idoux, 2022). We extend this literature by studying Toronto's explicit use of race- and gender-based quotas in secondary school admissions. Our empirical design further allows us to evaluate a large sequence of counterfactual admissions policies in equilibrium. By doing so, we quantify the trade-offs between equity- and efficiency-related objectives and show how these can be optimized from the perspective of a social planner, yielding concrete policy recommendations given possible affirmative action goals.

Finally, we contribute to a growing literature on estimating school-level value-added (VA). A small but robust set of studies focuses on how to recover credible VA measures (Dale and Krueger, 2002, 2014; Reardon and Raudenbush, 2009; Deming, 2014; Angrist et al., 2016, 2017, 2024; Abdulkadiroğlu et al., 2017, 2022;

Mountjoy and Hickman, 2021; Ainsworth et al., 2023). We build on strategies developed in Abdulkadiroğlu et al. (2017), Epple, Jha and Sieg (2018), and Angrist et al. (2024), together with insights from Allende (2019), to decompose VA into peer and non-peer components in a setting with voluntary school choice. We further estimate VA separately by student ability type, yielding a flexible measure of the returns a given school provides. This approach enables policy evaluation in terms of allocative efficiency across heterogeneous students. To our knowledge, the only other work to embed student-school match quality in a school choice model is Bau (2022), who focuses on family income and curriculum decisions in private schools. By contrast, we derive an ability-based measure of match quality that is endogenously determined by student sorting.

The rest of the paper is organized as follows: Section 2 describes the data sources, provides institutional background information, and details motivating descriptive evidence of the policy's impacts. Section 3 outlines our empirical strategy to recover school-level VA and presents the causal effects of the policy change on the distribution of VA. Section 4 develops the static equilibrium model and Section 5 outlines how the model is identified and estimated. Section 6 presents the results from the model estimation. Section 7 details our counterfactual policy analyses and discusses the findings. Finally, Section 8 concludes the paper.

2 Institutional Setting and Data

2.1 The Toronto Public School District

We focus on the district of Toronto, which is currently the largest school district in Canada and the seventh largest school district in North America. Toronto has a total annual enrollment of more than 200,000 students and serves a diverse student body. Approximately 30% of students self-identify as white, 40% as Asian, and 26% as either Black, Hispanic, Middle Eastern, or Indigenous. Nearly 15% of students identify as LGBTQ+ and 8% self-report as having a disability.[5] In 2021, the median family income in the area was $56,000 (USD). During the same year, 13% of the population lived below the poverty line, pointing to large income disparities common to urban areas.[6]

Toronto offers approximately 85 public secondary schools, which serve two-thirds of all high school students in the district.[7] Each student in the district is guaranteed a seat at their default secondary school

[5] These figures come from a 2023 survey administered by the district to all families (60% response rate).
[6] These statistics come from the 2021 Census conducted by the city of Toronto.
[7] Our analysis includes the 67 schools that have a 9th grade student body of at least 10 students. Approximately 25% of students in Toronto attend Catholic high schools and an additional 8% attend private schools. Both of these are unobservable in our data.

based on their home residence and school catchment areas. Toronto also offers 28 magnet programs, each of which is classified by a specialty focus, which we aggregate into three distinct groups: art, STEM, and International Baccalaureate (IB).[8] Due to the fact that nearly all magnet programs in the district begin in the 9th grade, along with the fact that 95% of magnet program applications are for seats allocated to 9th graders, our focus is on the transition from grade eight to grade nine.

2.2 The Policy Change to Magnet Program Admissions

In 2022, the Toronto District School Board (TDSB) approved a substantial change to the way magnet programs conducted admissions. Originally, programs operated under a decentralized, achievement-based admissions scheme in which applying students were required to submit holistic applications that included measures of academic achievement, such as middle school transcripts, entrance exam scores, interviews, statements of interest, and letters of recommendation. Each program had full discretion over their own admissions process, which depended on the program's focus.[9] Although this process was decentralized, TDSB limited students to apply to a maximum of two magnet programs in any given admissions cycle.[10]

Citing continued inequity in access to these programs, TDSB eliminated the achievement-based admissions process entirely beginning with the cohort of 9th graders who would enter magnet programs in the 2023–2024 academic year. Under the new achievement-blind centralized admissions scheme, students submitted rank-ordered lists (ROLs) of no more than two programs directly to the district. Programs could no longer ask students to submit a record of their academic histories, nor could they screen for ability using entrance exams, interviews, or auditions. Instead, students were admitted to programs via random lotteries administered by the district. These lotteries included additional race and gender quotas to promote student representation in competitive magnet programs. At each program, 20% of seats became reserved for underrepresented minority (URM) students, while 50% of all available seats at STEM programs became reserved for female applicants.

2.3 Details of the Lottery Mechanism

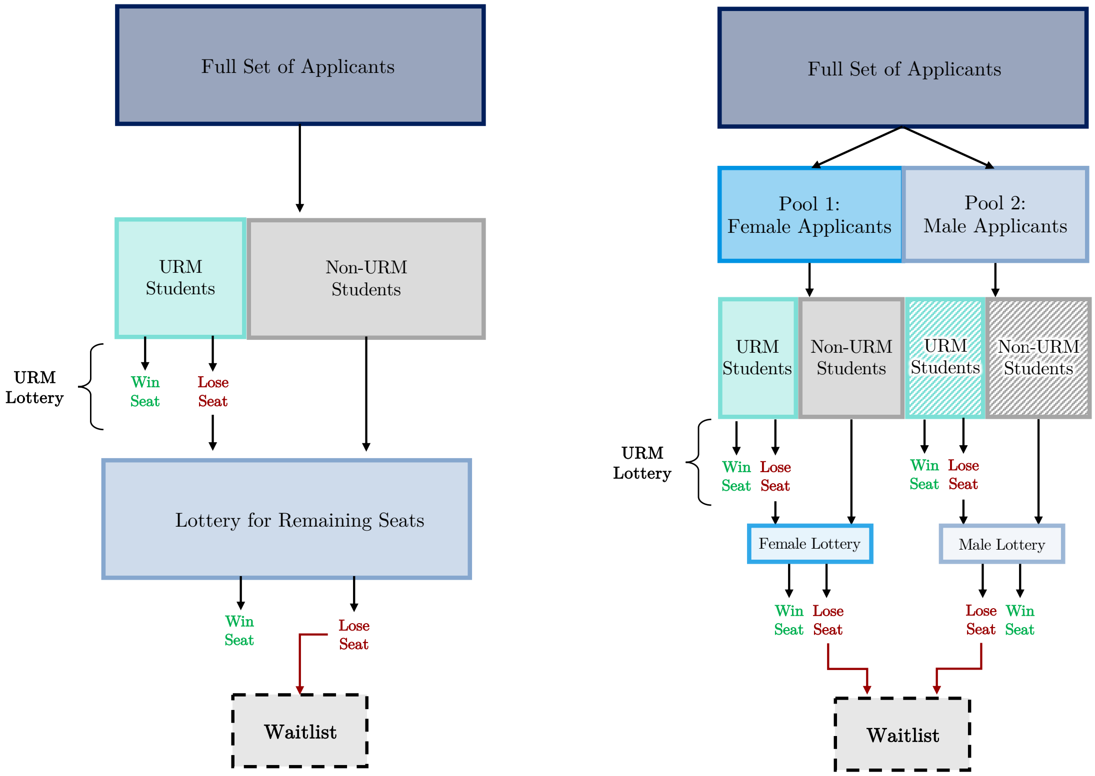
The achievement-blind admissions procedure can be described in the following five steps, which are also displayed diagrammatically in Figure 1:

[8] The district also offers specialized programs for athletic students, but we omit these programs from our analysis because they were not impacted by the policy change and did not screen admissions based on academic history in the baseline. Toronto also offers French immersion programs, but these typically begin in primary school and are therefore outside the scope of our analysis.

[9] For example, a visual arts magnet programs may have required students to submit a portfolio of their past artistic work.

[10] In the years that this process was decentralized, fewer than 1% of applicants submitted three or more applications, and the majority of students submitted only one application.

Figure 1: Flow Diagrams for Lottery Procedures at Different Magnet Programs



(a) Art and IB Programs

(b) STEM Programs

Notes: The above flow diagrams display the district's admissions design for the pool of applicants that rank a particular magnet program as their first or second choice. Figure 1a displays the mechanism for programs that only implement race quotas (Art and IB). Figure 1b displays the same but for programs that implement both race and gender quotas (STEM).

1. Eighth graders submit their magnet program ROLs to TDSB in the fall via an online portal.

2. A program's applicant pool is the set of students that rank that program *either* first or second. STEM programs split this pool into two groups—one for female applicants and one for male applicants—while non-STEM programs consider the pool in its entirety. All lotteries in the district are conducted simultaneously. Each student is assigned a random number that determines their admissions order.

3. The subset of URM students are assigned to seats first on the basis of their lottery numbers. If there are fewer URM applicants than allocated seats for URM students, all URM applicants receive an offer. Otherwise, URM students receive offers until 20% of total capacity is filled.

4. The remaining students (including URM students that did not receive an offer) then receive offers based on their lottery number until total capacity is filled. Students that lose the lottery are placed on the waitlist in the order of their lottery number.

5. Students that receive an offer from *both* their first and second choice automatically have their second choice offer rescinded. This newly available seat is filled by the next student at that program's waitlist. This process repeats until either every student has at most one offer or until capacity is filled at each program.

Students have one week to accept or reject their offer. In practice, students may initially accept their offer and then reject it before the next academic year. A small aftermarket occurs with the goal of filling all seats in the event of these late rejections from students.

2.4 Data

In collaboration with TDSB, we constructed three years of newly available administrative data on 9th graders. For the 2021–2022, 2022–2023, and 2023–2024 academic years, we observe the full universe of 9th graders' demographics, achievement, and enrollment. Demographic data include race, gender, and nationality. Achievement data include standardized test scores in reading and math from 3rd grade, as well as math test scores from 9th grade.[11]

The district also provided us with detailed enrollment records for all K-12 students. For each academic year, we observe 9th graders' enrollment decisions and their assigned default schools. The panel structure of the data allows us to recover each student's default middle school. In addition, about half of families participated in a district census survey, which adds information on parents' marital status and education.

We also obtained application data for the full set of 8th graders who applied to art, STEM, or IB magnet programs during the first year of the achievement-blind admissions scheme. These data include students' rank-ordered lists, lottery numbers, and offer and waitlist outcomes. Using unique student identifiers, we merge these records with enrollment decisions.

Finally, we assembled a three-year panel covering roughly 2,400 secondary school teachers in Toronto who worked at a magnet program during at least one year of our sample. While the data do not include teacher demographics or experience, they allow us to track teacher mobility across schools and to test for behavioral responses after the policy change. We present descriptive statistics related to this anonymized teacher data in Figure B1.

Several key characteristics of students and schools are not available in the administrative data. Although we do not directly observe students' home addresses, we do have residence-based middle and high school assignments for each student.[12] We collect publicly available catchment area boundaries for the

[11] Although students are also tested in 6th grade, the Covid-19 pandemic suspended these exams for the relevant cohorts.

[12] We also observe the designated elementary school for the subset of students that submit applications to magnet programs.

nearly 500 elementary, middle, and high schools in the district, which also provides us with exact geographical locations for each school, measured in latitude and longitude. We then overlay the recovered boundaries (one for each assigned school level) to create a finer zone of residency. We establish a student's home address as the centroid of their corresponding zone of residency.[13]

We use these imputed addresses to calculate travel times to each magnet program and to students' default high schools. Travel times are computed with the Google Maps API for local public transit, the primary mode of transportation for Toronto high school students. Times are averaged over two windows to capture the likely travel time when students commute to school: 45 minutes before the start of school and 15 minutes after the end.

Finally, the administrative data are missing information on race for 23%-33% of 9th graders, depending on the year. Similarly, 3rd grade math test scores are missing for about 27% of students. We assume this information is missing at random and impute these missing values by leveraging the corresponding racial distribution of students assigned to each default school, which proxies for neighborhood sorting. Test scores are similarly imputed using the available conditional distributions of test scores by default school and non-imputed race.

2.5 Analysis Sample

Table 1 presents summary statistics for our two main analysis samples. The first three columns report information on 9th graders in the 2022–2023 academic year, the last year under the achievement-based admissions policy. The final three columns display the same information for the 2023–2024 cohort, the first year of the achievement-blind policy.

In both years, we exclude from the analysis students with severe learning disabilities due to unobserved accommodations they may receive in traditional public schools, as well as the fact that many of these students do not take standardized exams. For the 2023–2024 cohort, we further restrict the magnet sample to those for whom we have linked application data. We remove a small fraction of students who enroll in magnet programs but have no ROL data available, resulting in a pre-reform magnet sample that is about 5% larger than the post-reform sample. Finally, we drop all students who either rank or enroll in athletic magnet programs, which were unaffected by the policy change and did not screen on academic ability.

Columns (1) and (4) of Table 1 show balance in district-wide cohort demographics between policies. Columns (2) and (5) restrict each analysis year to the subset of students that enroll in any magnet program. Columns (3) and (6) restrict only to those who enroll in a STEM program. Changes in magnet program

[13] See Appendix Figure B2 for an illustrative example of this imputation procedure.

Table 1: Descriptive Statistics for Analysis Samples

	2022–2023 School Year (Achievement-Based Admissions)			2023–2024 School Year (Achievement-Blind Admissions)		
	All (1)	Magnet (2)	STEM (3)	All (4)	Magnet (5)	STEM (6)
Female	0.488	0.585	0.462	0.483	0.598	0.484
Underrepresented Minority	0.256	0.157	0.174	0.274	0.296†	0.318†
Reading Score (Grade 3)	0	0.290	0.429	0	0.198†	0.230†
Math Score (Grade 3)	0	0.320	0.552	0	0.199†	0.325†
Math Score (Grade 9)	0	0.482	0.817	0	0.336†	0.545†
Distance Traveled (minutes)	19.49	35.27	35.03	19.42	32.77†	30.67†
College-Educated Parents	0.821	0.900	0.869	0.829	0.901	0.870
Two-Parent Household	0.764	0.834	0.849	0.761	0.803†	0.821
Observations	15,824	2,312	780	15,741	2,188	821

Notes: The above table displays descriptive statistics for the universe of 9th grade students attending Toronto public schools during the achievement-based admissions policy (2022–2023) and the achievement-blind admissions policy (2023–2024). Test scores are standardized to have mean 0 and standard deviation 1 across-district and within-year. “Magnet” includes all 9th grade students enrolled in either an art, STEM, or IB magnet program. “STEM” includes only those enrolled in a STEM magnet program. Travel times are calculated using Google Maps API with public transit routes. Data on parents comes from a sub-sample of families (49% of all students) that respond to annual student census surveys. The † symbol in columns (5) and (6) denote whether the difference relative to columns (2) and (3) are statistically significant at the 5% level (or lower).

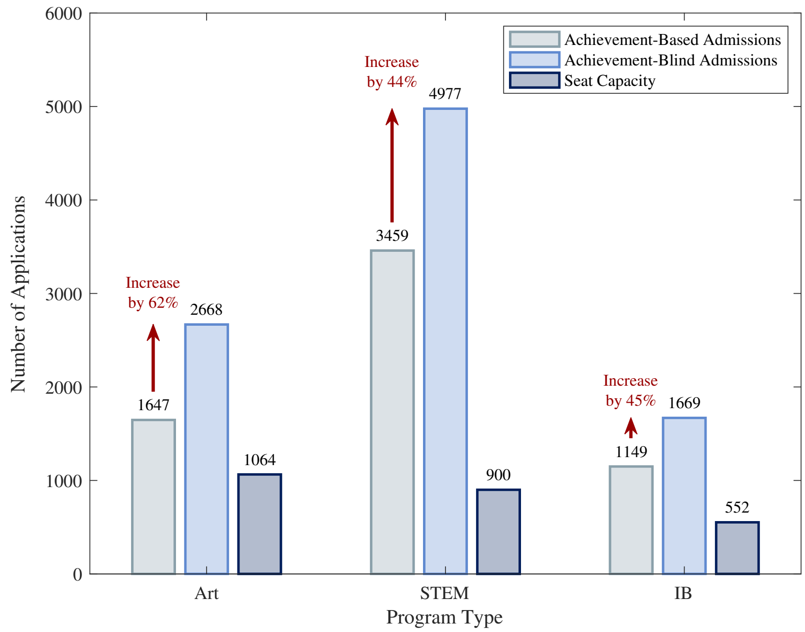
demographic composition capture changes in the way that students sorted into the magnet application system *and* changes in the way that programs admitted students. The policy change resulted in an increase in the representation of URM students in magnet programs by nearly 88% (from 16% to 30%). The incoming ability of enrolled students, as measured by lagged math test scores, declined by 38% between policies. Similarly, the achievement of 9th graders declined between policies, with 9th grade math scores decreasing by 30% across all magnet programs and 33% among STEM programs. These findings point to sizable changes in the composition of magnet-enrolled students and possible differences in the distribution of value-added across the district.

2.6 Descriptive Analysis

The elimination of screening mechanisms introduced by the achievement-blind admissions scheme led to a large increase in program demand, as measured by the total number of applications received. As shown in Figure 2, demand for magnet programs in Toronto greatly outpaces supply under either admissions regime. Following the reform, applications rose sharply—62% for art, 44% for STEM, and 45% for IB—pushing oversubscription rates to 2.5:1, 5.5:1, and 3:1, respectively.

We next investigate the behavior of 9th grade applicants during the achievement-blind admissions scheme. Table 2 displays the results of the lottery and subsequent enrollment decisions among the pool

Figure 2: Changes in Magnet Program Applications Between Policies



NOTES: The above figure plots the number of applications in the final year of the achievement-based admissions policy (2022-2023 SY, in light gray) and the first year of the achievement-blind policy (2023–2024 SY, in light blue). Dark gray bars denote the total number of available seats within each type of magnet program. Data for this figure come from publicly available reports on magnet program application trends in Toronto.

of students that apply to at least one magnet program (columns 1-3) and those that only apply to STEM programs (columns 4-6). We first examine overall trends before examining heterogeneity at the margin of race and gender. Panel A provides summary statistics for the results of the lottery. On average, 32% of students receive an initial offer from their favorite program and 12% receive one from their second-favorite program. The quota system gives URM students additional priority in the application cycle, which translates into a higher likelihood of receiving an offer. URM students are 13 percentage points (40%) more likely to receive an offer from their favorite program and 2 percentage points (18%) more likely to receive one from their second-favorite program. On the other hand, female applicants experience considerably fewer benefits from the STEM-specific quota. Compared to the average applicant, female students are only 7% more likely to receive a first-choice offer and 6% more likely to receive a second-choice one. Even among STEM-only applicants, the premium given to female applicants is considerably lower than that given to URM applicants.

Panel B reports the take-up rate of offers, i.e., the compliance rate of accepting an offer from either a first- or second-choice program after the initial allocation of offers. In general, students accept offers from their first-choice program 70% of the time and accept offers from the second-choice program 44% of the time.

Table 2: Application Outcomes Under Achievement-Blind Admissions

	All Rank-Ordered Lists			STEM-Only Rank-Ordered Lists		
	All Applicants (1)	URM Only (2)	Female Only (3)	All Applicants (4)	URM Only (5)	Female Only (6)
Panel A: Initial Lottery Outcomes						
Win Offer at First Choice	0.322	0.452	0.345	0.236	0.375	0.260
Win Offer at Second Choice	0.121	0.143	0.128	0.083	0.128	0.093
Receive No Offer	0.558	0.405	0.527	0.681	0.497	0.647
Panel B: Initial Offer Take-Up Rates						
First Choice	0.702	0.745	0.708	0.700	0.757	0.694
Second Choice	0.444	0.422	0.465	0.547	0.571	0.580
Panel C: Final Enrollment Decisions						
First Choice	0.327	0.436	0.350	0.244	0.367	0.264
Second Choice	0.068	0.070	0.072	0.053	0.075	0.060
Default School	0.414	0.324	0.388	0.510	0.386	0.481
Outside Option	0.191	0.170	0.191	0.193	0.172	0.196
Observations	5,739	1,304	3,110	2,317	493	944

Notes: The above table displays lottery outcomes for students that enter the application pool. The first three columns include all students that submit an ROL to the district, whereas the last three columns restrict to students that only rank STEM programs. Panel A presents summary statistics using the district's initial lottery in October of 2022. Panel B presents the take-up rate of each corresponding offer. For instance, the first entry indicates the average likelihood that any student receiving an offer from their first choice chooses to enroll in that program. Panel C presents summary statistics for the set of applicants in the beginning of the 2023–2024 academic year once enrollment decisions are finalized. Students that select the outside option are those who apply to the district's magnet programs from outside of the district. For these students, we only observe their lottery outcome and not their specific enrollment decision or their 9th grade test scores.

Lastly, Panel C details the realized enrollment decisions at the beginning of the 2023–2024 academic year. These statistics refer to the final allocation of seats once the assignment mechanism and aftermarket concludes. Ultimately, 40% of applicants enroll in a magnet program and 43% attend their default school. The remaining fraction of applicants attend either Toronto Catholic schools, private schools, or other schools outside of the district (all collectively denoting the outside option). For students that choose the outside option, we cannot observe any post-application decisions or test scores. In line with the findings of Panel A, URM students are more likely to enroll in a magnet program. Female applicants are marginally more likely than male applicants to enroll in a magnet program. These trends persist when we restrict to STEM-only applicants.

One possible explanation for why students may reject offers from programs that they choose to rank is that new information about programs is revealed to students between each stage of the application-enrollment cycle, as in Walters (2018) and Kapor, Neilson and Zimmerman (2020). Similarly, students' idiosyncratic tastes may change between the decisions to apply and to enroll. Another possible explanation is related to changes in the costs associated with applying. Whereas the achievement-based admissions policy required students to submit portfolios that demonstrated their ability (test scores, transcripts, letters of recommendation, etc.), the achievement-blind admissions policy has only the cost associated with filling out the application form and gathering information about schools. This change in costs may have increased demand from different types of students, such as non-compliers or those with limited program information.

The descriptive findings in this section point to large changes in the composition of students in magnet programs due to the policy change.[14] However, to better understand the ways in which this policy change *causally* impacted students and schools, we turn to Section 3 for a detailed discussion of our empirical strategy and findings.

3 The Causal Effects of Magnet Program Admissions

Our analysis so far has documented descriptive patterns in outcomes under alternative admissions policies. While informative, these results reflect raw differences in the data and cannot be given a causal interpretation. To rigorously assess the impact of admissions policies on the distribution of schools' ability to raise student achievement, we now turn to an empirical framework designed to isolate causal effects of admissions. A key input to this framework is a credible measure of school-level value-added (VA), defined as the contribution a school makes to its students' end-of-year 9th grade test scores relative to the district average.

Let $i = 1, ..., N$ index 9th graders and $j = 1, ..., J$ index magnet programs and traditional high schools. The conventional value-added model takes the form of a potential outcomes framework,

$$y_{ijt} = \alpha y_{it-1} + \boldsymbol{\gamma}'\boldsymbol{x}_{it} + \sum_{j=1}^{J} \widetilde{\theta}_{jt} d_{ijt} + \varepsilon_{it}, \tag{1}$$

where y_{ijt} is a measure of achievement for student i in year t, y_{it-1} is a lagged measure of student i's achievement, $\boldsymbol{x}_{it}$ is a vector of rich student-level controls, d_{ijt} is an enrollment indicator satisfying $\sum_{j=1}^{J} d_{ijt} = 1$, $\widetilde{\theta}_{jt}$ captures the return to test scores induced by attending school j in year t, and ε_{it} is an idiosyncratic error with mean zero.[15] This model provides a simple framework to measure the effect of attending school j relative to the average school in the district.

A causal interpretation of $\widetilde{\theta}_j$ requires quasi-experimental variation in enrollment that removes selection on unobservables not captured by $(y_{it-1}, \boldsymbol{x}_{it})$. In centralized school choice settings, researchers

[14] For additional descriptive analyses, see Appendix C.

[15] Following the value-added literature, lagged achievement is treated as a sufficient statistic for prior ability. The control vector $\boldsymbol{x}_{it}$ includes indicators for URM and female status, as well as a demeaned lagged reading score. Our preferred specification omits middle school fixed effects; including them leaves results essentially unchanged. School value-added estimates from a model with middle school fixed effects correlate 0.94 with our baseline estimates.

often exploit random tie-breaking to instrument for selective enrollment.[16] Our setting has centralized assignment for magnet programs, which are allocated via random lotteries in the post-reform period. The resulting lottery offers serve as natural instruments for magnet enrollment. For traditional high schools, there is no lottery assignment to leverage. Instead, we use default (zoned) assignment as an instrument for traditional enrollment. In Toronto, students are guaranteed a seat at their default school but may optionally elect to attend another public school, pending available seats. For this reason, assignment strongly but imperfectly predicts enrollment decisions. Under the standard IV conditions—relevance and the exclusion restriction that assignment affects outcomes only through enrollment—assignment provides identifying variation for the effect of traditional enrollment.

Crucially, our setting features *unordered treatments* from the fact that there is no single ranking of schools shared by all families. As Kirkebøen, Leuven and Mogstad (2016) emphasize, IV estimates can be difficult to interpret if idiosyncratic preferences are ignored.[17] Controlling for the full set of rank-ordered lists (ROLs) is infeasible in our setting, as we observe 385 distinct application types across nearly 6,000 students, many of which are unique to an individual. Instead, we follow the intuition of Abdulkadiroğlu et al. (2017) and adjust for preferences using program-specific assignment propensity scores constructed from students' ROLs and program-level admission rules. These propensity scores summarize the probability that a student would be offered a seat at each program and enter the model as controls, allowing us to isolate random offer shocks while flexibly absorbing selection on preferences in a low-dimensional way.

3.1 The Propensity Score

To address both self-selection into the magnet program lottery pool and heterogeneity in student-school matches, we extend the value-added model to further condition on the *ex ante* likelihood of receiving an offer from a particular magnet program. Unlike centralized assignment settings where students must submit complete ROLs to enter the public school system, students in Toronto are guaranteed a seat at their default school and may elect to optionally submit applications to at most two magnet programs. Our empirical strategy must therefore account for this channel of selective entry into the lottery pool.

Students who enter the pool reveal partial orderings of magnet programs, denoted by $\succ_i$, as captured by their submitted ROLs. Under the achievement-blind admissions scheme, magnet programs give differential priority to specific demographic groups through mandatory lottery quotas. In particular, all magnet programs were required to allocate at least 20% of their seats to URM students, while STEM programs were

[16] See, e.g., Hastings, Kane and Staiger (2006); Deming et al. (2014); Angrist et al. (2017); Abdulkadiroğlu et al. (2017); Gray-Lobe, Pathak and Walters (2023).

[17] See Appendix D for an illustrative example and detailed explanation.

additionally required to allocate 50% of their seats to female students. Let x_i^{URM} (x_i^f) indicate whether student i self-identifies as URM (female), and denote by z_{ij} an indicator for whether student i receives an initial offer from magnet program j through the random lottery procedure.[18] We define a student's *applicant type* as $\tau_i \equiv (\succ_i, x_i^f, x_i^{\mathrm{URM}})$, which jointly determines offer probabilities and forms the basis for our simulated propensity scores.

We construct the propensity score for student i at magnet program j as

$$p_{ij,r} \equiv \Pr(z_{ij} = 1 \mid \tau_i, r), \tag{2}$$

which captures the *ex ante* probability that student i receives an offer from j when ranking it r^{th} on their ROL. For any j that i does not rank, $p_{ij,r} = 0$. Because all magnet programs are oversubscribed, we avoid degenerate cases where $p_{ij,r} = 1$ for all i and r. In practice, we obtain propensity scores by programming and simulating the district-wide lottery procedure one million times, holding fixed students' preferences, school capacities, and lottery quotas.

Rosenbaum and Rubin (1983) show that conditioning on the propensity score eliminates omitted variable bias related to offers and types. In Table A2, we confirm that controlling for simulated propensity scores balances all observable student characteristics. Specifically, an uncontrolled model produces significant differences in offer frequencies across characteristics (at the 0.1% level), whereas once we condition on $p_{ij,r}$ no differences remain significant at even the 10% level. This supports the use of simulated propensities as valid controls for quasi-randomized offer assignment.

3.2 An Instrumental Variables Model of Value-Added

With the propensity scores in hand, we estimate a student-level two-stage least squares (2SLS) equation that treats the full vector of school indicators $\{d_{ij}\}_{j=1}^{J}$ as endogenous and recovers value-added measures $\{\theta_j\}_{j=1}^{J}$ for *both* magnet programs and traditional high schools. Let J denote the set of all schools, partitioned into magnet programs J^M and traditional high schools J^D, with $J^M \cap J^D = \varnothing$ and $J^M \cup J^D = J$. For magnet programs, the first stage uses randomized lottery offers z_{ij} as instruments; for traditional high schools, it uses each student's default (zoned) assignment z_{ij}^D, where $\sum_{j \in J^D} z_{ij}^D = 1$.[19] To accommodate unordered treatments and parsimoniously absorb selection on preferences, we also include as controls the J^M magnet program-specific simulated propensity scores $p_i \coloneqq \{p_{ij}\}_{j=1}^{J^M}$, together with prior achievement

[18] Table A1 shows that, among applicants, the lottery is random in observable characteristics that are uncorrelated with the student attributes prioritized in the mechanism.

[19] As a robustness check, we consider another instrument $z_{ij}^D \times \mathrm{dist}_{ij}$ in the style of Card (1993) that uses travel time at the assigned school as a shifter for realized enrollment. Results are similar when using this instrument.

(y_{it-6}) and baseline covariates $(\boldsymbol{x}_{it})$ that include race, gender, and lagged reading scores.

As written in Equation 1, the baseline specification imposes homogeneous treatment effects, meaning all students who attend school j are expected to receive an identical treatment level $\widetilde{\theta}_j$. Yet if certain schools are better matches for particular student types, this assumption can mask policy-relevant heterogeneity. We therefore allow VA to vary by ability, estimating separate 2SLS models for low and high ability students defined by lagged math performance (above/below zero on the baseline standardized scale).[20]

Putting all of these considerations together, we estimate the following fully-saturated equations via 2SLS:

$$d_{ijt} = \alpha^k y_{it-6} + \boldsymbol{\lambda}^{k\prime}\boldsymbol{x}_{it} + \sum_{j\in J^M} \pi_j^{M,k} z_{ij} + \sum_{j\in J^D} \pi_j^{D,k} z_{ij}^D + \boldsymbol{\rho}^{k\prime}\boldsymbol{p}_i + \eta_{ijt}^k, \tag{3}$$

$$y_{it} = \beta^k y_{it-6} + \boldsymbol{\gamma}^{k\prime}\boldsymbol{x}_{it} + \sum_{j\in J} \theta_{jt}^k \widehat{d}_{ijt} + \boldsymbol{\omega}^{k\prime}\boldsymbol{p}_i + \varepsilon_{it}^k, \tag{4}$$

where $k \in \{L, H\}$ indexes ability level group. We include the vector of simulated propensity scores $\boldsymbol{p}_i$ in both stages to absorb selection on application behavior and preferences in a low-dimensional way, ensuring that identification of $\{\theta_{jt}^k\}_{j=1}^J$ comes from the quasi-random variation in z_{ij} for $j \in J^M$ and in z_{ij}^D for $j \in J^D$. Standard errors are clustered at the school level.

The parameters of interest are $\{\theta_{jt}^k\}_{j=1}^J$, the district-wide set of value-added measures for each school j and student ability group k. Within the IV framework, these estimates are interpreted as local average treatment effects for compliers: students induced to enroll in a magnet program j because they won the lottery, and those who attend a traditional high school j' because they are zoned for it. This formulation provides a transparent way to interpret match quality across student types. For example, if $\theta_{\ell t}^L > \theta_{\ell t}^H$, then school ℓ is relatively more effective for low ability students.

3.3 The Decomposition of Value-Added

Although the heterogeneous value-added estimates provide a useful measure of VA, they do not reveal which underlying factors drive differences across schools. In particular, we are interested in the contribution of peers versus non-peer resources, because this distinction will matter for our counterfactual policy experiments that incorporate equilibrium sorting effects. To separate these components, we decompose VA in the spirit of Allende (2019).

[20] We define ability in this manner for consistency with our structural model. As a robustness check, we also form types using k-means on race, gender, and lagged scores. Results in Table A3 are qualitatively similar, albeit less precise due to the "low ability" group containing only one-third of the data.

Let $\boldsymbol{m}_{jt}$ denote the vector of school-level average characteristics of 9th grade peers, including the fraction of URM students, female students, and high ability students. Let $\boldsymbol{x}_j$ denote the vector of observed non-peer education inputs, including magnet program indicators and school building conditions, proxied by the demeaned number of health infractions recorded in the most recent Toronto School Health & Safety workplace inspection.[21]

After estimating Equation 4, we treat the recovered VA measures as data and project them on peer and non-peer inputs:

$$\theta_{jt}^{k} \equiv \theta_k(\boldsymbol{m}_{jt}, \boldsymbol{x}_j, \xi_j) = \boldsymbol{\theta}'_{m,k}\boldsymbol{m}_{jt} + \boldsymbol{\theta}'_{x,k}\boldsymbol{x}_j + \xi_j^k, \quad k \in \{L, H\}. \tag{5}$$

Here, $\boldsymbol{m}_{jt}$ denotes peer composition, $\boldsymbol{x}_j$ denotes non-peer school inputs, and ξ_j^k is an unobserved quality component known to students but unknown to the econometrician. We estimate Equation 5 by OLS; however, peer composition may be correlated with unobserved inputs such as teacher quality. To address this concern, we conduct two complementary 2SLS exercises in which $\boldsymbol{m}_{jt}$ is treated as endogenous.

First, we use lagged peer composition at school j as an instrument. The lottery assignment in the post-reform period supports the relevance of this instrument, but persistence in peer sorting raises concerns that lagged shares remain correlated with the error term. Second, following the logic of Berry, Levinsohn and Pakes (1995), we employ exogenous characteristics of nearby schools as instruments. Results from both approaches, reported in Table A5, are qualitatively consistent with the OLS estimates. However, weak instruments limit the precision of the 2SLS estimates, so we rely on OLS while recognizing the limitation of this approach.

Under this framework, the vector $\boldsymbol{\theta}_{m,k}$ captures *peer effects* for type k, that is, the extent to which VA responds to changes in student demographics. Similarly, $\boldsymbol{\theta}_{x,k}$ captures the *non-peer effects* of VA.[22]

We make two assumptions about this production function in order to trace out changes in VA between policies. First, we assume that the coefficients obtained by estimating Equation 5, as well as the corresponding functional form, are policy-invariant. In other words, changes to VA are captured directly by changes in education inputs and not changes in the relative productivity of any given input. Second, we assume that non-peer education inputs are fixed in the short run. This means that neither $\boldsymbol{x}_j$ nor ξ_j^k are indexed by t. Because our analysis compares VA between one year and the next, this assumption allows us

[21] These infractions include a mixture of building conditions like "ceiling damages" and safety concerns like "extension cord trip hazard." We show in Table A4 that the inclusion of additional school characteristics does not significantly change our results.

[22] Given data limitations, we risk overfitting by including interactions between peer and non-peer inputs. An empirical setting with more schools, or a panel structure with multiple years of lottery-based admissions, would provide greater power to identify such interaction terms.

to directly relate changes in VA to the resulting changes in peer composition. We support this assumption by documenting in Figure B1 that several key education inputs remained stable between policy years.[23]

Under these assumptions, we can write the change in ability-specific VA ($\Delta\theta_j^k$) as a function of the change in peer characteristics ($\Delta\boldsymbol{m}_j$) alone,

$$\Delta\theta_j^L = \boldsymbol{\theta}'_{\boldsymbol{m},\boldsymbol{L}}\Delta\boldsymbol{m}_j \text{ and } \Delta\theta_j^H = \boldsymbol{\theta}'_{\boldsymbol{m},\boldsymbol{H}}\Delta\boldsymbol{m}_j. \tag{6}$$

We emphasize that this method of measuring the change in VA between policies is necessary because the original policy does not feature quasi-random variation in school assignment. Therefore, we would obtain biased estimates of VA in the pre-period that would be difficult to compare to the 2SLS estimates from the post-period. We show in Table A6 that directly comparing our causal measures of test score returns during the achievement-blind policy to achievement-based VA estimates derived from a conventional OLS model generally overstates the relative change in VA induced by the policy. These findings are consistent with OLS estimates of magnet program returns being biased upwards from selection into treatment by high ability students.

3.4 Results

We present the summary estimation results of Equation 1 in Table A7. Our estimates for VA are similar between specifications, with a correlation of 0.74 (0.57) between the IV and OLS model estimates for low (high) ability VA levels. Figure B5 further demonstrates that VA is a key determinant of student demand: a one standard deviation increase in school-level test score returns for low ability students is associated with a 68% increase in oversubscription.[24]

Table 3 presents the results of the decomposition exercise (Equation 5).[25] These results reveal two key insights about the production of VA. First, there are important differences in the contribution of education inputs that explain why students with different abilities have greater or lower returns to the same school. An increase in the fraction of URM students at any given traditional high school by one percentage point is associated with a decrease in VA by 0.009σ for low ability students, but this same increase is associated with a decrease in VA of 0.014σ for high ability students. We do not find the fraction of female students to be particularly important in the production of VA. On the other hand, high ability peers are an important input for both low and high ability students, with a one percentage point increase in their representation

[23] Further discussions with TDSB confirmed that magnet programs were unable to adjust their curriculum in response to the policy change in the first year of its implementation.

[24] This is estimated from a univariate regression with estimated intercept 3.49 and coefficient 2.38 (p-value=0.088).

[25] Results from this exercise under alternative specifications of the VA production function are reported in Table A8.

Table 3: The Decomposition of the Returns to Value-Added by Education Inputs

	Peer Characteristics			Non-Peer Characteristics			
	Frac. URM (1)	Frac. Female (2)	Frac. High Ability (3)	Art (4)	STEM (5)	IB (6)	Building Condition (7)
Low Ability	−0.009	−0.000	0.016	0.305	−0.200	−0.251	−0.001
	(0.004)	(0.005)	(0.004)	(0.205)	(0.169)	(0.147)	(0.001)
High Ability	−0.014	0.014	0.017	−0.476	−0.037	−0.209	−0.001
	(0.005)	(0.014)	(0.005)	(0.678)	(0.148)	(0.183)	(0.001)

Notes: Student demographic characteristics are demeaned across district. URM refers to "underrepresented minority." Building condition refers to the demeaned number of annual safety and health infractions recorded in the most recent report conducted by the Toronto school district. Robust standard errors are reported in parentheses.

associated with an increase of low and high ability groups' VA level by 0.016σ and 0.017σ, respectively. With regard to non-peer effects, we recover imprecise measures of the contribution from magnet program types.

Second, we find that peer effects are the most important mechanism determining VA. A simple variance decomposition attributes roughly 46% (47%) of the total between-school variation in VA for low (high) ability students to differences in the share of high ability peers.[26] Variation in URM peer shares accounts for an additional 40% for either type. By contrast, all observed non-peer school characteristics together explain only 12% (9.5%) of the between-school variation in VA for low (high) types.[27]

With the VA production function in hand, we can estimate the effect of the policy change on short-run VA. To do so, we substitute into Equation 5 the 9th grade peer characteristics in the 2022–2023 academic year, when the achievement-based admissions regime was in its final year. We display these findings in Table 4. Results in Panel A come from the low ability production function, while Panel B displays the results for the high ability one. In each panel, we present findings under two different specifications for estimating VA: our preferred (2SLS) model and the conventional (OLS) model that does not address preference heterogeneity or selection bias.

In the discussion that follows, we focus only on the effects for low ability students, but emphasize that the findings are similar for the high ability students. Estimates using our preferred model show that the policy acted as a value-added equalizer. Traditional public high schools saw a rise in VA that aligned with a decline in magnet program VA. In particular, we find that traditional high school VA increased by 0.1σ while STEM program VA declined by 0.25σ. The effects on art and IB programs were also negative and sizable. The results are qualitatively similar when using the OLS estimates.

[26] We conduct a Shorrocks-Shapley decomposition of the regression's R^2 to arrive at these estimates.

[27] We do not have sufficient data to include teachers in this decomposition, meaning that the effect from teachers is not included in the observed school characteristics.

Table 4: Average Change in Value-Added by School Type

	All Schools	Traditional HS	Magnet Program Type: Art	STEM	IB
	(1)	(2)	(3)	(4)	(5)
Panel A: Low Ability					
2SLS Model	0.014	0.100	−0.146	−0.246	−0.108
	(0.021)	(0.015)	(0.065)	(0.059)	(0.075)
OLS Model	0.008	0.082	−0.107	−0.223	−0.110
	(0.020)	(0.016)	(0.072)	(0.061)	(0.075)
Panel B: High Ability					
2SLS Model	−0.003	0.104	−0.166	−0.305	−0.244
	(0.028)	(0.021)	(0.094)	(0.088)	(0.112)
OLS Model	0.003	0.082	−0.127	−0.223	−0.149
	(0.019)	(0.014)	(0.063)	(0.059)	(0.074)

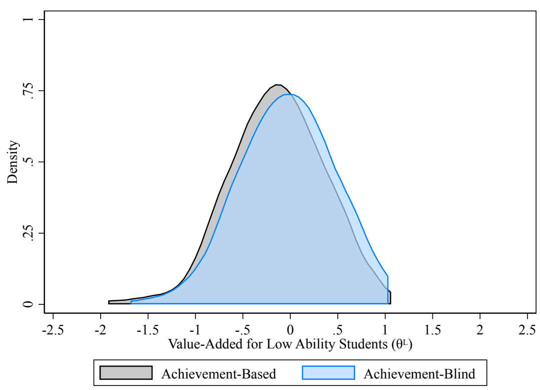
NOTES: The above table displays the relative change in value-added from the achievement-based policy to the achievement-blind policy, calculated as $\theta^k_{2024} - \theta^k_{2023}$ for θ^k_t the estimated return for ability type k in year t. The first row in each panel compares estimates using the 2SLS specification for our VA model, while the second row presents the same differences using estimates obtained by the conventional OLS model. Each cell is measured in standard deviations of 9th grade math test score achievement (σ). Standard errors in parentheses are derived from paired t-tests.

Figure 3 graphically displays these shifts in the distribution of value-added among traditional high schools (Panels (a) and (c)) and magnet programs (Panels (b) and (d)). On average, the elimination of admissions screening and the introduction of the lottery quotas led to an increase in the returns experienced by low (high) ability students at traditional schools by 0.1σ (0.09σ). At the same time, magnet program returns declined by 0.18σ (0.25σ) for low (high) ability students. Notably, the average disparity in VA between magnet programs and regular schools decreased by approximately 68%-76% as a result of this policy.[28]

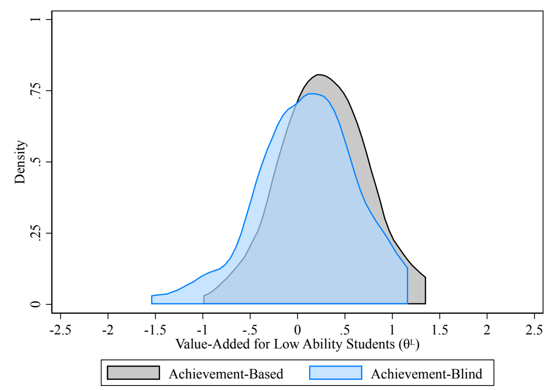
These distributional results have additional implications about possible match effects between students and programs. Namely, the reduction in overall magnet program VA could change which students have more to gain by attending these schools. In Figure 4, we compare the average test score gains experienced by low and high ability students under each regime. Although magnet program VA declined for both groups, we do not find evidence of a reversal in the pattern that high ability students gain more from attending STEM and IB programs on average, whereas low ability students gain more from attending art programs. However, the gains experienced by high ability students declined more in absolute terms, which suggests that the policy reduced their incentive to enroll in magnet programs marginally more than that of

[28] During the achievement-based policy, magnet programs had a 0.41σ (0.46σ) higher return for low (high) ability students on average. Under the new policy, this discrepancy became 0.13σ (0.11σ). Figure B6 displays these effects between school type and across years.

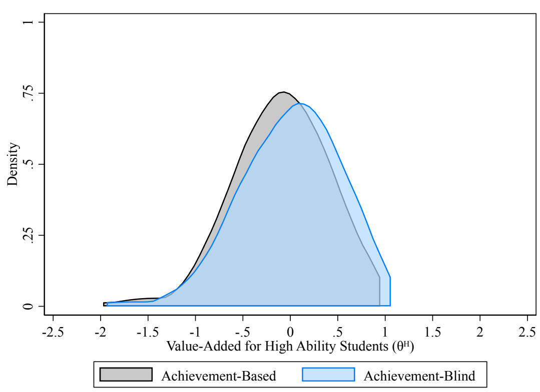
Figure 3: Change in the Distribution of Value-Added



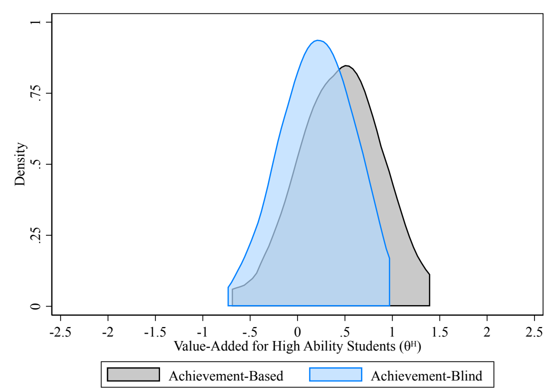
(a) Traditional High Schools (Low Ability)



(b) Magnet Programs (Low Ability)



(c) Traditional High Schools (High Ability)



(d) Magnet Programs (High Ability)

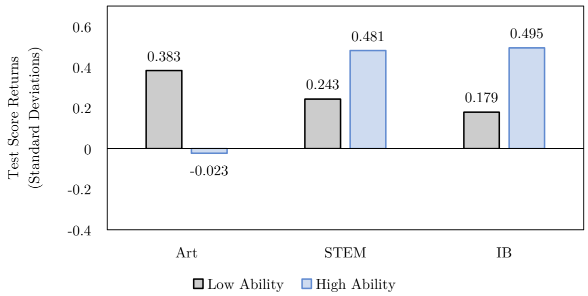
NOTES: The above figures display the change in the distribution of estimated VA between the final year of the achievement-based admissions policy and the first year of the achievement-blind one. Panels (a) and (c) display this change among the set of default high schools in Toronto while panels (b) and (d) display the same for the set of magnet programs. VA is measured in standard deviations of standardized 9th grade math test scores (σ).

low ability students. Figure B7 compares these returns at the program level and shows that 41% of programs offer larger gains for low ability students.[29] Together, these results point to heterogeneity in match quality and provide insight into the changing incentives of students to attend these specialized programs.

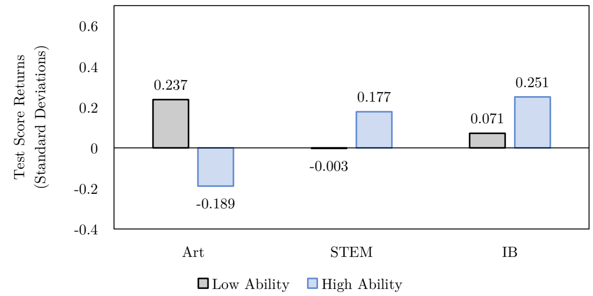
The results presented in this section show that the elimination of screening not only reshaped the overall distribution of VA but also altered relative returns across student types. The policy reduced—but did not eliminate—the advantage that high ability students receive from many magnet programs, pointing to meaningful heterogeneity in match quality. To identify which students gained or lost most from these shifts, we next turn to a structural model of student decision-making and preferences over schools.

[29] Figure B8 presents the same scatterplot using our imputed measures of achievement-based VA.

Figure 4: Returns to Programs by Ability Type and Admissions Policy



(a) Achievement-Based



(b) Achievement-Blind

NOTES: This figure displays the average estimated value-added for magnet programs during the achievement-blind regime by ability and by program type. Reported numbers are in terms of test score standard deviations, σ. The panels display these values respectively for the original admissions policy and the new admissions policy.

4 An Equilibrium Model of School Choice

To measure student welfare and evaluate counterfactual policies, we construct a static equilibrium model in which students make strategic application choices for secondary schooling, while magnet programs endogenously choose which students to admit. We formulate the model under both the achievement-based and achievement-blind admissions regimes. This equilibrium analysis, together with the reduced-form results in Section 3, is necessary for several reasons that we detail below.

First, the structural model allows us to measure changes in student utility that reduced-form estimates alone cannot capture. For example, even if the average magnet program experiences a decline in VA, average student welfare may still improve if students place greater value on other school characteristics, such as the share of URM students. Moreover, the model accounts for the fact that the effective match quality of a program is endogenous to student composition. As Figure B9 illustrates, a STEM program that initially delivers higher returns for high ability students may instead yield greater returns for low ability students after a sufficiently large increase in the share of URM enrollees.

Second, the reduced-form framework cannot generalize to a broader class of admissions rules. In other words, the previous analysis can only evaluate the observed policy reform. In contrast, the equilibrium model provides the flexibility to estimate the effects of a sequence of counterfactual admissions policies, using either regime as the baseline.

Finally, the initial analysis provides only partial insight into our central research question: how should an admissions system be designed in order to achieve equitable and efficient outcomes in the presence of strategic behavior and endogenous peer effects? The reduced-form estimates reveal important average effects of the implemented reform, but they cannot extrapolate to how alternative designs would reshape the equilibrium allocation of students across programs. Our structural model fills this gap by capturing both the demand side—students' heterogeneous preferences over programs—and the supply side—program admissions criteria that influence sorting and peer composition. By simulating counterfactual admissions policies within this equilibrium framework, we can evaluate how alternative rules shift program VA in the short-run, peer effects, and ultimately student welfare across different groups. In this way, the model allows us to directly confront the immediate distributional trade-offs inherent in designing admissions systems.

4.1 Environment

We consider a school district defined by its set of public school students entering 9th grade (I) and its set of secondary schools (J). The set of schools is the collection of default schools (J^D) that students are assigned to and magnet programs (J^M) that students may apply to. Let I_j denote the set of 9th grade students that enroll in school or program j. Although magnet programs are sometimes housed in secondary schools, we treat them as distinct to account for the fact that they tend to have non-standard curricula, a different student body, and, under the achievement-based admissions policy, the ability to screen students for entry. In the district, students are assigned a public school based on their residence and public school catchment areas. We denote the student's *default school* by $h_i \in J^D$.

During the fall of 8th grade, all students may apply to up to two magnet programs of their choosing. Let A_{ij} be an indicator for whether student i applies to program j. The application portfolio for student i is then $A_i = (A_{i1}, \dots, A_{iJ^M})$, which must satisfy the district's capacity rule, $\sum_{j \in J^M} A_{ij} \leq 2$.

Schools and programs are characterized by both observed and unobserved features. In particular, each j is defined by a pair of ability-specific VA measures $(\theta^L_{jt}, \theta^H_{jt})$, its location ℓ_j, the demographic composition of its student body $\boldsymbol{m}_{jt}$, and observed non-peer characteristics, including indicators for program type and building conditions, $\boldsymbol{x}_j$. We decompose VA following Equation 5, so that for each ability type $k \in \{L, H\}$, $\theta^k_{jt} = \theta_k(\boldsymbol{m}_{jt}, \boldsymbol{x}_j, \xi_j)$. This specification captures heterogeneous returns to value-added as well as endogenous sorting, as changes in peer composition directly map into VA levels for both low and high ability students. Consistent with our reduced-form framework, we assume that only peer inputs are variable in the short run. Finally, each school j faces a capacity constraint c_j that specifies the number of available seats.

Given this framework, we next detail the behavior of students and the endogenous admissions decisions of schools under both the original (achievement-based) and reformed (achievement-blind) systems. For readability, we suppress time subscripts in what follows.

4.2 The Student's Admissions Problem

Under either policy, student i receives utility u_{ij} from attending school j according to

$$u_{ij} = \alpha_i \theta_k(\boldsymbol{m}_j, \boldsymbol{x}_j, \xi_j) + \boldsymbol{\beta}_i' \boldsymbol{m}_j + \boldsymbol{\zeta}_i' \boldsymbol{x}_j - \psi D_{ij} + \varepsilon_{ij} + \varepsilon_{ij}^{\mathcal{O}}, \tag{7}$$

where $\theta_k(\boldsymbol{m}_j, \boldsymbol{x}_j, \xi_j)$ is the value-added for a student with ability $k \in \{L, H\}$, $\boldsymbol{m}_j$ is a vector of average student characteristics (the fraction of female students, URM students, and high ability students), $\boldsymbol{x}_j$ is a vector of non-peer components (indicators for magnet program type and the building condition), D_{ij} is the travel time (in minutes) using public transit from a school's location ℓ_j to the student's residential address ℓ_i, and ε_{ij} is an i.i.d. preference shock that follows a normal distribution with mean zero and standard deviation, σ_ε. The term $\varepsilon_{ij}^{\mathcal{O}}$ is a separate normal shock realized *after* programs make offers. In the spirit of Walters (2018) and Kapor, Neilson and Zimmerman (2020), we include this shock to rationalize changes in student behavior between application and enrollment stages.

We index the preference parameters for VA, peers, and non-peer characteristics by i, allowing them to vary with student-level characteristics that, based on empirical correlations in the data (Table A9), are likely relevant at the margin. In practice, we parameterize α_i as

$$\alpha_i = \alpha_0 + \alpha_1 x_i^f + \alpha_2 x_i^{\text{URM}} + \alpha_3 x_i^H. \tag{8}$$

In the above, x_i^f is an indicator for student i being female, x_i^{URM} is an indicator for being an underrepresented minority, and x_i^H is an indicator for being high ability. The baseline term α_0 can be interpreted as the preference for VA among non-URM male students with low ability. All other preference parameters follow the above form, except for ψ, which we assume to be homogeneous across all students. We impose the normalization $\psi = 1$ so that every other preference parameter carries a "willingness-to-travel" interpretation. To allow for flexible heterogeneity in the compliance rate of offer acceptances, we parameterize the standard deviation of the offer shock as $\sigma_{\mathcal{O},i} = \exp(\sigma_{\mathcal{O},0} + \sigma_{\mathcal{O},1} x_i^f + \sigma_{\mathcal{O},2} x_i^{\text{URM}} + \sigma_{\mathcal{O},3} x_i^H)$.

For completeness, we normalize the utility received from the outside option (leaving the district's public school system or attending a non-public school in Toronto) to zero and write $u_{i0} = \varepsilon_{i0}$. The utility a student receives from their default school (h_i) is therefore written as u_{ih}, with u defined as in Equation 7.

Achievement-Blind Admissions. We follow Calsamiglia, Fu and Güell (2020) and cast the student's magnet program application decision as a portfolio problem. In our setting, ROLs may have up to length two. Before filling out an application, students formulate their *expected reservation value* from either not entering the magnet application system or losing the lottery at all of their listed programs. We define this reservation value by

$$v_i := \max\{\widetilde{u}_{ih}, \varepsilon_{i0}\}, \tag{9}$$

where $\widetilde{u}_{ih}$ is the *expected* utility received from i's default school prior to admissions. The reason for this expectation is that the realized composition of peers at each school $j \in J$ depends on district-wide decisions of students and the subsequent random allocation of offers to programs. We discuss how these expectations are formed in Section 4.5.

Recall that students may rank up to two of the 28 total magnet programs offered by the district. Although this yields many possible combinations, the number of choices is computationally tractable and we can therefore tackle the optimization problem directly. Let $\mathbf{P}(J^M; 2)$ denote the set of all possible permutations of the J^M magnet programs conditional on a ROL capacity of two. The students' portfolio problem can be written as

$$\max_{A_i \in \mathbf{P}(J^M;2)} \Pi(A_i, \mathbf{\Omega}_i), \tag{10}$$

where A_i is the submitted application, $\Pi(A_i, \cdot)$ is the expected value of A_i, and $\mathbf{\Omega}_i$ is the vector of characteristics that govern student i's preferences and priorities.

Although Equation 10 embodies a generalized application problem, the mechanism we study is unlikely to be strategy-proof. One obvious reason for this is the district's imposed two-program limit on students' rank-ordered lists, which forces students to prioritize programs with better admission odds over those they would actually prefer to attend. It is therefore imperative to embed strategic behavior into our structural model. Note that the portfolio problem, as written, does not take into account students' strategic decision-making behavior. To adjust for this, we reformulate the problem as a sequential one that incorporates students' chances of admissions. Although the completed application is submitted at one time, students rank programs sequentially, which allows us to solve the model via backward induction (Calsamiglia, Fu and Güell, 2020). This sequential framework can also be rationalized by the fact that students have a much lower chance of receiving an offer at a program ranked second compared to the same program ranked first.[30]

With this in mind, we can rewrite Equation 10 recursively. Following the notation in Section 3.1, let

[30] We depart from Calsamiglia, Fu and Güell (2020) by not allowing for any students to be non-strategic. We make this decision for two reasons. First, as stated before, the policy studied is unlikely to be strategy-proof. Second, because entering the magnet pool is completely optional, students are more likely to be strategic in our context than in the Barcelona case studied in Calsamiglia, Fu and Güell (2020), where 96% of families were estimated to be strategic.

$\{p_{ij,1}, p_{ij,2}\}$ be the corresponding propensity scores for receiving an offer from program j in the first or second round, respectively. Program j_2^* solves the student's second-choice application decision,

$$V_2(\mathbf{\Omega}_i) = \max_{j_2 \in \{J^M \setminus j_1\} \cup \{\varnothing\}} \Big\{ p_{ij,2}\widetilde{u}_{ij} - \delta\kappa_i \mathbf{1}\{j_2 \in J^M \setminus j_1\} + (1 - p_{ij,2})v_i \Big\}. \tag{11}$$

Note that the propensity scores are simulated according to the district's true lottery rules and therefore account for the fact that student i does not receive an offer at their top choice. We let $\{\varnothing\}$ denote the choice to not rank any magnet program second. In practice, we substitute v_i for $\widetilde{u}_{ij}$ if this is chosen, which leads to the degeneration $V_2(\mathbf{\Omega}_i) = v_i$. Once again, $\widetilde{u}_{ij}$ refers to the *expected* utility student i will receive from attending j, where the uncertainty is a result of unknown peer composition realizations. Lastly, κ_i captures the marginal cost of submitting an application under the decentralized system and $\delta \in [0, 1]$ is the application cost discount factor associated with the second round of rankings.

We iterate backwards to the initial stage and recast Equation 10 as

$$V_1(\mathbf{\Omega}_i) = \max_{j_1 \in J^M \cup \{\varnothing\}} \Big\{ p_{ij,1}\widetilde{u}_{ij} - \kappa_i \mathbf{1}\{j \in J^M\} + (1 - p_{ij,1})V_2(\mathbf{\Omega}_i) \Big\}. \tag{12}$$

The student's continuation value is the utility they expect to receive from their first-ranked program, weighted by the *ex ante* probability of receiving an offer, plus the expected value of not receiving an offer (Equation 11). We parameterize the initial application cost as $\kappa_i = \kappa_0 + \kappa_1 x_i^f + \kappa_2 x_i^{\text{URM}} + \kappa_3 x_i^H$ to allow for possible heterogeneity in the costs faced by different types of students. Similar to before, we say that choosing $\{\varnothing\}$ means choosing the expected reservation value and substitute v_i for $\widetilde{u}_{ij}$.

Achievement-Based Admissions. During this period, magnet programs accepted students in a one-shot manner after receiving the final set of applications (see Equation 16). We assume that programs rank the set of applying students according to a *program-specific* notion of observed student quality. Unlike previously studied settings such as Boston or New York City, these admissions decisions did not depend on a standardized entrance exam but instead encompassed a more holistic array of student characteristics that predict achievement. Following the intuition of models of college admissions that depend on incoming ability levels and expected outcomes (Epple, Romano and Sieg, 2006; Fu, 2014), we therefore construct quality measures for each student according to

$$q_{is} = (\gamma_{0,s} + \gamma_{1,s} x_i^{\text{URM}} + \gamma_{2,s} x_i^f)\mathbb{E}[y_i \mid \boldsymbol{x}_i] + \upsilon_{is}, \tag{13}$$

where $s \in \{\text{Art, STEM, IB}\}$ and $\mathbb{E}[y_i \mid \boldsymbol{x}_i]$ is the student's expected test score with $\mathbb{E}[\varepsilon_i^k \mid \boldsymbol{x}_i] = 0$. The $\gamma_{0,s}$ term captures the preference that a magnet program of type s places on a student's expected 9th grade test scores, while $\gamma_{1,s}$ and $\gamma_{2,s}$ denote possible differential admissions weights that programs place on expected achievement from URM and female students, respectively. The term υ_{is} is a residual quality component observed by programs but unobserved to the econometrician, assumed standard normal. This quality shock encompasses other material that programs may have screened on independent of expected achievement and observable demographics.

Let $\bar{q}_j$ be the endogenous marginal quality level that magnet program j accepts, described in detail below in Section 4.3. Students receive an admissions offer whenever $q_{ij} \geq \bar{q}_j$ and are rejected otherwise. Given the decentralized nature of this policy, and the fact that Toronto still restricted students to apply to no more than two programs, we continue to impose students as strategic agents. The first-choice program solves

$$j_1^* = \underset{j \in J^M \cup \{\varnothing\}}{\operatorname{argmax}} \Pr(q_{ij} \geq \bar{q}_j)\widetilde{u}_{ij} - \kappa_{ij}\mathbf{1}\{j \in J^M\}. \tag{14}$$

As before, the choice of $\{\varnothing\}$ represents the choice to not apply to any program and instead accept either the default school or the outside option. Equation 14 then represents a discrete choice problem with $J^M + 1$ alternatives.[31] Here, κ_{ij} is the marginal cost of applying to some program j, which is allowed to depend on the program itself.[32] Under the decentralized program, we do not impose the same discount cost as before because students have to submit applications individually to programs, which may include substantially different requirements. Instead, the second-choice program solves

$$j_2^* = \underset{j \in \{J^M \setminus j_1\} \cup \{\varnothing\}}{\operatorname{argmax}} \Pr(q_{ij} \geq \bar{q}_j)\widetilde{u}_{ij} - \kappa_{ij}\mathbf{1}\Big\{j \in \{J^M \setminus j_1\}\Big\} - \kappa_{ij_1}. \tag{15}$$

We include the first-choice cost (κ_{ij_1}) to account for the fact that students will apply to a second program only if their expected utility net the *total cost* of submitting two applications exceeds their reservation value. The decentralized nature of this admissions regime prohibits us from writing the student's application problem recursively as before.

As a final note, we emphasize that the two policies we study are unlikely to be strategy-proof. The modeling assumption that students are strategic is therefore crucial to accurately capture student behav-

[31] We allow students to apply to two schools but do not discount the cost of the second choice. Our lack of application data during this period further prohibits us from identifying an application cost parameter.

[32] Art programs ask applying students to submit different items which may be more or less costly than the types of items that a STEM program may ask for. In practice, we estimate only the achievement-blind costs and then impose an exogenous increase based on the changes to program demand between policies. See Section 5.3 for details.

ior in this environment. This assumption comes at a cost: when students consider their *expected* utility, their idiosyncratic taste shocks are scaled by their idiosyncratic probability of admissions. Therefore, the convenient choice probability form derived in demand models with additive logit shocks cannot be derived in our framework (Train, 2009). This fact motivates our decision to model idiosyncratic taste shocks as normally distributed.

4.3 The Magnet Program's Problem

Achievement-Based Admissions. Facing a fixed set of education inputs in the short run, a program's objective is to maximize the quality of its attending students. Thus, the equilibrium set of admitted students in program j, denoted by I_j^*, is the solution to

$$\begin{aligned} \max_{I_j \subseteq I_j^A} \quad & \sum_{i \in I_j} q_{ij} \\ \text{s.t.} \quad & |I_j| \leq c_j, \end{aligned} \tag{16}$$

where I_j^A is the set of students applying to program j and q_{ij} is defined as in Equation 13. The assumption that programs maximize their *total* student quality guarantees that the constraint binds whenever $|I_j^A| > c_j$. In other words, as long as a program receives at least as many applications as there are available seats, the capacity constraint implies a cutoff rule for admission. If $|I_j^A| \leq c_j$, program j admits all applicants and $I_j = I_j^A$. If $|I_j^A| > c_j$, program j admits its top c_j applicants.

Achievement-Blind Admissions. Under the achievement-blind scheme, the magnet program's problem (Equation 16) is replaced with the quota-based lottery procedure outlined in Section 2.2. All programs, when possible, must allocate a fraction $\lambda^{\text{URM}} \in [0, 1]$ of total available seats to students self-identifying as an underrepresented minority. Additionally, all STEM-focused programs must allocate a fraction $\lambda^f \in [0, 1]$ of their seats to female students. Accordingly, programs face the constraints

$$\sum_{i \in I_j} x_i^{\text{URM}} \geq \lambda^{\text{URM}} c_j, \tag{17}$$

and, for STEM programs,

$$\sum_{i \in I_j} x_i^f \geq \lambda^f c_j, \tag{18}$$

where x_i^{URM} and x_i^f are indicators for underrepresented minority and female students, respectively. In the district's baseline policy, $\lambda^{\text{URM}} = 0.2$ and $\lambda^{f} = 0.5$.[33] Under this scheme, programs have no ability to screen or exert discretion in admissions decisions. With one or both of Equation 17 and Equation 18 enforced, the district conducts lotteries for all applying students at each program.

4.4 The Student's Enrollment Problem

Once magnet programs solve their admissions problem (or after the lottery procedure concludes), they issue offers to students. Receipt of an offer triggers the realization of an offer shock, $\varepsilon_{ij}^{\mathcal{O}}$, at the program making the offer. Unlike the baseline preference shock ε_{ij}, which is drawn prior to the application stage, the offer shock is only realized when an offer arrives and does not affect initial application decisions. Following Walters (2018) and Kapor, Neilson and Zimmerman (2020), this shock rationalizes the empirically observed fact that some students decline offers from programs they rank highly, capturing idiosyncratic changes in preferences that are revealed only once an offer is in hand.

Denote by $\mathcal{O}_i$ the offer set of student i. With updated information, the student chooses the school that solves the following discrete choice problem:

$$j^* = \underset{j \in \mathcal{O}_i \cup \{h_i, 0\}}{\operatorname{argmax}} \left\{ \alpha_i \theta_k(\widehat{\boldsymbol{m}}_j^E, \boldsymbol{x}_j, \xi_j) + \boldsymbol{\beta}_i' \widehat{\boldsymbol{m}}_j^E + \boldsymbol{\zeta}_i' \boldsymbol{x}_j - \psi D_{ij} + \varepsilon_{ij} + \varepsilon_{ij}^{\mathcal{O}} \right\}. \tag{19}$$

Students form enrollment-stage expectations over peer composition, $\widehat{\boldsymbol{m}}_j^E$, which may differ from their application-stage expectations.[34] The choice set always includes the student's guaranteed default option h_i (the zoned school) and the outside option (indexed by 0). Students who receive no offers—either because they lose in all lotteries or never enter the application pool—select between $\widetilde{u}_{ih}$ and the outside option.

4.5 The Formulation of Rational Expectations

Students form rational expectations over the endogenous components that govern decision-making at each stage of the admissions process. In our model, these components are (i) admissions probabilities and (ii) the peer characteristics that enter utility (Equation 7). The former depends directly on the admissions policy in place, while the latter reflects equilibrium peer composition.

During the achievement-blind regime, students' beliefs about admission odds must be consistent with the propensity scores implied by the district-wide application pool, $\{p_{ij,1}, p_{ij,2} \mid \mathcal{A}\}_j$. Conversely, under

[33] Note that the constraints may not bind due to either an insufficient number of applications from students belonging to a targeted group *or* from an insufficient number of those students accepting offers.

[34] These differences in expectations across application and enrollment stages are driven by heterogeneity in the variance of the offer shock. However, given the strong channel of selection present at the application stage, a model that enforces enrollment-level rational expectations at the application stage would not generate substantially different results.

the achievement-based regime, admission probabilities depend on the relative quality of students in the applicant set I_j^A. We assume students have rational expectations about the quality of the marginally admitted student at each program j, denoted $\bar{q}_j$. In other words, students correctly anticipate the expected achievement level of the marginally admitted student, as well as their demographic attributes, but do not know v_{is}, which is private information shared between student i and program j. Rewriting the quality parametrization from Equation 13 as $q_{ij} = \boldsymbol{\gamma}_j' \boldsymbol{X}_i + v_{is}$ for program type s, and using the distributional assumption on v_{is}, the expected probability that student i earns admission to j under achievement-based admissions is

$$\Pr(q_{ij} \geq \bar{q}_j) = 1 - \Phi\left(\frac{\boldsymbol{\gamma}_j'(\overline{\boldsymbol{X}}_j - \boldsymbol{X}_i)}{\sqrt{2}}\right), \tag{20}$$

where $\overline{\boldsymbol{X}}_j$ are the characteristics of the marginally admitted student at j.[35]

In addition, students form rational expectations over peer composition. That is, when evaluating admissions and enrollment choices, they anticipate the equilibrium distribution of classmates at each program. Since VA depends on peer characteristics, this also implies rational expectations over VA in equilibrium. An analogous consistency condition is imposed when estimating preferences under the achievement-blind scheme.

4.6 Definition of Equilibrium

With the environment fully described, we can define the policy-specific equilibrium in this economy.

Achievement-Based Admissions. An equilibrium in this economy is a set of magnet program admissions probabilities, $\{\Pr(q_{ij} \geq \bar{q}_j)\}_{i,j}$, peer characteristics at the application and enrollment stage, $\{\widehat{\boldsymbol{m}}_j^A, \widehat{\boldsymbol{m}}_j^E\}_j$, program offers, $\{I_j^A\}_j$, and enrollment decisions, $\{d_{ij}\}_{i,j}$, such that:

- Admissions probabilities and peer characteristics are consistent with rational expectations;
- Program offers are the solution to the magnet program's problem (Equation 16);
- Students' expectations over marginally admitted students $\{\bar{q}_j\}_j$ are consistent with the realized set of marginal admits;
- Enrollment decisions are the solution to the student's enrollment problem (Equation 19).

Achievement-Blind Admissions. An equilibrium in this economy is a set of rank-specific magnet program admissions probabilities, $\{p_{ij,1}, p_{ij,2}\}_{i,j}$, peer characteristics at the application and enrollment stage,

[35] The derivation of this formulation is provided in Appendix E.

$\{\widehat{\boldsymbol{m}}_j^A, \widehat{\boldsymbol{m}}_j^E\}_j$, application portfolios, $\{A_i\}_i$, and enrollment decisions, $\{d_{ij}\}_{i,j}$, such that:

- Application portfolios are the solution to the student's application problem (Equation 12) and are induced by rational expectations over both peer characteristics and offer probabilities consistent with the lottery mechanism;
- Enrollment decisions are the solution to the student's enrollment problem (Equation 19) and are induced by rational expectations over peer characteristics at the enrollment stage.

5 Estimation and Identification

Denote by $\boldsymbol{\Theta} \coloneqq \{\alpha, \boldsymbol{\beta}, \boldsymbol{\zeta}, \sigma_\varepsilon, \sigma_\mathcal{O}, \kappa, \delta\}$ the collection of 42 parameters that govern utility. We estimate $\boldsymbol{\Theta}$ using data from the achievement-blind admissions policy, following the procedure described in Section 5.1. We then estimate the supply-side parameters, $\boldsymbol{\gamma}$, using data from the achievement-based admissions policy (Section 5.2).

5.1 Estimation for Achievement-Blind Admissions

In principle, solving the student's problem requires a nested fixed-point algorithm. The outer loop solves for students' beliefs over peer characteristics, while the inner loop solves for the *ex ante* probability of admission. To make estimation feasible, we alleviate this computational burden by imposing the equilibrium observed in the data. Specifically, we require that students' beliefs over peer characteristics and admissions probabilities are consistent with the application-level student demographics and the group-averaged propensity scores obtained in Section 3. To ensure parameter estimates are consistent with these expectations, we include school-level demographics for every district school as additional moments in the estimation procedure. By contrast, in counterfactual exercises we cannot impose data-based expectations and instead recover rational expectations via fixed-point iteration.[36]

Given a candidate parameter vector $\boldsymbol{\Theta}_0$, we simulate student application decisions R times. For each simulated application profile, we then simulate the district lottery to generate offers. Each offer draws an idiosyncratic offer shock from the distribution specified before. Students update their enrollment-stage

[36] The algorithm works as follows: for a given set of beliefs $\{p_{ij,1}, p_{ij,2}, \widehat{\boldsymbol{m}}_j^A\}$, students solve their application problem (Equation 12), yielding a district-wide set of optimal applications $\boldsymbol{\mathcal{A}}^*$ and corresponding peer characteristics $\{\boldsymbol{m}_j^A\}_j$. Updating admissions probabilities directly would require repeatedly simulating the district-wide lottery, which is computationally burdensome. Instead, we approximate the lottery with a linear probability model that conditions on student characteristics, program indicators, and programs' relative oversubscription. In Appendix F, we show that this projection closely replicates the true lottery probabilities, and thus serves as a computationally tractable substitute. The process iterates until peer characteristics (at the application and enrollment stages) converge.

beliefs about peer characteristics to match the equilibrium composition observed in the data. Applicants then choose between the magnet program in their offer set (if any), their default school, and the outside option. Non-applicants choose between their default school and the outside option. As students reject offers, programs send new offers down their waitlists until either no applicants remain or capacity is filled.

5.1.1 Simulated Method of Moments

We obtain estimates of $\boldsymbol{\Theta}$ by matching both application and enrollment behavior in the data. Importantly, the application problem includes the decision of whether to enter the pool at all, which allows us to incorporate the full population of students in the estimation procedure. Let M denote the vector of application and enrollment moments observed in the data, and $M(\boldsymbol{\Theta})$ the corresponding moments generated by the model given utility parameters $\boldsymbol{\Theta}$. We estimate $\boldsymbol{\Theta}$ using simulated method of moments (SMM), solving

$$\widehat{\boldsymbol{\Theta}} = \underset{\boldsymbol{\Theta}}{\operatorname{argmin}} \, (M - M(\boldsymbol{\Theta}))' W (M - M(\boldsymbol{\Theta})), \tag{21}$$

where W is a chosen weighting matrix.[37] Table A10 summarizes the 318 moments used to identify the 42 utility parameters. Standard errors are obtained via the delta method (see Appendix G).

5.2 Estimation for Achievement-Based Admissions

Given our assumption that student preferences are invariant across policies, we estimate the admissions weights $\boldsymbol{\gamma}$ (see Equation 13) after recovering the demand-side parameters $\boldsymbol{\Theta}$. To do so, we exploit enrollment data from the 2022–2023 school year, the final year under the achievement-based policy.

Recall that students anticipate an offer from magnet program j with probability $\Pr(q_{ij} \geq \bar{q}_j)$. Although application data are not available for this cohort, the previously estimated demand-side parameters allow us to model students' application and enrollment choices, as well as programs' offer decisions, in tandem. In the simulations, students may submit up to two applications. Programs then issue offers by solving their admissions problem, and the process repeats until each program's capacity is filled or no students remain in the pool.

We estimate $\boldsymbol{\gamma}$ via simulated method of moments (SMM). Let $\widetilde{M}$ denote the vector of moments observed in enrollment data during the achievement-based year, and let $\widetilde{M}(\boldsymbol{\gamma})$ denote the corresponding moments

[37] We establish $W = I$ but scale the enrollment moments across all 92 schools by enrollment size, in which the largest school is given a weight of 1.

generated in simulation for a given set of admissions weights $\boldsymbol{\gamma}$. We obtain estimates from

$$\widehat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma}}{\operatorname{argmin}} \, (\widetilde{M} - \widetilde{M}(\boldsymbol{\gamma}))' \widetilde{W} (\widetilde{M} - \widetilde{M}(\boldsymbol{\gamma})), \tag{22}$$

where $\widetilde{W}$ is a weighting matrix.[38] Standard errors for $\widehat{\boldsymbol{\gamma}}$ are again computed via the delta method.

5.3 Identification

Preferences are identified by naturally occurring variation observed in the achievement-blind application data. For instance, the baseline preference for value-added can be attributed to students applying to observably similar schools that differ only in their VA level. It is further identified by the tendency of students at low-VA default schools to apply to magnet programs. Heterogeneous preference parameters can similarly be identified by how these decisions vary at the margin of, e.g., race or gender. There are two primary reasons to believe that there is sufficient variation in our data to identify each of the preference parameters. First, we observe 385 unique application types, accounting for 49% of the total number of application combinations that are theoretically possible. Second, our estimation procedure embodies the decision to submit a rank-ordered list, giving us greater insight into which factors may influence an observably similar student to apply or not. This granularity in data at the extensive and intensive margin also allows us to distinguish between application costs and preferences. Finally, the variance of the post-application shock is identified by the tendency of different demographic groups to reject an offer from programs they rank.

In terms of estimating parameters during the achievement-based application regime, we first rely on our main identifying assumption that preference parameters are policy-invariant. This allows us to hold fixed all elements of the demand-side of the market, which pins down the remaining variation in enrollment behavior to supply-side preferences for students.[39] In other words, discrepancies in the allocation of students with observably similar characteristics (and therefore similar utility orderings) can be explained by program-specific preferences for attributes that are independent of utility. For instance, two students with the same taste for their favorite program j may each attend different schools in equilibrium based on how heavily that program values expected achievement and whether that program penalizes students of a given demographic group.[40]

The lack of application data during this period requires us to make an additional assumption about

[38] As before, we consider $W = I$ except that we scale enrollment moments down by relative school size.
[39] Differences in district-wide peer compositions are accounted for when we enforce students to have rational expectations over peer composition at each school, which rules out this variation from impacting our supply-side estimates.
[40] Identification of programs' preferences requires us to normalize the variance of students' unobserved quality to 1 at all program types.

how application costs changed between policies. It is natural to assume that costs were higher during the achievement-based policy, as students had to exert additional effort to collect administrative information (transcripts, letters of recommendation, etc.), as well as to complete additional tasks like interviewing and writing statements of interest. Given the evidence in Figure 2, we assume that students have elastic demand for magnet programs. In particular, we fix the own-price elasticity of demand for magnet programs to a conservative value of −2. This assumption generates achievement-based costs for art, STEM, and IB programs that are 45%, 29%, and 30% higher, respectively, than the estimated achievement-blind costs. In Figure B14 and its subsequent discussion, we show the robustness of our findings to this elasticity assumption.

We conclude this section by highlighting an important assumption underlying both our estimation as well as our counterfactual analyses. We normalize the utility of the outside option to ε_{i0} under every admissions regime, but this may ignore important equilibrium effects. Private schools or nearby public schools outside the district of Toronto may respond to stark policy changes in Toronto in order to competitively attract students. In this sense, a more appropriate parameterization of the outside option may take the form $u_{i0} = g(\tau, \boldsymbol{x}_i) + \varepsilon_{i0}$ for g some function that depends on τ, the policy in place, and $\boldsymbol{x}_i$, the set of student characteristics governing utility. Given administrative data limitations, as well as limitations in publicly available data on schools that encompass the outside option, we are unable to incorporate this parameterization. However, we do not observe immediate changes in the education inputs of Toronto public schools, which suggests that the policy may not have had immediate effects on the outside option either. If this is the case, our fixed specification of the outside option will not bias our policy recommendations.[41]

6 Estimation Results

6.1 Parameter Estimates

Table 5 displays the estimated structural parameters. In Panel A, we show the estimated values for the parameters that govern utility, including preference parameters, standard deviation terms for taste and offer shocks, and the achievement-based application costs. Panel B displays the estimated preferences that magnet programs have for student attributes during the achievement-based regime.

Consistent with Abdulkadiroğlu et al. (2020), we find that students value VA positively, though the

[41] Furthermore, TDSB has communicated with us that an estimated 5% of students in Toronto attend private secondary schools, meaning the market is small and unlikely to disrupt the market we study. Furthermore, Catholic schools and traditional high schools are not likely close substitutes.

Table 5: Parameter Estimates

Panel A: Utility Parameters				
	Baseline (1)	× Female (2)	× URM (3)	× High Ability (4)
Value-Added (α)	26.99 (0.162)	17.62 (0.204)	20.48 (0.249)	−14.66 (0.169)
Fraction Female (β^f)	182.14 (1.105)	438.80 (1.832)	240.69 (1.628)	156.53 (1.621)
Fraction URM (β^{URM})	98.67 (1.368)	−19.96 (8.131)	439.39 (3.992)	−206.34 (1.879)
Fraction High Ability (β^H)	332.36 (1.603)	123.55 (1.596)	203.98 (2.529)	299.14 (2.565)
Art (ζ^{Art})	−96.06 (0.376)	37.01 (0.239)	−55.95 (0.867)	−31.76 (0.351)
STEM (ζ^{STEM})	−65.78 (0.288)	6.317 (0.135)	55.85 (0.378)	53.34 (0.421)
IB (ζ^{IB})	−118.25 (0.545)	3.956 (0.302)	−38.40 (0.311)	32.47 (0.192)
Building Condition (ζ^b)	−29.64 (0.427)	−0.016 (3.978)	63.22 (0.655)	36.26 (0.394)
Taste Shock Std. Dev. (σ_ε)	204.79 (0.483)			
Offer Shock Std. Dev. ($\sigma_\mathcal{O}$)	4.691 (0.100)	−1.339 (0.207)	−0.556 (1.145)	0.053 (0.019)
First-Rank Application Cost (κ)	46.69 (0.184)	−15.24 (0.071)	−0.814 (4.104)	−22.40 (0.109)
Second-Rank Cost Depreciator (δ)	0.726 (0.003)			
Panel B: Program Preferences				
	Art (1)	STEM (2)	IB (3)	
Baseline—Expected Achievement (γ_0)	0.741 (0.039)	0.901 (0.047)	0.535 (0.028)	
× URM (γ_1)	−0.090 (0.017)	0.039 (0.015)	−0.057 (0.011)	
× Female (γ_2)	0.104 (0.023)	0.066 (0.023)	0.094 (0.014)	

Notes: Parameter estimates come from estimating the corresponding equilibrium model. Preference parameters and application costs (demand-side parameters, Panel A) are estimated using the achievement-blind admissions data. This includes students outside of Toronto who apply but are not accepted. Supply-side program preference parameters (Panel B) are estimated using enrollment data on students from the achievement-based admissions scheme. Panel A parameters are measured in travel minutes, other than δ, which is a scalar parameter. VA as measured in utility has a standard deviation of approximately 0.414 (0.695) for low (high) ability students, meaning that baseline preferences for VA can be interpreted as a willingness to travel 0.414α minutes for a 1σ increase in VA. Peer attributes are measured in percent (from 0 to 100), meaning they are interpreted as a willingness to travel 0.01β minutes for a 1pp increase in student attribute m. Building conditions are measured continuously from 0 to 1, meaning the estimates capture the willingness to avoid the school with the worst conditions (1) to that with the best (0). Offer shock parameters are exponentiated in the model. Panel B parameters are measured in reference to a standard deviation in a standard normal distribution. Standard errors, calculated from the Delta method, are presented below each point estimate in parentheses.

magnitudes are modest.[42] At baseline, low ability students are willing to travel about 11 minutes to attend

[42] Although we assume students observe all school characteristics, attenuated preferences for VA may reflect students' imprecise knowledge of this attribute (Corradini, 2023).

a school with one additional standard deviation in test score gains.[43] Female and URM students place the highest value on VA, being willing to travel an additional 7-8 minutes for the same gain.

We also find strong evidence of homophily along every demographic margin. The baseline student is willing to travel 3.3 minutes for a school with one percentage point more high ability peers, while high ability students are willing to travel an additional 3 minutes. Similarly, baseline students show little taste for URM peers, whereas URM students are willing to travel over 4 additional minutes for each percentage point increase in URM enrollment.[44]

We also document heterogeneity in preferences for non-peer characteristics. Program indicator coefficients capture tastes for features unrelated to peers, VA, or buildings—such as curriculum or instructional style. Baseline students (non-URM male students with low ability) exhibit overall distaste for magnet programs. In contrast, female students prefer art programs, URM and high ability students favor STEM programs, and high ability students display positive marginal preferences for IB programs. Both URM and high ability students are less sensitive to building conditions than others. Baseline estimates further suggest that submitting the first application under achievement-blind admissions cost about 47 minutes on average, with an additional 34 minutes for the second. These costs are lower for high ability and female students.

On the supply side, programs displayed heterogeneous priorities during the achievement-based regime. On average, STEM programs placed the highest baseline value on predicted 9th grade end-of-year math performance. Art (IB) programs' preferences for baseline future achievement are 18% (41%) lower than that of STEM programs. Our estimates suggest that art and IB programs placed less emphasis on URM students' future performance, whereas all programs exhibit positive marginal preferences for female students' short-run achievement.

6.2 Model Fit

The model replicates the data well, as evident by its ability to match both the moments targeted in the SMM procedures and auxiliary untargeted moments analyzed after estimation. On the demand side, the model reproduces key moments related to application decisions (Table A11), offer acceptances (Table A12), and enrollment choices (Figure B10). On the supply side, estimation closely matches broad enrollment patterns (Table A13), as well as school-specific enrollment outcomes (Figure B11). Importantly, we are able

[43] School-level value-added for each ability level has a standard deviation in the model of approximately 0.4 and 0.7, respectively. To facilitate interpretation, we transform this measure to the positive reals. Multiplying the baseline entry by 0.4 yields the interpretable coefficient of interest for low ability students.

[44] Because peer composition m_j is measured on the unit interval, coefficients in Table 5 refer to moving from 0% to 100% of a given peer attribute. A one-point increase therefore corresponds to a 0.01 change in m_j.

to replicate school-level enrollment shares at the margin of race, gender, and ability with high precision across both policy years and across all school types. In comparing simulated moments to targeted ones during the achievement-based regime, we find that the model understates the enrollment of URM students in magnet programs, especially in art and IB programs.

Figure B12 reports untargeted moments on enrollment decisions during the achievement-blind regime. Using the estimated preference parameters, we simulate the full application, lottery, and aftermarket process, and then compare the distribution of lagged achievement among magnet-enrolled students. Panel (a) displays this comparison for lagged math test scores, while Panel (b) does the same for lagged reading test scores (each scaled 0.1–4.9). The simulated and observed distributions align closely, showing that the model accurately reproduces how students of different ability levels select into magnet programs. Importantly, we target only enrollment differences at the high versus low ability margin, without using actual lagged test scores in estimation. Lagged reading scores, in particular, enter the analysis only here as an untargeted check. Taken together, the targeted and untargeted moments indicate that the model captures overall trends in student decision-making well, providing a credible foundation for the counterfactual simulations that follow.

7 Counterfactual Analyses

7.1 Overview

In this section, we first examine how the Toronto policy change affected average student utility. To do so, we take the universe of 9th graders who enrolled under the achievement-blind policy and simulate their outcomes had they instead been exposed to the original achievement-based admissions design. This analysis, presented in Section 7.2, provides insight into the broader consequences of the policy change.

We then turn to the question of how to optimally design a district-wide admissions scheme for oversubscribed programs. The optimal design depends on the district's objectives and the extent to which it prioritizes potential equity-efficiency trade-offs. In Section 7.4, we consider a set of competing social objectives, each capturing a different notion of "equity" or "efficiency." For each objective, we compare outcomes under the pre-reform and post-reform schemes.

Finally, we benchmark these findings against variants of the Gale-Shapley deferred acceptance (DA) algorithm. We first consider a generalized DA with unrestricted rank-order lists in which programs rank students to maximize expected 9th grade achievement. This conventional DA keeps programs' preferences as those estimated previously, but filters the admissions process through a DA mechanism. We then examine an achievement-blind variant in which program preferences are replaced by randomized lotteries. To complement these, we study a constrained-optimization DA in which schools rank students by their marginal contribution to value-added, subject to a district-wide URM quota designed to prevent segregation from worsening. Together, these three variants allow us to compare mechanisms that preserve existing achievement-based admissions, mechanisms that eliminate achievement-based preferences entirely, and mechanisms that directly target both efficiency and equity objectives.

7.2 Effects of the Policy Change on Student Utility

We assess the welfare impact of Toronto's admissions reform by comparing average student utility under the achievement-blind and achievement-based regimes. A natural approach would recover realized *ex post* utilities from enrollment outcomes under each policy combined with the estimated preference parameters. This is infeasible in our setting due to the fact that we do not observe unsuccessful applications or outcomes when multiple applications were submitted under the achievement-based policy. Using enrollment alone would therefore upward-bias average utilities for that regime. Instead, we hold the achievement-blind cohort fixed and simulate outcomes under both policies.[45] By fixing the underlying student composition, any change in welfare can be attributed directly to the policy change.

Table 6 presents the effects of adopting the achievement-blind admissions policy on average student utility across an array of different groups of the sample.[46] For each listed group, column 1 reports average effects and column 2 reports the corresponding percent change relative to the achievement-based levels. We conduct this analysis over 100 simulations and report the 5th and 95th percentile effects from these simulations beneath each average effect.

Overall, the policy was welfare-improving, with student utility increasing by 2% on average. We uncover heterogeneity by demographic groups, with non-URM students gaining more than URM students and female students gaining more than male students. Notably, high ability URM students experience gains (3%) while low ability URM students experience losses (−1%). This result can be explained by higher ability students attending magnet programs at a higher rate, leaving the share of URM students lower at traditional schools. For URM students who stay at these schools, their utility slightly declines on average.

Beyond these demographic differences, the structural model allows us to identify ninth graders who, while observed under the achievement-blind regime, would have been admitted under the achievement-based policy. We label these students "Pre-Reform Admits" in Table 6. They experience large welfare

[45] This requires a fixed-point routine to ensure rational expectations over program-specific demographic shares in the "out-of-sample" achievement-based counterfactual.

[46] Figure B13 displays the full district-wide distribution of utilities under both policies.

Table 6: Estimated Effects of Policy Change on Student Utility

	Mean (1)	Percentage Change (2)
All students	12.81 [11.34, 15.06]	2.02% [1.79%, 2.36%]
URM	6.15 [2.76, 10.26]	0.68% [0.30%, 1.13%]
Non-URM	15.27 [13.13, 17.79]	2.85% [2.45%, 3.31%]
Female	19.48 [16.61, 22.64]	2.46% [2.09%, 2.85%]
Male	6.46 [4.48, 8.99]	1.33% [0.93%, 1.85%]
High Ability	18.20 [15.45, 21.02]	2.63% [2.23%, 3.04%]
Low Ability	7.19 [4.86, 9.99]	1.24% [0.84%, 1.73%]
URM; High Ability	31.19 [24.02, 39.56]	3.14% [2.42%, 3.97%]
URM; Low Ability	−8.80 [−11.65, −5.14]	−1.03% [−1.36%, −0.60%]
Non-URM; High Ability	15.01 [12.32, 18.20]	2.44% [2.00%, 2.96%]
Non-URM; Low Ability	15.59 [12.37, 19.15]	3.61% [2.86%, 4.44%]
Pre-Reform Admits	−357.71 [−383.44, −335.45]	−38.30% [−40.10%, −36.65%]
Pre-Reform Non-Admits	35.41 [32.95, 38.78]	5.73% [5.34%, 6.28%]
Post-Reform Female Admits	317.52 [305.84, 333.72]	35.20% [33.74%, 37.38%]
Post-Reform URM Admits	320.52 [303.66, 342.00]	30.06% [28.37%, 32.12%]

Notes: Entries report the distribution of $u^1 - u^0$ within each group, where u^1 is the average utility under the achievement-blind policy and u^0 is the average utility under the achievement-based policy. Bracketed values beneath each row report the [5th, 95th] percentiles across 100 simulations. The "Percentage Change" column reports the percentage change in utility from adopting the achievement-based policy. "Pre-Reform" refers to the achievement-blind regime.

losses—about 38% of their mean achievement-based utility—reflecting both lost offers due to randomized, non-screening admissions and lower within-program utility under the altered peer composition. By contrast, the set of non-applicants and applicants who would not have received offers ("Pre-Reform Non-Admits") gain by nearly 6% on average, which is again consistent with both increased access and improved VA at default schools for these students.

At the same time, even though mean effects for URM and female students are relatively modest, those who enroll in magnet programs under the achievement-blind policy see substantial gains from the reform. Female magnet enrollees gain roughly 30% on average, and URM magnet enrollees gain about 28%.

Table A14 repeats the exercise holding fixed the marginal application cost across policies (set to the level recovered in our demand estimation). This counterfactual isolates mechanical features of the admissions rule from reallocation gains while still accounting for endogenous decision-making in response to the cost adjustment. Under a fixed-cost specification, mean utility falls by about 2%, indicating that allowing policy-dependent application costs is important for capturing the realized welfare improvements.

7.3 Sensitivity of Results

Figure B14 illustrates how the welfare estimates vary under alternative assumptions about the demand elasticity. We re-calibrate program-specific application costs to match the observed increase in applications implied by each elasticity. In the baseline, we set the elasticity to −2, which yields application costs under the achievement-based regime that are 45% higher for art, 29% higher for STEM, and 30% higher for IB. We then vary the elasticity over the range [−6, −1] and solve the equilibrium for each value.[47] Across this range, the implied welfare gains remain close to those in our baseline, indicating that the results are not particularly sensitive to this parameter choice.

We also assess robustness to model specification. Specifically, we recompute equilibrium outcomes in three restricted models: (i) one that holds peer characteristics fixed rather than updating them through the fixed-point algorithm, (ii) one in which students apply based only on expected utility, ignoring admissions probabilities, and (iii) one in which both features of our preferred model are eliminated. The results, reported in Table A15, show that omitting either endogenous peer effects or strategic behavior substantially overstates welfare gains. Although our baseline model implies that the policy increased average utility by about 2%, the restricted model without either mechanism implies gains of more than 5%. Although each model consistently shows that the policy weakly improved student utility, these comparisons underscore the importance of accounting for peer feedback and strategic application behavior, and show that these modeling assumptions are not innocuous.

To assess sensitivity to equilibrium selection in the presence of endogenous peer effects, we recompute policy effects after perturbing students' expectations about underrepresented-minority (URM) peer shares. For each school j, expected URM share is initialized at $m_j^* + \widetilde{m}_j$, where m_j^* is the baseline counterfactual share and $\widetilde{m}_j \sim \mathcal{N}(0, \sigma_m^2)$ is mean-zero noise. We consider $\sigma_m \in \{0.01, 0.02, 0.03, 0.04, 0.05\}$. Small

[47] We do not consider inelastic levels due to the fact that application costs quickly explode and we cannot rationalize admissions patterns observed in data. For example, $\epsilon = -1$ implies that art program application costs were 163% higher pre-reform.

perturbations leave our main conclusions intact; for example, $\sigma_m = 0.02$ yields an average utility gain of 1.9%. However, larger perturbations can overturn the results: $\sigma_m = 0.05$ produces a net welfare loss of 1.3%. These patterns are consistent with the possibility of multiple equilibria given students' strong preferences for similar peers. To hold the equilibrium selection rule fixed across simulations, all counterfactuals use the same starting vector of characteristics $\boldsymbol{m}_j$, so that differences reflect policy changes rather than algorithmic path dependence.

7.4 The District's Possible Objectives

Quantifying the equity-efficiency trade-off of a given policy first requires explicit definitions of "equity" and "efficiency" in the context of public education. We consider six distinct social objectives that capture possibly competing goals. Three focus on equity: improving demographic representation in magnet programs, balancing student representation between magnets and traditional high schools, and minimizing achievement gaps between demographic groups. The other three focus on efficiency: maximizing student-school match quality, maximizing student utility, and maximizing magnet program VA. We summarize these six objectives alongside their mathematical formulations in Table 7.

As a reminder, we evaluate each objective under both the pre-reform and post-reform designs. Under the former, the district chooses admissions weights ($\boldsymbol{\gamma}$); under the latter, it chooses race and gender quotas ($\lambda := \{\lambda^{\text{URM}}, \lambda^{f}\}$).[48] For brevity, we use $\boldsymbol{\Gamma}(\tau)$ to denote the relevant choice variables under policy τ. In the main text, we emphasize one equity-oriented social objective (school-level segregation) and one efficiency-oriented social objective (student–school match quality). Detailed descriptions for the full set of objectives are provided in Appendix H.

School-level segregation (EQ I). A policy that seeks to improve demographic representation in magnet programs aims for parity between each magnet's composition and the district-wide demographic distribution. We measure segregation using same-group exposure, the probability that a member of a group meets another member of the same group, following Idoux (2022) and Margolis, Dench and Hashim (2020). For program j, we define $S_{g,j} = SGE_{g,j} - P_g$, where $SGE_{g,j}$ is group g's same-group exposure at school j and P_g is the group's district-wide share. The segregation index is then $|S_{g,j}|$, which measures the distance between actual and integrated compositions. Minimization of school-level segregation is written as

$$\min_{\boldsymbol{\Gamma}(\tau)} \sum_{j=1}^{J^M} \sum_{g} |S_{g,j}|. \tag{23}$$

[48] We continue to assume that gender quotas apply only to STEM programs, while race quotas apply to all programs.

Table 7: Overview of Counterfactual Policy Objectives

N	Object of Interest	Policy Objective	Mathematical Representation
EQ I	School-level segregation	Demographic parity at the school-level	$\min_{\mathbf{\Gamma}(\tau)} \sum_{j=1}^{J^M} \sum_g \lvert S_{g,j} \rvert$
EQ II	Program-wide segregation	Demographic parity between magnets/non-magnets	$\min_{\mathbf{\Gamma}(\tau)} \sum_g \lvert SGE_g(J^M) - SGE_g(J^D) \rvert$
EQ III	Achievement gaps	Close test score performance gaps between groups	$\min_{\mathbf{\Gamma}(\tau)} \sum_g \lvert \overline{Y}_g - \overline{Y}_{\not{g}} \rvert$
EF I	Student-school match quality	Efficiently allocate students to schools on ability basis	$\max_{\mathbf{\Gamma}(\tau)} \frac{1}{J^M} \sum_{j=1}^{J^M} \{(\rho_{L,j} - \rho_{H,j}) \times (\theta_L(\boldsymbol{m}_j, \boldsymbol{x}_j, \xi_j) - \theta_H(\boldsymbol{m}_j, \boldsymbol{x}_j, \xi_j))\}$
EF II	Student utility	Maximize students' average well-being	$\max_{\mathbf{\Gamma}(\tau)} \frac{1}{N} \sum_{i=1}^{N} u_{ij}$
EF III	Magnet program VA	Maximize each program's overall VA	$\max_{\mathbf{\Gamma}(\tau)} \frac{1}{J^M} \sum_{j=1}^{J^M} \theta_L(\boldsymbol{m}_j, \boldsymbol{x}_j, \xi_j) + \theta_H(\boldsymbol{m}_j, \boldsymbol{x}_j, \xi_j)$

Notes: This table outlines the six social objectives considered in this paper. Column 2 presents the names of the objective as presented in the main text, column 3 describes the objective of the stated policy, and column 4 provides the mathematical formulation of the respective social objective. Notation is defined as in the text, but repeated here for clarity. $\mathbf{\Gamma}(\tau)$ denotes the choice variables conditional on the policy (τ) in place. Achievement-based (achievement-blind) choice variables are programs' admissions weights (lottery quotas). $S_{g,j}$ is the segregation index of group g at school j. $SGE_g(X)$ is the same group exposure of group g at school type $X \in \{J^M, J^D\}$, denoting magnet programs and traditional high schools, respectively. $\overline{Y}_g$ is the average test score across the district of group g. Utility is represented by u_{ij}. The VA of school j experienced by students of ability type k is denoted by $\theta_k(\boldsymbol{m}_j, \boldsymbol{x}_j, \xi_j)$, with $\rho_{k,j}$ the fraction of k types at j.

Student-school match quality (EF I). An efficiency-oriented objective accounts for both program VA and which students benefit most from attending. A social planner maximizes match quality by solving

$$\max_{\mathbf{\Gamma}(\tau)} \frac{1}{J^M} \sum_{j=1}^{J^M} (\rho_{L,j} - \rho_{H,j}) \big(\theta_L(\boldsymbol{m}_j, \boldsymbol{x}_j, \xi_j) - \theta_H(\boldsymbol{m}_j, \boldsymbol{x}_j, \xi_j)\big), \tag{24}$$

where $\rho_{k,j}$ is the realized fraction of students at school j who belong to ability group k. The objective function reflects the alignment between program VA differences and the distribution of student ability. It is positive when groups with comparative advantage in a program (relative to other groups) are more heavily represented there, and negative otherwise. Larger negative values indicate greater mismatch at the margin of ability.

7.5 Tracing Out the Equity-Efficiency Frontier

The policies, as written, treat each objective in isolation. In reality, a social planner may consider a weighted combination of one equity-related objective (EQ) and one efficiency-related objective (EF). Let $\omega \in [0, 1]$ denote the weight placed on equity and $(1 - \omega)$ the weight on efficiency. The planner's problem can be written as

$$\max_{\mathbf{\Gamma}(\tau)} \Big\{ \max_{\omega \in [0,1]} -\omega EQ + (1 - \omega) EF \Big\}, \tag{25}$$

where lower values of EQ indicate more equitable outcomes, hence the negative sign. The solution $\{\mathbf{\Gamma}^*(\tau), \omega^* \mid EQ, EF\}$ gives the optimal admissions scheme and equity priority for the specified objectives.[49]

Since equity and efficiency measures are on different scales, we normalize each outcome relative to its baseline value in the data:

$$\widetilde{EQ}_{\mathbf{\Gamma}(\tau)} = \frac{EQ_{\mathbf{\Gamma}(\tau)} - EQ_0}{|EQ_0|}; \quad \widetilde{EF}_{\mathbf{\Gamma}(\tau)} = \frac{EF_{\mathbf{\Gamma}(\tau)} - EF_0}{|EF_0|}, \tag{26}$$

where EQ_0 and EF_0 are the respective baseline values. Rewriting Equation 25 using this transformation gives

$$\max_{\mathbf{\Gamma}(\tau)} \left\{ \max_{\omega \in [0,1]} -\omega \widetilde{EQ}_{\mathbf{\Gamma}(\tau)} + (1-\omega) \widetilde{EF}_{\mathbf{\Gamma}(\tau)} \right\}. \tag{27}$$

In practice, we solve Equation 27 for all 18 combinations of objectives (three equity objectives × three efficiency objectives × two policy schemes).

7.6 Counterfactual Policy Results

We first focus on the achievement-blind regime. Here, the district optimizes over URM and female quotas to balance equity and efficiency, subject to the quota constraint $\lambda^f + \lambda^{\text{URM}} \leq 1$. As before, we emphasize the two objectives of the main text: minimizing school-level segregation and maximizing student-school match quality.[50] Figure 5 displays the effect of each quota combination on these two outcomes. Panel (a) plots the effect on representation (Equation 23), while Panel (b) plots the effect on match quality (Equation 24). Each cell is measured as a percent change relative to the baseline quotas of $\lambda^{\text{URM}} = 0.2$ and $\lambda^f = 0.5$.

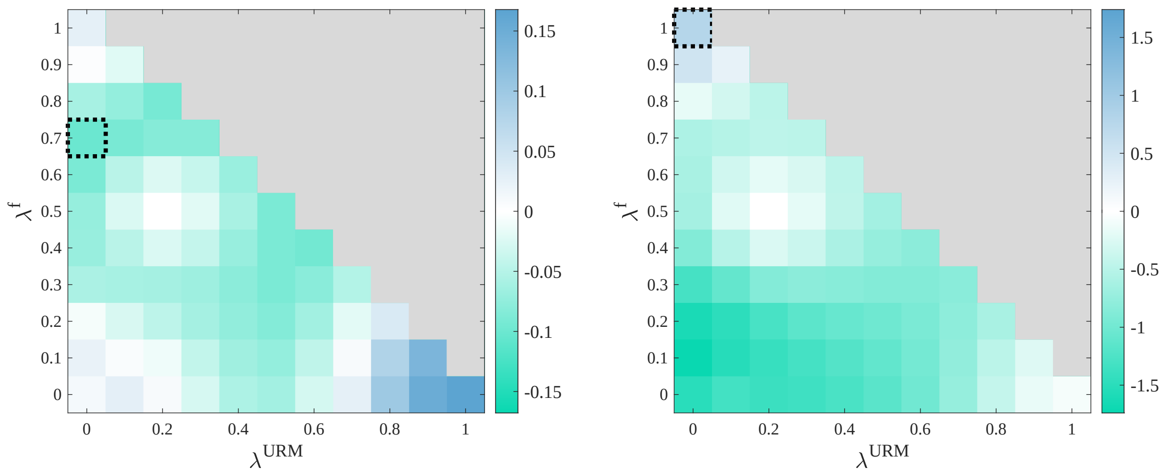
Two patterns stand out. First, the optimal policy depends heavily on the chosen objective: segregation is minimized at $\lambda^*_{EQ} = \{0, 0.7\}$, while match quality is maximized at $\lambda^*_{EF} = \{0, 1\}$. Second, optimizing one outcome worsens the other. At the equity-relevant optimum, segregation declines by 10% but match quality declines by 58%. Conversely, at the efficiency-relevant optimum, match quality improves by 79% and segregation increases by 3%.

We next turn to the achievement-based counterfactuals. Here, the district mandates the vector of nine admissions weights ($\boldsymbol{\gamma}$) that programs use to rank students. We define a grid of social weights $\widetilde{\omega} \in \{0,\ 0.25,\ 0.5,\ 0.75,\ 1\}$ and, for each value, solve the district's problem (Equation 27) across all nine equity-efficiency pairs. A grid search then identifies the $\widetilde{\omega}$-specific vector of optimal admissions weights. The

[49] We approximate Equation 25 using a discretized grid of ω values, compute $\mathbf{\Gamma}^*(\tau; \widetilde{\omega})$ for each grid point, and then select the $\widetilde{\omega}$ that maximizes the objective.

[50] Figure B15 reports similar results for the other four objectives.

Figure 5: Optimal Lottery Quotas



(a) School-Level Segregation **(b)** Match Quality

NOTES: Heat maps display percent changes relative to the simulated outcome with a URM quota of 0.2 and a female STEM quota of 0.5. Positive percent changes are in blue, negative in green, and the baseline is white. The black dotted cell marks the district's optimal policy choice.

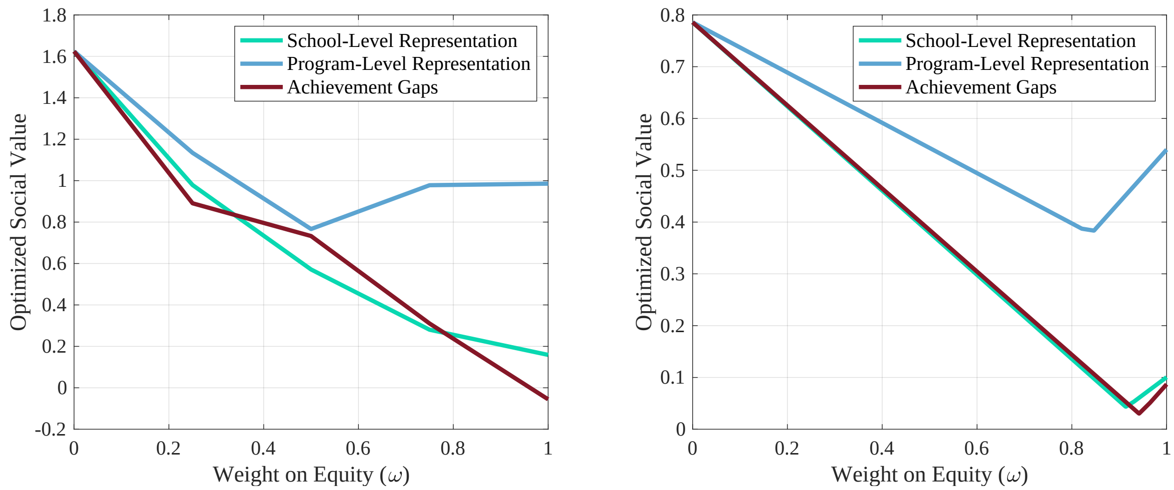
resulting solution is $\{\gamma^*, \omega^*\}$, consisting of program-specific weights plus the chosen social weight on equity. We report the solution for each combination of social objectives in Table A16.

Figure 6 plots the trade-off between match quality and each equity objective.[51] Panel (a) shows the frontier under the achievement-based regime. Panel (b) shows the same for the achievement-blind regime. Both imply substantial scope for improving match quality, with $\omega^* = 0$ (pure efficiency) emerging as the solution in every case that includes match quality. For the segregation-match quality frontier, maximizing match quality would require art programs to de-emphasize preferences for baseline achievement, STEM programs to penalize URM applicants in admissions, and IB programs to greatly increase their preferences for URM and female applicants.[52] Consistent with the achievement-blind findings in Figure 5, this design worsens segregation as well, here by 45%.

These results underscore the central challenge: whether it is possible to design an admissions system that simultaneously improves match quality while reducing segregation levels. Our analysis suggests that such simultaneity is not attainable within the frameworks of either Toronto policy we study.

[51] See Figure B16 for all other combinations.
[52] Table A16 shows the full set of optimal weights.

Figure 6: Match Quality Trade-Off Frontier



(a) Achievement-Based **(b)** Achievement-Blind

Notes: Each line plots the best-response frontier between match quality and a given equity objective, conditional on the stated equity weight. Panel (a) reports achievement-based results; Panel (b) reports achievement-blind. The maximum point on each line gives the attainable social value for the chosen pair of objectives.

7.7 Improvements from Deferred Acceptance

We have shown that both policies implemented in Toronto have scope for improvement. However, even optimized versions of these policies may remain suboptimal compared to other widely used allocation mechanisms. To investigate this, we consider three versions of the Deferred Acceptance (DA) algorithm.[53] In the first, which we refer to as the achievement-based DA, programs' preferences are defined as those estimated in Section 6. In the second, which we refer to as the achievement-blind DA, we replace student quality rankings with random rankings generated by a simple lottery. The final DA envisions programs as ranking students on the basis of their marginal contribution to VA, subject to the district's URM quota. This iteration, the "constrained-optimization" DA, seeks to maximize a possibly more efficient measure (relative to expected achievement) without exacerbating segregation. In all three DA counterfactuals, we extend the maximum ROL length to 28 and shut off application costs to ensure that the mechanism is strategy-proof (Pathak and Sönmez, 2013).[54] Appendix I provides generalized details on these mechanisms.

[53] DA has two useful properties—stability and truthful reporting. Because our demand model features peer effects, we implement DA as the inner step of a fixed-point procedure over peer composition. At the fixed point, the DA outcome is stable with respect to the induced rank-ordered lists and program priorities, and truthful reporting is approximately optimal in our large-capacity setting. For the achievement-blind version, statements hold conditional on a fixed lottery seed. See Appendix I for details.

[54] Strategy-proofness requires that applications be costless, meaning we establish κ as 0 in each of these counterfactuals. If we interpret the marginal costs to ranking programs as capturing the costs associated with learning about programs, then a natural interpretation for this adjustment is that Toronto pairs DA with an information campaign, like that recently conducted in New York City (Corradini, 2023).

Table 8: Effects of Admissions Schemes on Equity- and Efficiency-Relevant Objectives

	Pre-Reform	Post-Reform	Deferred Acceptance		
			Conventional	Blind	Constrained
	(1)	(2)	(3)	(4)	(5)
Panel A: Equity Objectives					
EQ I. School-level segregation	15.82	13.37	10.87	10.70	9.33
EQ II. Program-wide segregation	0.209	0.324	0.322	0.370	0.225
EQ III. Achievement gaps	0.553	0.510	0.645	0.658	0.418
Panel B: Efficiency Objectives					
EF I. Match quality	11.01	6.45	11.86	3.37	16.46
EF II.* *Ex post* utility	633.3	680.0	676.8	705.4	678.9
EF III. Program VA	5.271	5.132	5.475	6.181	6.157

Notes: The above table displays the six social objectives under four different admissions regimes. In each, we simulate student sorting using the achievement-blind administrative data. Column 1 presents the six objectives under Toronto's original achievement-based policy using the estimated supply-side preferences. Column 2 considers Toronto's new, achievement-blind policy with a 20% URM quota at all programs and a 50% female quota at all STEM programs. Column 3 displays the results of the achievement-based DA that fixes programs' preferences to those found in estimation. Column 4 shows the findings of the achievement-blind iteration of DA that replaces programs' preferences with a simple random lottery. Column 5 provides the results from the constrained-optimization DA with a URM quota set to 20% of seat capacity. * We report *ex post* utility which omits application costs from enrollment-stage utility. Including costs artificially lowers the utilities under the two Toronto policies, as we omit costs in the DA algorithm to preserve its property of strategy-proofness.

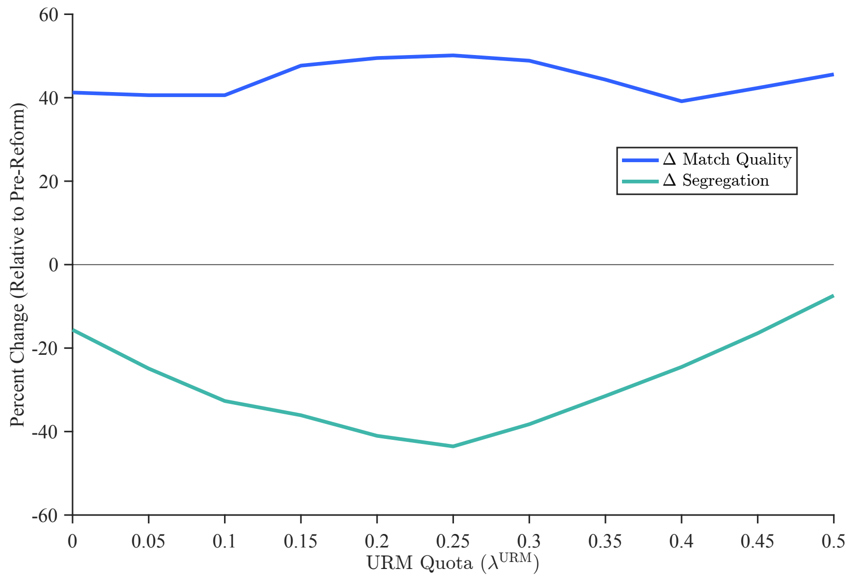
Table 8 compares Toronto's two enacted policies (columns 1 and 2) to the three DA counterfactuals (columns 3-5). All policies are simulated using data on the achievement-blind 9th grade cohort. Panel A presents our three social objectives related to equity, while Panel B presents those related to efficiency. Recall that improvements to the equity-relevant (efficiency-relevant) objectives occur when values are decreased (increased).

Across equity measures, all DA variants reduce segregation more than Toronto's reform, which is striking given reduced segregation was the explicit goal of the reform. We see little change in program-wide segregation from any of the tested policies. Notably, the constrained-optimization DA delivers the largest reduction in segregation and the greatest narrowing of achievement gaps among all tested policies. This finding related to minimizing achievement gaps is important and highlights the gains that can be made by admitting students to selective programs on the basis of value-added versus outright achievement.

We next examine match quality. Toronto's policy substantially reduced match quality (−41%), although the achievement-blind DA performs even worse (−69%) in this dimension. The total elimination of screening consistently worsens match quality. By contrast, the conventional DA and the constrained-optimization DA generate sizable gains (8% and 50%, respectively). Even with a 20% URM quota, admitting students by their marginal contribution to value-added delivers the largest improvement in match quality.

Turning to other efficiency outcomes, all four centralized mechanisms raise average student *ex post*

Figure 7: Segregation-Match Quality Trade-Off under Optimization-Constrained DA



NOTES: The above figure plots the equilibrium percent change in student-school match quality (blue) and segregation in magnet programs (green) relative to the achievement-based policy simulation and across different values of magnet program quotas for URM students. This DA algorithm envisions programs as ranking students on the basis of their marginal contribution to value-added.

utility by at least 7%.[55] Magnet program VA falls under Toronto's reform but improves under both the achievement-blind and constrained-optimization DA mechanisms.

Overall, the constrained-optimization DA improves equity and efficiency more than any other policy we test. We lastly study how the URM quota shapes these gains. Figure 7 plots equilibrium changes in match quality and segregation (relative to the original policy) across quota values, embedding each counterfactual in a fixed-point algorithm so peer characteristics are consistent with student expectations. Segregation declines as the quota increases up to district-wide parity ($\lambda^{\text{URM}} = 0.25$), and then steadily increases as URM students become overrepresented. However, even at large quota values, segregation remains below the original Toronto level. In fact, across the full range of quotas we examine, both match quality and segregation improve relative to the baseline. The best performance occurs at a 25% quota, where segregation is minimized and match quality is maximized.

Our findings underscore that admissions design can reconcile objectives often treated as competing. Toronto's reform did reduce segregation, but in doing so worsened match quality. However, mechanisms that elicit preferences and use DA reduce segregation further while improving match quality and student welfare. In particular, a constrained-optimization DA that ranks applicants by marginal VA—subject to a

[55] We focus on *ex post* utility here because costs are set to zero in DA but allowed to be non-zero in the two Toronto policies. Thus, we measure utility at enrollment and ignore cost differences. In Table 6, where rank lengths are held fixed, we incorporate application costs and directly compare realized utilities across Toronto policies.

modest URM quota—delivers the largest gains in match quality and the steepest declines in segregation, without degrading other outcomes. This mechanism has the added benefit of minimizing achievement gaps beyond any other policy tested. This suggests that pairing DA with VA priorities, calibrated by simple equity constraints, can produce large improvements in multiple, socially-relevant outcomes.

8 Conclusion

Public school districts across the world have begun eliminating achievement-based admissions in an effort to increase representation among disadvantaged students. Despite their growing prevalence, these policies remain controversial and relatively understudied. This paper fills that gap by analyzing the equilibrium effects of a large change to admissions criteria on both students and schools. We conduct our analyses using the setting of Toronto, where high school magnet program admissions abruptly shifted from a decentralized, achievement-based system to a centralized, achievement-blind one. We provide descriptive evidence that the reform substantially increased the representation of URM students in magnet programs, but also lowered the average ability levels of incoming students.

To understand the broader consequences of the reform, we estimate heterogeneous value-added models in an IV framework and study how the policy changed both program VA. We find that magnet program VA declined while traditional high school VA improved, effectively acting as a VA equalizer. We then develop and estimate a static equilibrium model of high school enrollment to measure impacts on student utility and to evaluate counterfactual admissions schemes. Our model shows that Toronto's reform increased average student utility, with especially large gains for those benefiting from the quota system. We highlight key trade-offs between policies that prioritize representation and those that prioritize match quality. Importantly, we demonstrate that more sophisticated admissions mechanisms, such as variants of the Deferred Acceptance (DA) algorithm, can raise match quality while also reducing segregation. We find that a DA variant that ranks students on the basis of their marginal contribution to VA achieves the largest gains in both equity and efficiency among the policies that we examine.

This paper is a first step toward a broader analysis of the equilibrium consequences of admissions reforms. Future work will extend the framework dynamically to study program responses to changing cohorts, student persistence, and longer-run outcomes such as college readiness and post-secondary enrollment. Together, this line of work will provide a richer understanding of both the immediate and long-run consequences of admissions reforms and will inform the design of admissions systems that balance equity with efficiency.

References

Abdulkadiroğlu, Atila, and Tayfun Sönmez. 2003. "School choice: A mechanism design approach." *American economic review*, 93(3): 729–747.

Abdulkadiroğlu, Atila, Joshua D Angrist, Peter D Hull, and Parag A Pathak. 2016. "Charters without lotteries: Testing takeovers in New Orleans and Boston." *American Economic Review*, 106(7): 1878–1920.

Abdulkadiroğlu, Atila, Joshua D Angrist, Yusuke Narita, and Parag A Pathak. 2017. "Research design meets market design: Using centralized assignment for impact evaluation." *Econometrica*, 85(5): 1373–1432.

Abdulkadiroğlu, Atila, Joshua D Angrist, Yusuke Narita, and Parag Pathak. 2022. "Breaking ties: Regression discontinuity design meets market design." *Econometrica*, 90(1): 117–151.

Abdulkadiroğlu, Atila, Parag A Pathak, and Christopher R Walters. 2018. "Free to choose: Can school choice reduce student achievement?" *American Economic Journal: Applied Economics*, 10(1): 175–206.

Abdulkadiroğlu, Atila, Parag A Pathak, Jonathan Schellenberg, and Christopher R Walters. 2020. "Do parents value school effectiveness?" *American Economic Review*, 110(5): 1502–39.

Agarwal, Nikhil, and Paulo Somaini. 2018. "Demand analysis using strategic reports: An application to a school choice mechanism." *Econometrica*, 86(2): 391–444.

Agarwal, Nikhil, and Paulo Somaini. 2020. "Revealed preference analysis of school choice models." *Annual Review of Economics*, 12: 471–501.

Ainsworth, Robert, Rajeev Dehejia, Cristian Pop-Eleches, and Miguel Urquiola. 2023. "Why do households leave school value added on the table? The roles of information and preferences." *American Economic Review*, 113(4): 1049–1082.

Allende, Claudia. 2019. "Competition under social interactions and the design of education policies." *Job Market Paper*.

Allende, Claudia, Francisco Gallego, Christopher Neilson, et al. 2019. "Approximating the equilibrium effects of informed school choice."

Angrist, Joshua D, Peter D Hull, Parag A Pathak, and Christopher R Walters. 2017. "Leveraging lotteries for school value-added: Testing and estimation." *The Quarterly Journal of Economics*, 132(2): 871–919.

Angrist, Joshua, Peter Hull, Parag A Pathak, and Christopher Walters. 2024. "Credible school value-added with undersubscribed school lotteries." *Review of Economics and Statistics*, 106(1): 1–19.

Angrist, Joshua, Peter Hull, Parag Pathak, and Christopher Walters. 2016. "Interpreting tests of school VAM validity." *American Economic Review*, 106(5): 388–392.

Arcidiacono, Peter, and Michael Lovenheim. 2016. "Affirmative action and the quality–fit trade-off." *Journal of Economic Literature*, 54(1): 3–51.

Arcidiacono, Peter, Esteban M Aucejo, Hanming Fang, and Kenneth I Spenner. 2011. "Does affirmative action lead to mismatch? A new test and evidence." *Quantitative Economics*, 2(3): 303–333.

Arcidiacono, Peter, Michael Lovenheim, and Maria Zhu. 2015. "Affirmative action in undergraduate education." *Annu. Rev. Econ.*, 7(1): 487–518.

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

Barrow, Lisa, Lauren Sartain, and Marisa De La Torre. 2020. "Increasing access to selective high schools through place-based affirmative action: Unintended consequences." *American Economic Journal: Applied Economics*, 12(4): 135–163.

Barry, Ellen. 2021. "Boston Overhauls Admissions to Exclusive Exam Schools." *The New York Times*.

Barseghyan, Levon, Damon Clark, and Stephen Coate. 2019. "Peer preferences, school competition, and the effects of public school choice." *American Economic Journal: Economic Policy*, 11(4): 124–158.

Bau, Natalie. 2022. "Estimating an equilibrium model of horizontal competition in education." *Journal of Political Economy*, 130(7): 1717–1764.

Belasco, Andrew S, Kelly O Rosinger, and James C Hearn. 2015. "The test-optional movement at America's selective liberal arts colleges: A boon for equity or something else?" *Educational Evaluation and Policy Analysis*, 37(2): 206–223.

Bennett, Christopher T. 2022. "Untested admissions: Examining changes in application behaviors and student demographics under test-optional policies." *American Educational Research Journal*, 59(1): 180–216.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. “Automobile Prices in Market Equilibrium.” *Econometrica*, 63(4): 841–890.

Bleemer, Zachary. 2022. “Affirmative action, mismatch, and economic mobility after California’s Proposition 209.” *The Quarterly Journal of Economics*, 137(1): 115–160.

Bleemer, Zachary. 2023. “Affirmative action and its race-neutral alternatives.” *Journal of Public Economics*, 220: 104839.

Borghesan, Emilio. 2022. “The Heterogeneous Effects of Changing SAT Requirements in Admissions: An Equilibrium Evaluation.”

Calsamiglia, Caterina, Chao Fu, and Maia Güell. 2020. “Structural estimation of a model of school choices: The boston mechanism versus its alternatives.” *Journal of Political Economy*, 128(2): 642–680.

Campos, Christopher, and Caitlin Kearns. 2022. “The Impact of Neighborhood School Choice: Evidence from Los Angeles’ Zones of Choice.” *Available at SSRN 3830628*.

Card, David. 1993. “Using geographic variation in college proximity to estimate the return to schooling.”

Corradini, Viola. 2023. “Information and access in school choice systems: Evidence from new york city.” Technical report, working paper.

Crema, Angela. 2024. “School Competition, Classroom Formation, and Academic Quality.”

Dale, Stacy B, and Alan B Krueger. 2014. “Estimating the effects of college characteristics over the career using administrative earnings data.” *Journal of human resources*, 49(2): 323–358.

Dale, Stacy Berg, and Alan B Krueger. 2002. “Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables.” *The Quarterly Journal of Economics*, 117(4): 1491–1527.

Deming, David J. 2014. “Using school choice lotteries to test measures of school effectiveness.” *American Economic Review*, 104(5): 406–411.

Deming, David J, Justine S Hastings, Thomas J Kane, and Douglas O Staiger. 2014. “School choice, school quality, and postsecondary attainment.” *American Economic Review*, 104(3): 991–1013.

Dur, Umut, Parag A Pathak, and Tayfun Sönmez. 2020. “Explicit vs. statistical targeting in affirmative action: Theory and evidence from Chicago’s exam schools.” *Journal of Economic Theory*, 187: 104996.

Ellison, Glenn, and Parag A Pathak. 2021. "The efficiency of race-neutral alternatives to race-based affirmative action: Evidence from Chicago's exam schools." *American Economic Review*, 111(3): 943–975.

Elsen-Rooney, Michael. 2024. "NYC to restore a controversial admissions boost for Manhattan students, Banks says." *Chalkbeat.*

Epple, Dennis, Akshaya Jha, and Holger Sieg. 2018. "The superintendent's dilemma: Managing school district capacity as parents vote with their feet." *Quantitative Economics*, 9(1): 483–520.

Epple, Dennis, and Richard E Romano. 1998. "Competition between private and public schools, vouchers, and peer-group effects." *American Economic Review*, 33–62.

Epple, Dennis, Richard Romano, and Holger Sieg. 2006. "Admission, tuition, and financial aid policies in the market for higher education." *Econometrica*, 74(4): 885–928.

Epple, Dennis, Richard Romano, and Holger Sieg. 2008. "Diversity and affirmative action in higher education." *Journal of Public Economic Theory*, 10(4): 475–501.

Fack, Gabrielle, Julien Grenet, and Yinghua He. 2019. "Beyond truth-telling: Preference estimation with centralized school choice and college admissions." *American Economic Review*, 109(4): 1486–1529.

Fruehwirth, Jane Cooley. 2013. "Identifying peer achievement spillovers: Implications for desegregation and the achievement gap." *Quantitative Economics*, 4(1): 85–124.

Fu, Chao. 2014. "Equilibrium tuition, applications, admissions, and enrollment in the college market." *Journal of Political Economy*, 122(2): 225–281.

Gray-Lobe, Guthrie, Parag A Pathak, and Christopher R Walters. 2023. "The long-term effects of universal preschool in Boston." *The Quarterly Journal of Economics*, 138(1): 363–411.

Hastings, Justine S, Thomas J Kane, and Douglas O Staiger. 2006. "Gender and performance: Evidence from school assignment by randomized lottery." *American Economic Review*, 96(2): 232–236.

Idoux, Clemence. 2022. *Integrating new york city schools: The role of admission criteria and family preferences.* MIT Department of Economics.

Kapor, Adam. 2020. "Distributional effects of race-blind affirmative action."

Kapor, Adam J, Christopher A Neilson, and Seth D Zimmerman. 2020. "Heterogeneous beliefs and school choice mechanisms." *American Economic Review*, 110(5): 1274–1315.

Kirkebøen, Lars J, Edwin Leuven, and Magne Mogstad. 2016. "Field of study, earnings, and self-selection." *The Quarterly Journal of Economics*, 131(3): 1057–1111.

Kutscher, Macarena, Shanjukta Nath, and Sergio Urzúa. 2023. "Centralized admission systems and school segregation: Evidence from a national reform." *Journal of Public Economics*, 221: 104863.

Luflade, Margaux. 2018. "The value of information in centralized school choice systems." *Duke University*, 4(6): 7.

Magnet Schools of America. 2019. "A Snapshot of Magnet Schools in America."

Margolis, Jesse, Daniel Dench, and Shirin Hashim. 2020. "The impact of middle school integration efforts on segregation in two New York City districts."

Mezzacappa, Dale. 2021. "Philly overhauls selective admissions policy in bid to be antiracist." *Chalkbeat*.

Mountjoy, Jack, and Brent R Hickman. 2021. "The returns to college (s): Relative value-added and match effects in higher education." National Bureau of Economic Research.

Neilson, Christopher. 2013. "Targeted vouchers, competition among schools, and the academic achievement of poor students." *Job Market Paper*, 1: 62.

Otero, Sebastián, Nano Barahona, and Cauê Dobbin. 2021. "Affirmative action in centralized college admission systems: Evidence from Brazil." *Unpublished manuscript*.

Pathak, Parag A, and Tayfun Sönmez. 2013. "School admissions reform in Chicago and England: Comparing mechanisms by their vulnerability to manipulation." *American Economic Review*, 103(1): 80–106.

Reardon, Sean F, and Stephen W Raudenbush. 2009. "Assumptions of value-added models for estimating school effects." *Education Finance and Policy*, 4(4): 492–519.

Rosenbaum, Paul R, and Donald B Rubin. 1983. "The central role of the propensity score in observational studies for causal effects." *Biometrika*, 70(1): 41–55.

Rothstein, Jesse 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.

Shapiro, Eliza. 2021. "Adams Commits, With Few Details, to Keeping Gifted Program in Schools." *The New York Times*.

Train, Kenneth E. 2009. *Discrete choice methods with simulation.* Cambridge university press.

Walters, Christopher R. 2018. “The demand for effective charter schools.” *Journal of Political Economy*, 126(6): 2179–2223.

Zhu, Pengyu, Yi Zhang, and Juan Wang. 2023. “Canceling the admission priority of private schools enlarges housing price gap in public school districts: Evidence from Shanghai’s new admission policy.” *Real Estate Economics*, 51(1): 49–67.

Online Appendix

A Additional Tables

Table A1: Balance in Observable Characteristics by Lottery Outcomes

	No Initial Offer (1)	Any Initial Offer (2)	Difference (3)
3rd Grade Math Score	3.473	3.381	0.092
	(0.013)	(0.015)	
3rd Grade Reading Score	3.628	3.595	0.033
	(0.010)	(0.012)	
Two-Parent Household	0.851	0.806	0.045
	(0.007)	(0.009)	
College-Educated Parents	0.907	0.887	0.020
	(0.007)	(0.009)	
Parent Survey on Record	0.459	0.477	−0.018
	(0.009)	(0.010)	
Born in Canada	0.764	0.760	0.004
	(0.009)	(0.009)	
Observations	3,201	2,376	

NOTES: The above table displays group means with corresponding standard deviations in parentheses. The first column displays information for those who submit a rank-ordered list but do not receive an offer from any school in the first round. The second column displays the same for those who receive an offer either from their first or second choice in the first round. Raw test scores are reported from Toronto's 0.1-4.9 scale. Parent information comes from the 2017 parent survey. Observations refer to the greatest number available among each of the outcomes. Some outcomes are available for fewer students. For example, the corresponding number of observations with non-missing information on their country of origin is 2,402 and 2,076, respectively.

Table A2: Evidence of Balance from Propensity Scores

	No Controls (1)	Propensity Score Controls (2)
Underrepresented Minority	0.056***	−0.006
	(0.011)	(0.012)
Female	0.133***	0.021
	(0.012)	(0.015)
Lagged Math Score	0.202***	0.015
	(0.023)	(0.029)
Lagged Reading Score	0.204***	0.024
	(0.022)	(0.028)
Two Parent HH	−0.052***	−0.007
	(0.012)	(0.013)
College-Education Parents	0.064***	0.010
	(0.011)	(0.013)
Born in Canada	0.034***	0.009
	(0.010)	(0.013)
Observations	14,481	14,481

NOTES: The above table reports coefficients and robust standard errors (in parentheses) from linear regression models where the stated student characteristic is the dependent variable and the regressor of interest is an indicator for whether the student received a lottery offer. Column (1) presents results from an uncontrolled, univariate specification. Column (2) adds 28 propensity score controls (one for each magnet program). The stated number of observations refers to the number in the URM regression, although certain specifications like parental information have fewer observations due to missing parent survey responses.

Table A3: Group-Specific Decomposition of Conventional Value-Added

	Peer Characteristics			Non-Peer Characteristics			
	Frac. URM (1)	Frac. Female (2)	Frac. High Ability (3)	Art (4)	STEM (5)	IB (6)	Building Condition (7)
Panel A: Ability Groups							
Low Ability	−0.007	−0.003	0.015	0.388	−0.181	−0.202	−0.001
	(0.005)	(0.013)	(0.006)	(0.382)	(0.193)	(0.189)	(0.001)
High Ability	−0.014	0.016	0.015	−0.543	−0.007	−0.181	−0.001
	(0.007)	(0.027)	(0.008)	(1.025)	(0.199)	(0.296)	(0.001)
Panel B: k-Means Clustering							
Group 1	0.051	−0.113	0.075	3.174	−2.218	−0.544	−0.001
	(0.048)	(0.098)	(0.053)	(2.413)	(1.905)	(0.546)	(0.003)
Group 2	−0.013	0.005	0.013	−0.154	0.061	−0.052	−0.001
	(0.007)	(0.017)	(0.008)	(0.614)	(0.179)	(0.183)	(0.001)

NOTES: The table presents decomposition results for value-added (estimated via 2SLS) under alternative group definitions. Panel A reports results for our preferred specification, which splits the sample based on lagged math achievement. Panel B reports results from an alternative specification that uses k-means clustering (with $k = 2$) based on race, gender, and lagged math and reading achievement. Group 1 (2) is 42% (20%) URM, 48% (48%) female, and 4% (82%) high ability. In Panel B, "Frac. High Ability" is redefined to the fraction of group 1 students to be consistent with the k-means clusters. Similarly, we use lagged group 2 shares as our instrument in the decomposition.

Table A4: Decomposition with Additional School Characteristics

	Low Ability Returns		High Ability Returns	
	(1)	(2)	(3)	(4)
Art	0.388	0.420	−0.543	−0.482
	(0.382)	(0.383)	(1.025)	(0.991)
STEM	−0.181	−0.175	−0.007	0.030
	(0.193)	(0.190)	(0.199)	(0.174)
IB	−0.202	−0.182	−0.181	−0.159
	(0.189)	(0.192)	(0.296)	(0.311)
Building Condition	−0.001	−0.002	−0.001	−0.002
	(0.001)	(0.001)	(0.001)	(0.002)
Frac. URM Students	−0.007	−0.008	−0.014	−0.016
	(0.004)	(0.005)	(0.007)	(0.006)
Frac. Female Students	−0.003	−0.004	0.016	0.014
	(0.013)	(0.012)	(0.027)	(0.026)
Frac. High Ability Students	0.015	0.015	0.015	0.013
	(0.006)	(0.007)	(0.008)	(0.007)
Lead Level Risk		0.128		0.085
		(0.135)		(0.154)
Accessibility		−0.029		0.065
		(0.065)		(0.080)
Pool		0.078		0.146
		(0.066)		(0.101)
Observations	92	92	92	92
R^2	0.474	0.490	0.411	0.428

NOTES: Columns (1) and (3) reproduce the estimates from our preferred IV specification. Peer characteristics are demeaned across district. URM refers to "underrepresented minority." Building condition refers to the demeaned number of annual health infractions recorded in the most recent report conducted by the Toronto school district. Accessibility is an indicator for whether the school building is deemed wholly handicapped accessible. Lead Level Risk is an indicator for whether the level of lead tested in the school's water system was deemed hazardous. Pool is an indicator for whether the school campus has an athletic pool on site. Robust standard errors are reported in parentheses.

Table A5: Decomposition Results by Estimation Strategy

	Peer Characteristics			Non-Peer Characteristics			
	Frac. URM (1)	Frac. Female (2)	Frac. High Ability (3)	Art (4)	STEM (5)	IB (6)	Building Condition (7)
Panel A: OLS							
Low Ability	−0.009	−0.000	0.016	0.305	−0.200	−0.251	−0.001
	(0.004)	(0.005)	(0.004)	(0.205)	(0.169)	(0.147)	(0.001)
High Ability	−0.014	0.014	0.017	−0.476	−0.037	−0.209	−0.001
	(0.005)	(0.014)	(0.005)	(0.678)	(0.148)	(0.183)	(0.001)
Panel B: Lagged Peer IV							
Low Ability	−0.007	−0.003	0.015	0.388	−0.181	−0.202	−0.001
	(0.005)	(0.013)	(0.006)	(0.382)	(0.193)	(0.189)	(0.001)
High Ability	−0.014	0.016	0.015	−0.543	−0.007	−0.181	−0.001
	(0.007)	(0.027)	(0.008)	(1.025)	(0.199)	(0.296)	(0.001)
Panel C: Nearby School IV							
Low Ability	−0.020	0.015	0.030	−0.073	−0.406	−0.684	−0.001
	(0.013)	(0.022)	(0.015)	(0.599)	(0.320)	(0.463)	(0.001)
High Ability	−0.022	0.050	0.038	−1.425	−0.407	−0.926	−0.001
	(0.016)	(0.033)	(0.021)	(0.995)	(0.416)	(0.612)	(0.002)

Notes: Student demographic characteristics are demeaned across district. URM refers to "underrepresented minority." Building condition refers to the demeaned number of annual safety and health infractions recorded in the most recent report conducted by the Toronto school district. Panel A uses current-year peer composition and non-peer covariates in an ordinary least squares specification. Panel B instruments current-year peer composition with the same school's prior-year peer composition. Panel C uses BLP-style rival instruments: inverse-distance–weighted averages of nearby schools' program flags and health measure within a 1-10 mile ring. Robust standard errors are reported in parentheses.

Table A6: Naive Differencing of Value-Added Between Years

	All Schools	Traditional HS	Magnet Program Type		
			Art	STEM	IB
	(1)	(2)	(3)	(4)	(5)
Low Ability	−0.000	0.172	0.149	−0.726	−0.484
	(0.053)	(0.033)	(0.144)	(0.207)	(0.112)
High Ability	0.000	0.111	−0.258	−0.284	−0.180
	(0.056)	(0.054)	(0.371)	(0.110)	(0.108)

NOTES: The above table displays the relative change in value-added from the achievement-based policy to the achievement-blind policy, calculated as $\theta^k_{2024} - \theta^k_{2023}$ for θ^k_t the estimated return for ability type k in year t. We use the IV estimates for the achievement-blind regime and conventional OLS estimates recovered using the achievement-based enrollment data for the achievement-based measure. Each cell is measured in standard deviations of 9th grade math test score achievement (σ). Standard errors in parentheses are derived from paired t-tests.

Table A7: OLS Covariate Estimates for 9th Grade Math Scores

	Low Ability (1)	High Ability (2)
Lagged Math	0.325	0.456
	(0.018)	(0.025)
Lagged Reading	0.024	−0.048
	(0.013)	(0.014)
URM	−0.156	−0.297
	(0.022)	(0.029)
Female	−0.097	−0.059
	(0.020)	(0.018)
Observations	6,313	8,122

Notes: Each estimate comes from a value-added model without a constant. Lagged test scores are standardized across district and refer to 3rd grade achievement levels. Columns (1) and (2) display results from the propensity-adjusted model that includes 28 propensity score controls and 28 zero-probability indicators. The first (second) column restricts to those classified as low (high) ability. Peer characteristics are demeaned across district. URM refers to "underrepresented minority." Building condition refers to the demeaned number of annual health infractions recorded in the most recent report conducted by the Toronto school district. Columns (3) and (4) report the same information for regressions that omit propensity scores and applicants' default school. Standard errors clustered at the school level are reported in parentheses.

Table A8: Alternative Decomposition Specifications

	Low Ability Returns			High Ability Returns		
	(1)	(2)	(3)	(4)	(5)	(6)
Art	0.290			−0.142		
	(0.156)			(0.442)		
STEM	0.042			0.216		
	(0.173)			(0.105)		
IB	0.117			0.291		
	(0.094)			(0.138)		
Building Condition	−0.001			−0.001		
	(0.002)			(0.001)		
Frac. URM Students		−0.010	−0.011		−0.009	−0.005
		(0.005)	(0.007)		(0.007)	(0.030)
Frac. Female Students		0.006	0.005		0.002	−0.019
		(0.004)	(0.011)		(0.005)	(0.033)
Frac. High Ability Students		0.010	0.010		0.017	0.036
		(0.005)	(0.009)		(0.005)	(0.029)
URM × High			0.0003			−0.002
			(0.0008)			(0.002)
URM × Female			−0.001			−0.004
			(0.001)			(0.006)
High × Female			0.0004			0.005
			(0.0019)			(0.005)
URM × High × Female			−0.0000			−0.001
			(0.0003)			(0.001)
Observations	92	92	92	92	92	92

NOTES: Peer characteristics are demeaned across district. URM refers to "underrepresented minority." Building condition refers to the demeaned number of annual health infractions recorded in the most recent report conducted by the Toronto school district. Coefficients in columns (1) and (4) are derived from simple OLS regressions. Coefficients in all other columns are obtained from a decomposition that instruments for each peer effect with the corresponding lagged peer composition. Robust standard errors are reported in parentheses.

Table A9: Correlations Between Student Characteristics and Magnet Attributes

	Frac. URM (1)	Frac. Female (2)	Frac. High Ability (3)	Art (4)	STEM (5)	IB (6)
URM	0.050	0.015	0.035	0.055	0.090	0.021
	(0.009)	(0.011)	(0.023)	(0.029)	(0.026)	(0.019)
Female	0.005	0.045	−0.001	0.120	−0.040	−0.008
	(0.007)	(0.015)	(0.010)	(0.043)	(0.018)	(0.011)
High Ability	0.004	−0.038	0.283	−0.011	0.090	0.062
	(0.020)	(0.007)	(0.131)	(0.068)	(0.048)	(0.034)
Observations	5,605	5,605	5,605	5,605	5,605	5,605

NOTES: Each column reports coefficients from a separate regression of the listed magnet attribute on student characteristics among the set of applicants during the achievement-blind policy. Rows correspond to URM, female, and high ability indicators. Standard errors clustered at the magnet level are reported in parentheses.

Table A10: Main Sources of Identification for Model Parameters

Model Parameter Group	Main Source of Identifying Variation
Student Preferences	Fraction of applicants with first-choice of type k
	Fraction of applicants with second-choice of type k
	Fractional transition of applications with $\{k, \ell\}$ ROL type
	OLS coefficients from program-type, rank-specific regressions:
	$\text{Rank}_i^{r,k} = \alpha_0^{r,k} + \alpha_1^{r,k} x_i^f + \alpha_2^{r,k} x_i^{\text{URM}} + \alpha_3^{r,k} x_i^H + \epsilon_i^{r,k}$
Application Components	Fraction of demographic group X that enter application pool
	Linear probability model coefficients for offer acceptance:
	$\text{Accept}_i = \beta_0 + \beta_1 x_i^f + \beta_2 x_i^{\text{URM}} + \beta_3 x_i^H + \vartheta_i$
Enrollment	m_j, the fraction of some characteristic of students at school $j \in J$
Program Preferences	Average characteristic X of enrolled students in program type k

NOTES: This table displays the generalized moments used to identify the corresponding parameters. We use 318 moments to identify 42 demand-side parameters and 303 moments to identify the 12 supply-side parameters. In the above table, X is a placeholder that typically refers to some or all of the following: all students, URM students, female students, high ability students, and lagged test scores. The x_i terms are defined as in the main text. In the above, k denotes program type and $r \in \{1, 2\}$ refers to the program rank on the ROL.

Table A11: Model Fit—Targeted Achievement-Blind Application Moments

	Data (1)	Model (2)
Fraction Applying		
Female	0.55	0.53
URM	0.23	0.30
High Ability	0.53	0.55
Overall	0.33	0.52
Fraction of Applicants Ranking X First		
Art	0.30	0.21
STEM	0.54	0.61
IB	0.16	0.18
Fraction of Applicants Ranking X Second		
Art	0.16	0.10
STEM	0.33	0.38
IB	0.13	0.17
No Second Choice	0.38	0.35
Second-Choice Program if First-Choice Art		
Art	0.33	0.02
STEM	0.12	0.32
IB	0.05	0.11
No Second Choice	0.51	0.55
Second-Choice Program if First-Choice STEM		
Art	0.09	0.12
STEM	0.43	0.36
IB	0.15	0.23
No Second Choice	0.33	0.30
Second-Choice Program if First-Choice IB		
Art	0.07	0.11
STEM	0.41	0.55
IB	0.19	0.05
No Second Choice	0.33	0.29

NOTES: This table compares the stated moments related to *application decisions* in the administrative data (column 1) with those simulated using our model parameters (column 2).

Table A12: Model Fit—Achievement-Blind Offer Moments

	Data (1)	Model (2)
Enrollment Rate		
Default High School	0.81	0.77
Outside Option	0.06	0.04
Compliance Rate		
Female	0.71	0.60
URM	0.72	0.56
High Ability	0.79	0.58

NOTES: This table compares the stated moments related to *enrollment decisions* in the administrative data (column 1) with those simulated using our model parameters (column 2). In addition to those presented in this table and Table A8, our simulated method of moments (SMM) procedure for student preferences matches OLS coefficients of the propensity of each demographic type to rank a given program first or second. We suppress these 18 moments for ease of visualization, but they are available upon request.

Table A13: Model Fit—Achievement-Based Enrollment Moments

	Data (1)	Model (2)
Fraction Magnet-Enrolled in Art	0.45	0.44
Fraction Magnet-Enrolled in STEM	0.33	0.36
Fraction Magnet-Enrolled in IB	0.22	0.20
Fraction URM Students (Art)	0.16	0.04
Fraction URM Students (STEM)	0.18	0.15
Fraction URM Students (IB)	0.12	0.04
Fraction Female Students (Art)	0.68	0.59
Fraction Female Students (STEM)	0.46	0.30
Fraction Female Students (IB)	0.59	0.54
Fraction High Ability Students (Art)	0.52	0.50
Fraction High Ability Students (STEM)	0.73	0.63
Fraction High Ability Students (IB)	0.70	0.76
Median Std. Lagged Math Score (Art)	0.71	0.69
Median Std. Lagged Math Score (STEM)	0.77	0.76
Median Std. Lagged Math Score (IB)	0.75	0.77
Median Std. Lagged Reading Score (Art)	0.73	0.73
Median Std. Lagged Reading Score (STEM)	0.79	0.75
Median Std. Lagged Reading Score (IB)	0.77	0.78

NOTES: This table compares the moments in the achievement-based enrollment data to those in our simulation. Lagged scores in the model are readjusted to belong to [0,1]. In addition to the above moments, we target overall enrollment patterns into magnet programs by program type.

Table A14: Estimated Effects of Policy Change on Student Utility

	Mean (1)	Percentage Change (2)
All students	−11.21	−1.70%
	[−12.85, −9.20]	[−1.95%, −1.39%]
URM	−25.53	−2.72%
	[−29.42, −20.97]	[−3.13%, −2.23%]
Non-URM	−5.94	−1.07%
	[−7.72, −3.99]	[−1.39%, −0.72%]
Female	−9.71	−1.18%
	[−12.19, −6.93]	[−1.48%, −0.84%]
Male	−12.64	−2.51%
	[−14.58, −10.63]	[−2.90%, −2.10%]
High Ability	−1.82	−0.26%
	[−4.40, 0.99]	[−0.62%, 0.14%]
Low Ability	−21.02	−3.47%
	[−23.32, −18.12]	[−3.84%, −2.99%]
URM; High Ability	5.94	0.58%
	[−1.60, 13.81]	[−0.16%, 1.35%]
URM; Low Ability	−44.32	−4.97%
	[−47.62, −40.45]	[−5.33%, −4.54%]
Non-URM; High Ability	−3.72	−0.59%
	[−6.42, −0.79]	[−1.01%, −0.12%]
Non-URM; Low Ability	−8.78	−1.92%
	[−11.90, −5.58]	[−2.61%, −1.22%]
Pre-Reform Admits	−382.31	−39.69%
	[−406.88, −360.97]	[−41.41%, −38.17%]
Pre-Reform Non-Admits	11.31	1.76%
	[8.95, 14.64]	[1.40%, 2.28%]
Post-Reform Female Admits	289.42	31.11%
	[276.20, 305.51]	[29.49%, 33.04%]
Post-Reform URM Admits	287.46	26.15%
	[269.93, 310.59]	[24.50%, 28.27%]

NOTES: Entries report the distribution of $u^1 - u^0$ within each group, where u^1 is the average utility under the achievement-blind policy and u^0 is the average utility under the achievement-based policy. Bracketed percentages beneath each row report the same statistics as a share of the group's mean achievement-based utility. "Pre-Reform" refers to the achievement-blind regime. "Pre-Reform Admits" are the counterfactual magnet admits, and "Pre-Reform Non-Admits" are the students who do not enroll in a magnet program in the counterfactual.

Table A15: Policy Effects Under Iterative Elimination of Model Mechanisms

	(1)	(2)	(3)	(4)
Δ Avg. Utility	12.81	19.49	36.61	31.71
	[11.34, 15.06]	[18.04, 20.93]	[34.25, 39.81]	[29.81, 33.24]
Endogenous Peer Effects	✓	✗	✓	✗
Strategic Behavior	✓	✓	✗	✗

NOTES: The above table displays the estimated changes in welfare using the achievement-blind administrative data and the set of estimated preference parameters. The first column reproduces our main results. The second column eliminates endogenous peer effects in achievement-based utility by not running the fixed-point algorithm for peers and value-added. The third column shuts off students' strategic behavior. In the model, this is equivalent to establishing all probabilities of admission to one. The fourth column combines these two iterations of the model. 5th and 95th percentiles of the average utility differences across 100 simulations are presented in brackets.

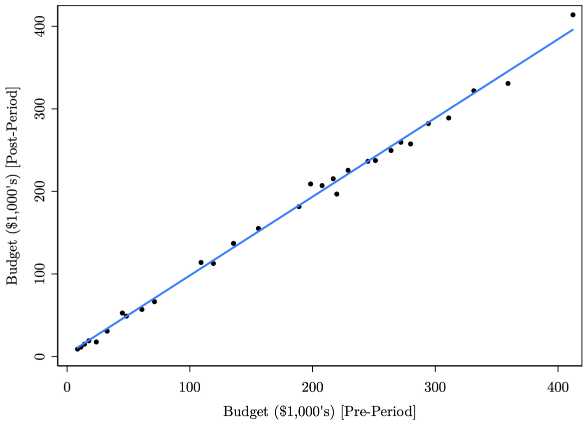
Table A16: Optimal Policy Design by Baseline Policy and Equity-Efficiency Combination

	Optimal Equity Weight, ω^*	Optimal Policy Design, $\boldsymbol{\Gamma}^*(\tau)$		
Panel A: Achievement-Based				
		Baseline	URM Status	Female Status
School Representation vs. Utility	1	$\{1.45, 1.20, 1.97\}$	$\{0.33, 0.04, 0.001\}$	$\{-1.13, -0.37, 0.05\}$
Program Representation vs. Utility	1	$\{0.78, 1.11, 0.64\}$	$\{0.81, 0.19, 0.07\}$	$\{-0.03, -0.12, 0.25\}$
Achievement Gaps vs. Utility	1	$\{0.78, 1.25, 1.16\}$	$\{-1.99, 0.18, 0.18\}$	$\{0.85, 0.36, 0.57\}$
School Representation vs. Program VA	0	$\{0.96, 1.17, 0.87\}$	$\{1.33, -0.12, -0.08\}$	$\{-0.57, -0.19, 0.21\}$
Program Representation vs. Program VA	1	$\{0.78, 1.11, 0.64\}$	$\{0.81, 0.19, 0.07\}$	$\{-0.03, -0.12, 0.25\}$
Achievement Gaps vs. Program VA	1	$\{0.78, 1.25, 1.16\}$	$\{-1.99, 0.18, 0.18\}$	$\{0.85, 0.36, 0.57\}$
School Representation vs. Match Quality	0	$\{-0.60, 0.82, 0.80\}$	$\{0.27, -0.84, 0.35\}$	$\{0.40, 0.94, 1.14\}$
Program Representation vs. Match Quality	0	$\{-0.60, 0.82, 0.80\}$	$\{0.27, -0.84, 0.35\}$	$\{0.40, 0.94, 1.14\}$
Achievement Gaps vs. Match Quality	0	$\{-0.60, 0.82, 0.80\}$	$\{0.27, -0.84, 0.35\}$	$\{0.40, 0.94, 1.14\}$
Panel B: Achievement-Blind				
		URM Quota	Female Quota	
School Representation vs. Utility	1	0	0.7	
Program Representation vs. Utility	1	0.2	0.8	
Achievement Gaps vs. Utility	1	0.6	0.2	
School Representation vs. Program VA	1	0	0.7	
Program Representation vs. Program VA	1	0.2	0.8	
Achievement Gaps vs. Program VA	1	0.6	0.2	
School Representation vs. Match Quality	0	0	1	
Program Representation vs. Match Quality	0	0	1	
Achievement Gaps vs. Match Quality	0	0	1	

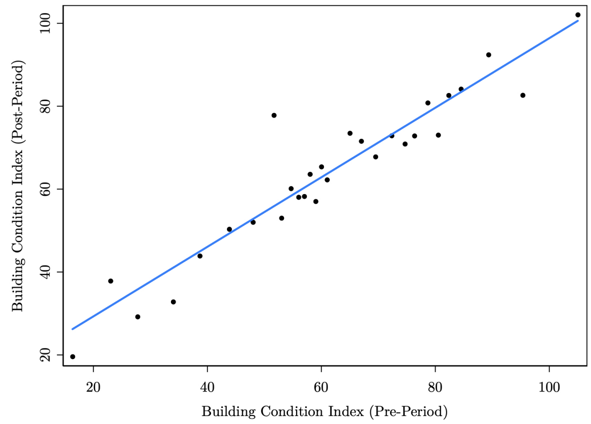
NOTES: The above table displays the optimal admissions policy design for a given combination of equity and efficiency and for a given baseline policy. Panel A displays the decentralized, program-specific admissions weights that optimize the stated objective. The weights correspond to $\{\text{Art}, \text{STEM}, \text{IB}\}$. Panel B displays the centralized lottery quotas that optimize the stated objective. URM quotas are applied to all magnet programs while female quotas are applied to STEM programs only.

B Additional Figures

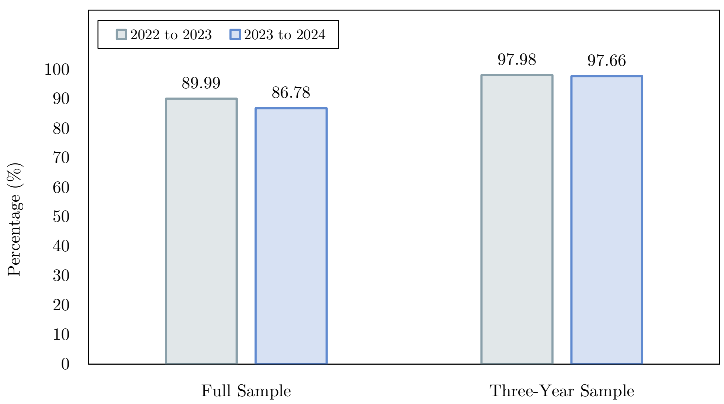
Figure B1: Relative Stability of Education Inputs



(a) Annual Preliminary Budget



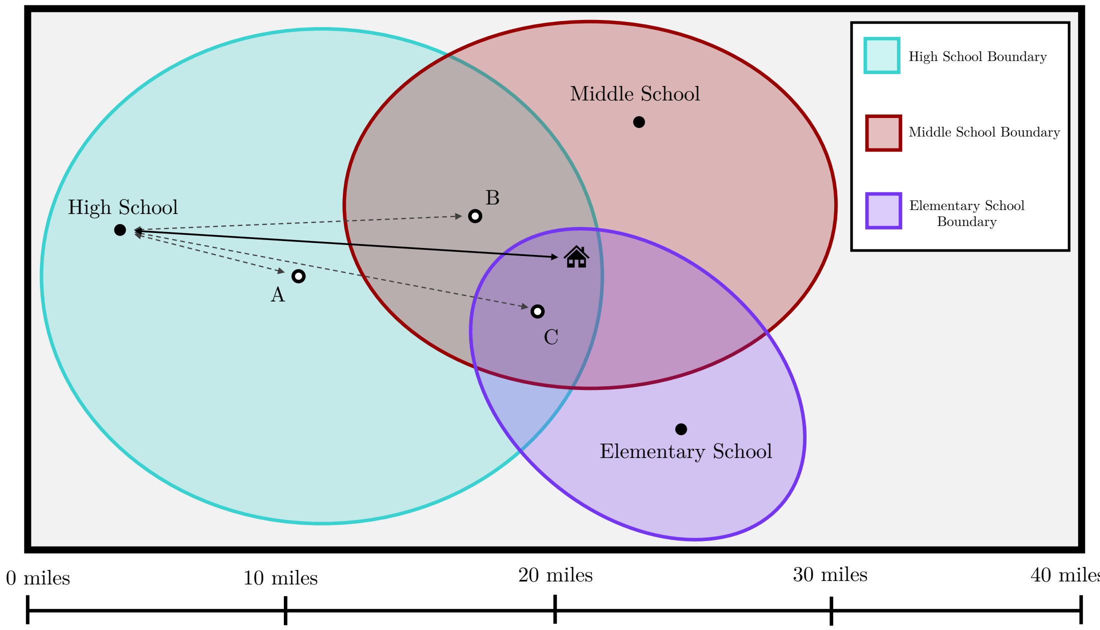
(b) Building Condition Index



(c) Teachers (Staying Rate)

Notes: The above figures display evidence of stability in education inputs across admissions policies. The first two panels show binscatter plots with corresponding lines of best-fit for the listed school characteristic in the pre-period (x-axis) and post-period (y-axis). Figure B12a displays these results for the annual school budget designated at the start of the academic year. Figure B12b displays the same for the district-assigned index for the school's building condition. The final panel (Figure B1c) shows the implicit rates of teacher turnover at magnet programs in the district. The full sample includes all teachers that work at a magnet program at any point during the three-year period (2022, 2023, and 2024). The "Three-Year Sample" restricts to those with non-missing entries for all three years. The percentages displayed refer to the proportion of the relevant sample that stays at the same school from one year to the next.

Figure B2: Illustrative Example of Home Address Imputation



Notes: The above figure provides a stylized example of how we impute students' home addresses. The students' true (unobserved) address is denoted by the house icon. The student attends the high school shown, whose latitude and longitude are given by the black dot in the green oval. Point A represents the imputed home address if only the default high school were available in data. Point B shows the imputed address if both high school and middle school assignment were available. Finally, Point C is the imputed address if all levels of schooling were available in data. Dotted lines represent crude estimates of distance traveled between "home" and school. The solid line captures the true distance. These distances are for illustrative purposes only. Distances used in the paper are generated on real travel routes using Google API.

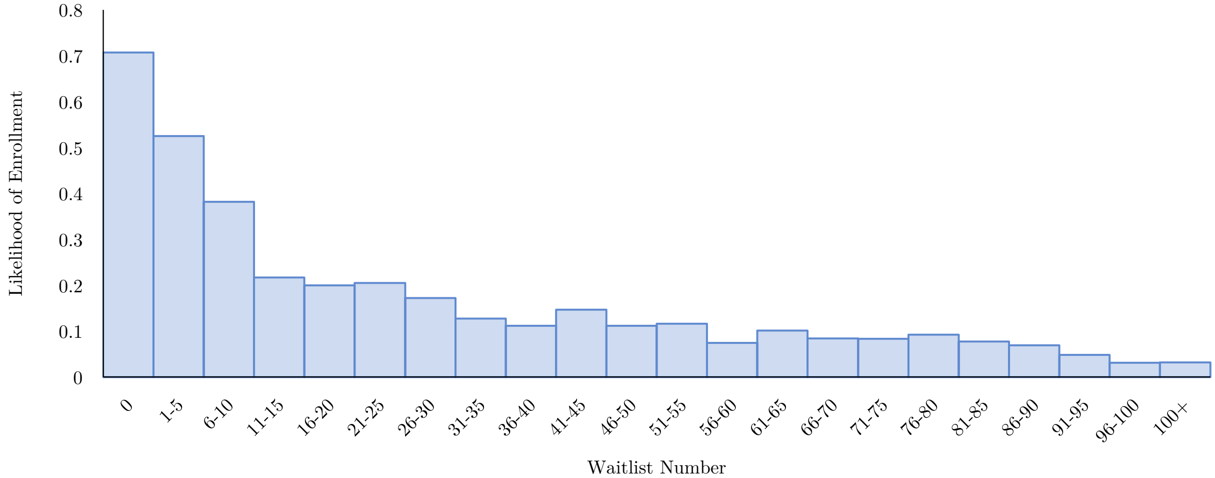
Figure B3: Distribution of Imputed Home Locations

(a) 2022-2023 School Year

(b) 2023–2024 School Year

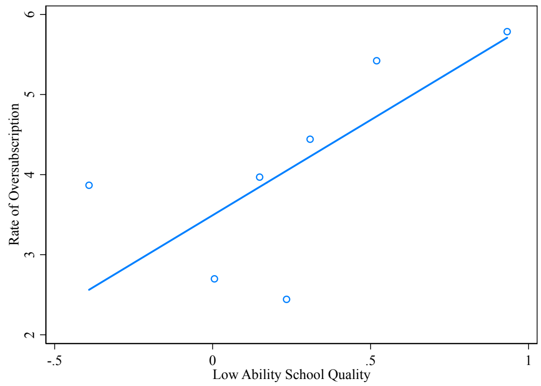
Notes: The above figures display the home locations of 9th graders in the district, which we impute using information on default high school, middle school, and elementary school assignment, along with each school's corresponding catchment area. Out-of-district students are imputed as living on the border of Toronto nearest their enrolled school (in the case of enrolling in Toronto) or highest ranked magnet program (in the case of losing admissions to Toronto public schools). When students rank two programs, we establish their residency as the midpoint on the border between those two schools. Panel (a) displays this distribution for the universe of 9th graders in the 2022-2023 school year (final year of achievement-based admissions) and Panel (b) displays the same for 9th graders in the 2023–2024 school year (first year of achievement-blind admissions).

Figure B4: Likelihood to Enroll in First-Choice Magnet Program Conditional on Waitlist Number

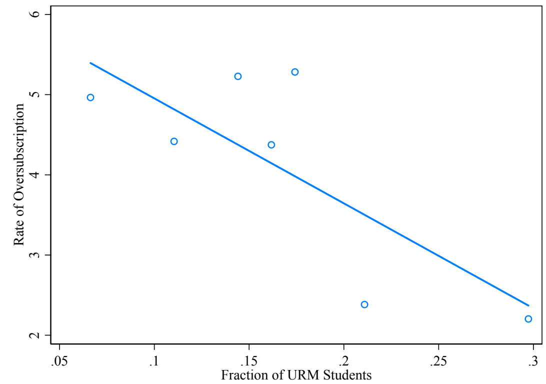


Notes: The above figure plots the mean probability that a student enrolls in their top-ranked magnet program, conditional on receiving the stated waitlist number. Here, 0 refers to receiving an offer (not being waitlisted). 100+ includes all students that receive a waitlist of 101 or greater. In the data, the largest waitlist number is 708. We include only those who rank at least one program to construct the figure.

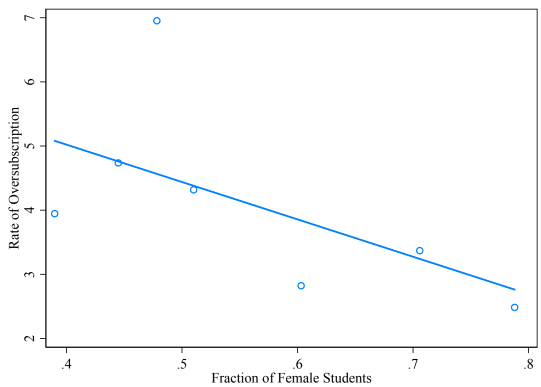
Figure B5: Program Demand and Endogenous School Characteristics



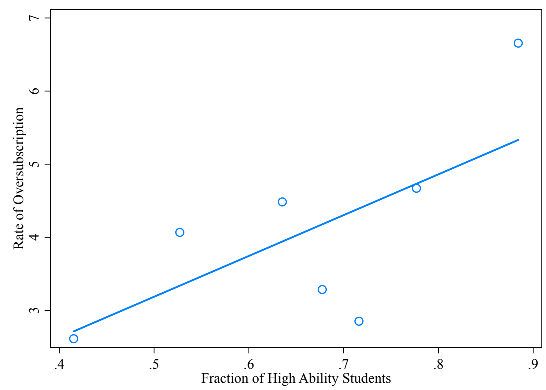
(a) Value-Added



(b) URM Students



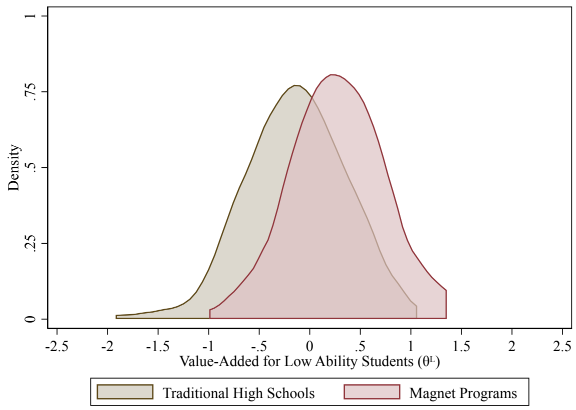
(c) Female Students



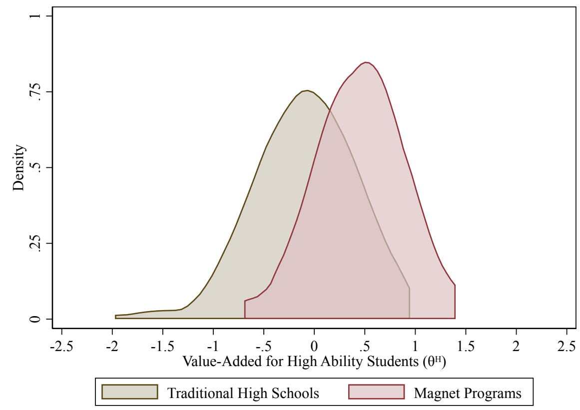
(d) High Ability Students

Notes: The above figures display binscatters between 2022-2023 school-level attributes and 2023–2024 oversubscription levels, measured as the total number of submitted applications divided by the total number of seats available at each program. Panel A displays a positive relationship between demand and school VA. Panel B (C) displays a negative relationship between demand and the relative fraction of URM (female) students. Finally, Panel D displays a positive relationship between overall magnet program demand and the fraction of high ability students, as measured by lagged math scores.

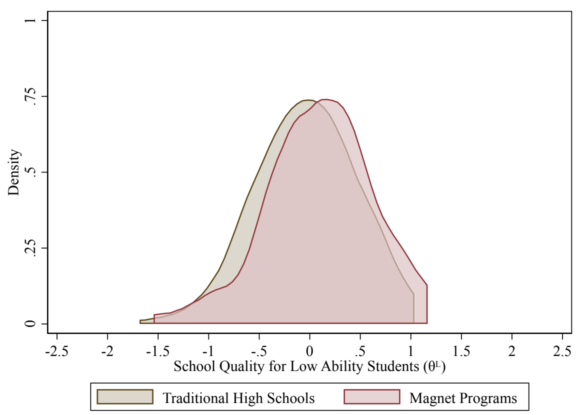
Figure B6: Change in the Distribution of Value-Added, by School Type



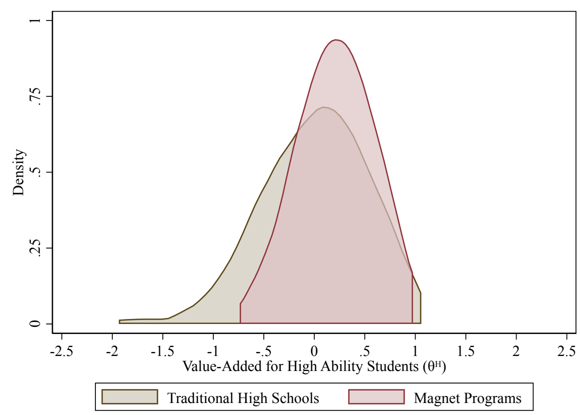
(a) Low Ability VA, Achievement-Based



(b) High Ability VA, Achievement-Based



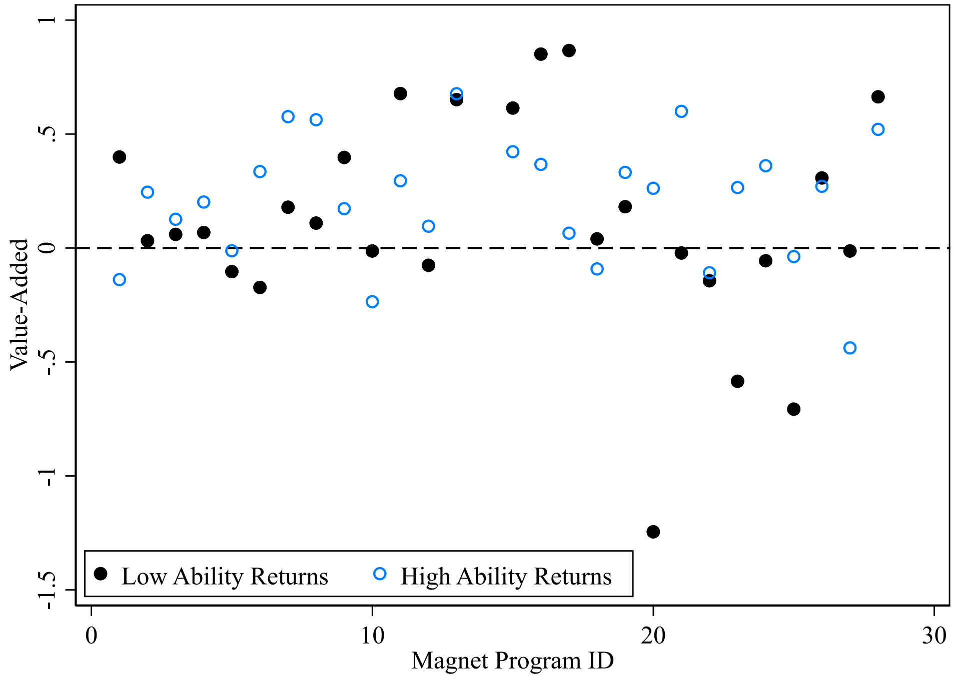
(c) Low Ability VA, Achievement-Blind



(d) High Ability VA, Achievement-Blind

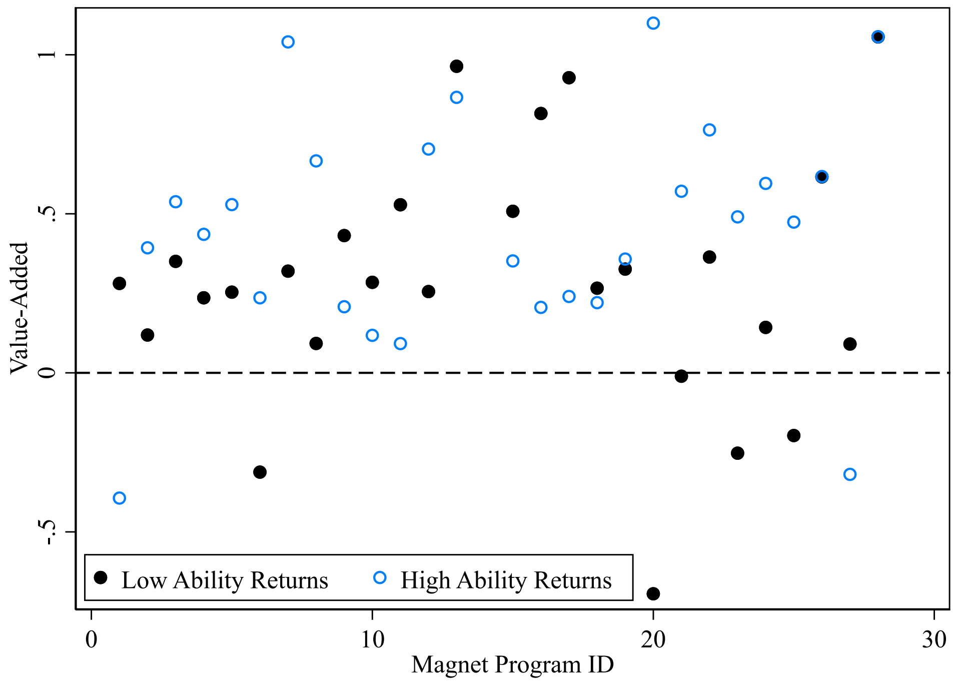
Notes: The above figures display the change in the distribution of estimated value-added between the final year of the achievement-based admissions policy and the first year of the achievement-blind one. Panels (a) and (b) display the distribution of low and high ability VA in the 2022-2023 academic year while (c) and (d) display the same for the 2023–2024 academic year. Value-added is measured in standard deviations of standardized 9th grade math test scores (σ).

Figure B7: Achievement-Blind Magnet Program Value-Added by Ability Type



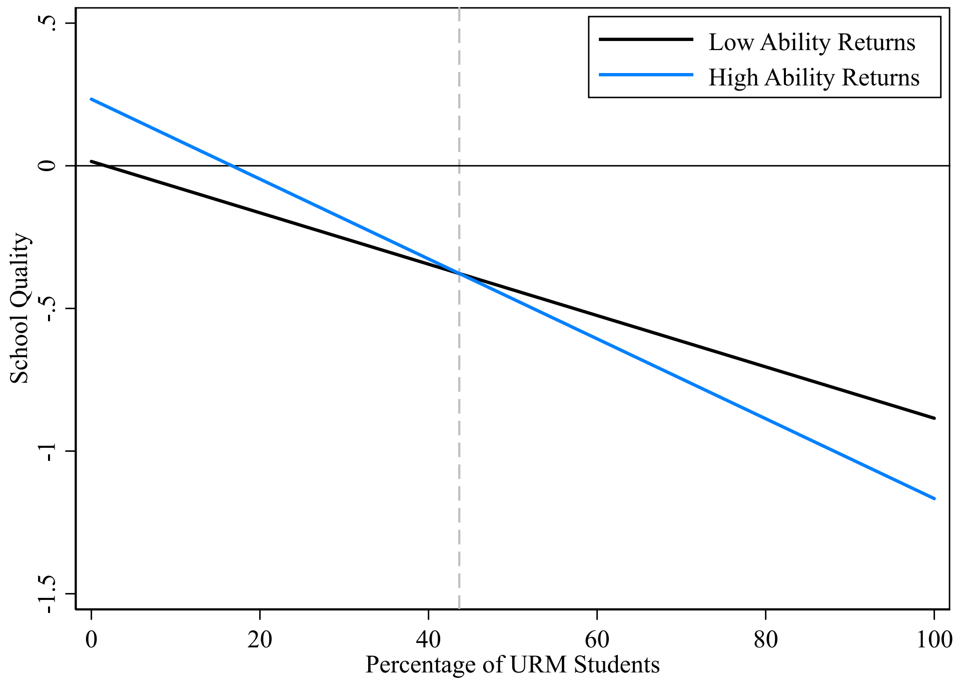
NOTES: The above figure shows a two-way scatterplot of the low and high ability VA measures for 27 magnet programs in 2023–2024. VA estimates are derived from 2SLS regressions that control for lagged test scores, race, gender, and assignment propensity scores.

Figure B8: Achievement-Based Magnet Program Value-Added by Ability Type



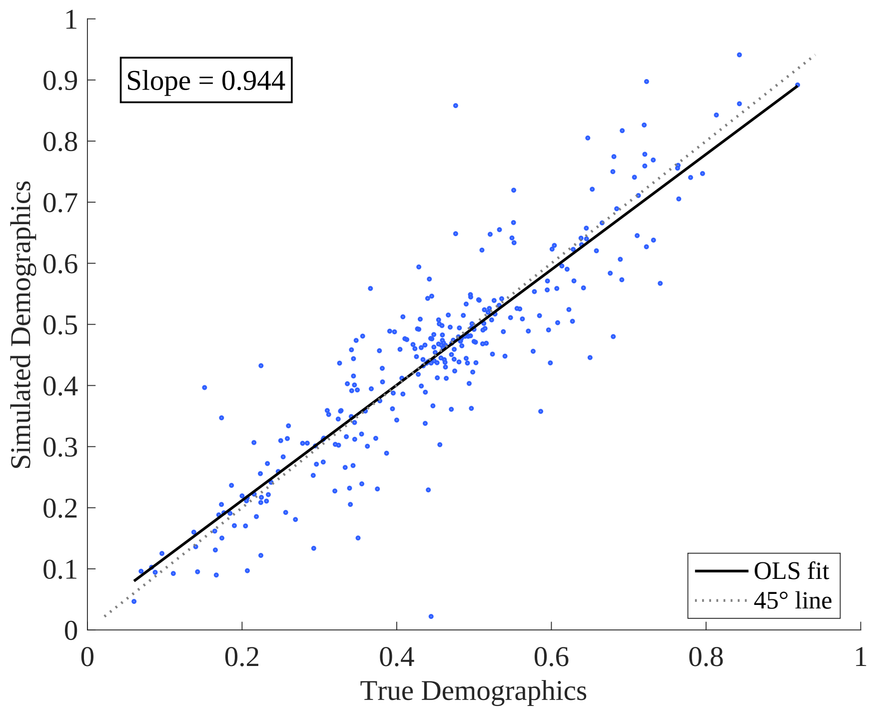
NOTES: The above figure shows a two-way scatterplot of the low and high ability VA measures for 27 magnet programs in 2022-2023. VA estimates are derived from propensity-adjusted OLS regressions that control for lagged test scores, race, gender, and assignment propensity scores.

Figure B9: Endogeneity of Match Quality



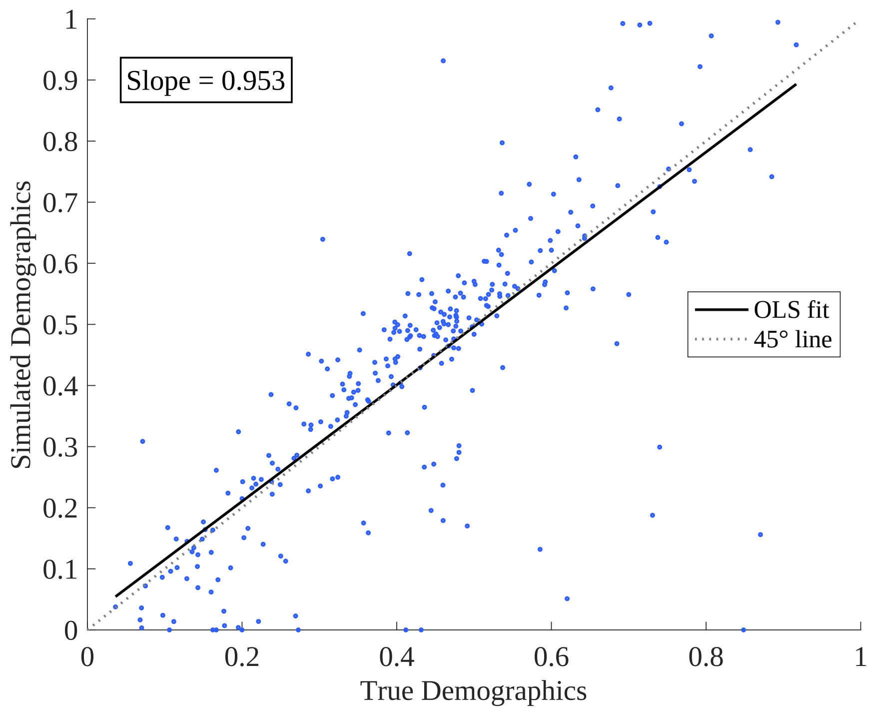
Notes: The above figure plots the value-added for low and high ability students of one anonymous magnet program in Toronto for a given percentage of URM students in the school. We use the coefficients obtained from our decomposition exercise and hold fixed all non-URM attributes of the corresponding production function. The vertical, dotted line plots the crossing point when the program has higher returns for low ability students (approximately 44%). Note that in data, the fraction of URM students in magnet programs ranges from 16.5% to 49%, meaning this illustrative exercise is meaningful and relevant.

Figure B10: Evidence of Convergence to Equilibrium Peer Characteristics (Achievement-Blind)



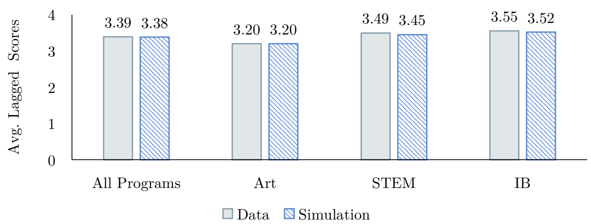
NOTES: The above compares the true peer characteristics (fraction female, URM, and high ability) at the school level to the simulated levels averaged across 100 simulations. Each circle plots this coordinate across the 92 schools in Toronto and for each of the three school-level peer characteristics. The dotted line is 45 degrees from the origin. The black line is a fitted OLS line. We recover a statistically significant coefficient of 0.94 (and intercept of 0.02), indicating close replication of the equilibrium levels observed in data.

Figure B11: Evidence of Convergence to Equilibrium Peer Characteristics (Achievement-Based)

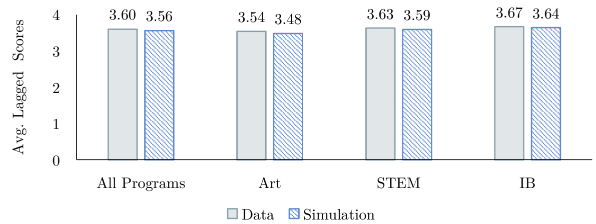


NOTES: The above compares the true peer characteristics (fraction female, URM, and high ability) at the school level to the simulated levels averaged across 100 simulations. Each circle plots this coordinate across the 92 schools in Toronto and for each of the three school-level peer characteristics. The dotted line is 45 degrees from the origin. The black line is a fitted OLS line. We recover a statistically significant coefficient of 0.95 (and intercept of 0.02), indicating close replication of the equilibrium levels observed in data.

Figure B12: Comparison of Untargeted Moments in Data and Simulation



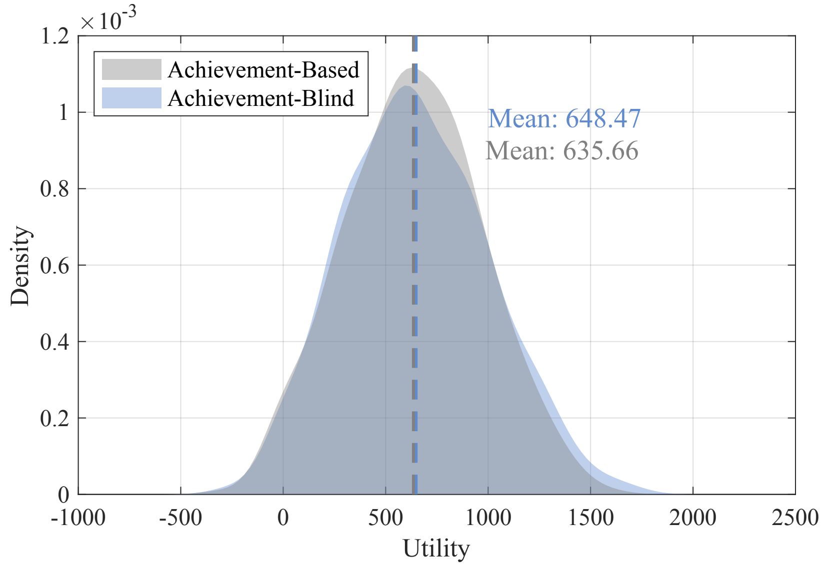
(a) Lagged Math Scores by Enrollment



(b) Lagged Reading Scores by Enrollment

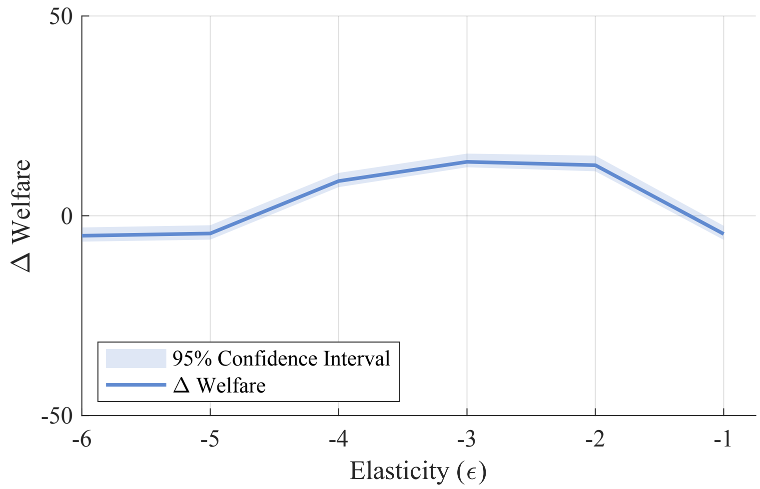
Notes: The above figure plots the average level of lagged math (Panel (a)) and reading (Panel (b)) in magnet programs overall and by magnet program type. Gray bars plot the means calculated in administrative data, whereas the slashed blue bars plot the simulated values.

Figure B13: Policy Effect on the Distribution of Student Utility



NOTES: The above figure plots the distribution of district-wide student utility between the two studied policies. Utility under the original policy (achievement-based) is shown in gray. Utility under the new policy (achievement-blind) is shown in blue. We use data from the achievement-blind regime for both. This strategy eliminates spurious effects driven by small changes in the demographic composition of students across Toronto from one year to the next.

Figure B14: Policy Impact on Average Welfare as a Function of Student Elasticity



NOTES: The above figure plots the average difference in utility across the two policies. We fix students' own-price elasticity to recover the reduction in application costs at each type of magnet program. These differences in costs generate differences in welfare gains. 95% confidence intervals are derived from two-sample t-tests of the difference in welfare between policies.

Discussion of Figure B14

For ease of notation, let $\kappa(0)$ denote the application cost under the original policy and $\kappa(1)$ the new cost. Recall that we impose a linear relationship between costs such that $\kappa(0) = \alpha\kappa(1)$. We assume in the main text that $\epsilon_{A,\kappa} = -2$. Using the formula for elasticity, for art programs (which experienced an average increase in demand of 62%), this translates to

$$-2 = \frac{\%\Delta A}{\%\Delta\kappa} \implies \%\Delta\kappa = -31\% \implies \frac{\kappa(1) - \alpha\kappa(1)}{\alpha\kappa(1)} = -0.31.$$

Rearranging, we find that $\alpha = 1.45$. This implies that art programs costs were 45% higher under the original policy. For STEM programs, this elasticity implies that costs were 29% higher before. For IB programs, the costs were 30% higher. We repeat this procedure to adjust costs under various specifications of $\epsilon_{A,\kappa}$.

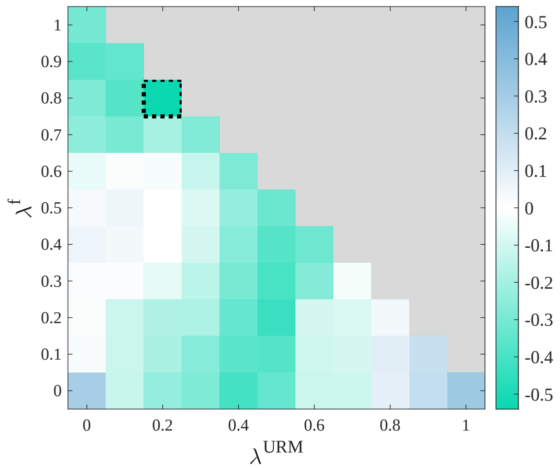
Note too that the observed change in demand places bounds on elasticity. In particular, we know that the largest change in demand was among art programs (+62%) and that the change in costs has to be non-positive, i.e., $\%\Delta\kappa \leq 0$. However, the cost cannot decrease by more than 100%, meaning $-1 \leq \Delta\kappa \leq 0$. Rearranging the elasticity formula and dividing by 100% gives us

$$\Delta\kappa = \frac{0.62}{\epsilon_{A,\kappa}} \implies -1 \leq \frac{0.62}{\epsilon_{A,\kappa}} \leq 0 \implies -\infty \leq \epsilon_{A,\kappa} \leq -0.62.$$

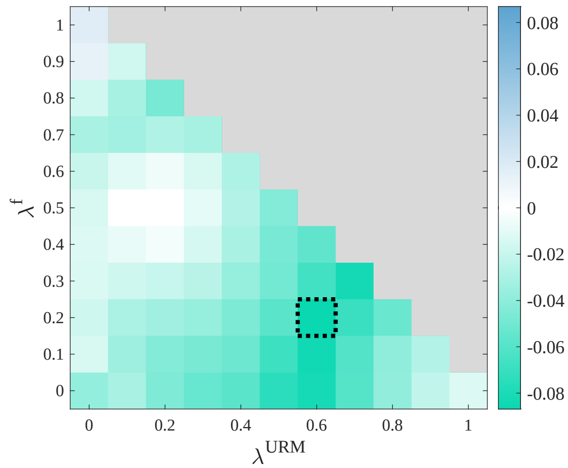
The other programs' changes in demand imply elasticity bounds that are less restrictive than the above.

Students' preferences help pin down further restrictions on ϵ. For instance, setting $\epsilon = -1$ cannot reproduce trends in the data, as far too few students apply to and enroll in art programs. This is due to the fact that unit elasticity in our setting would imply that art programs' costs were 2.63 times the current cost originally. For this reason, we set $\epsilon_{A,\kappa} = -2$ as a conservative estimate. This value is able to replicate the data.

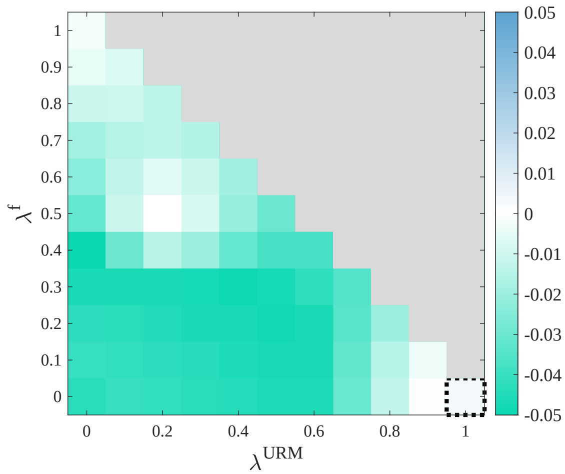
Figure B15: Additional Optimal Lottery Quotas



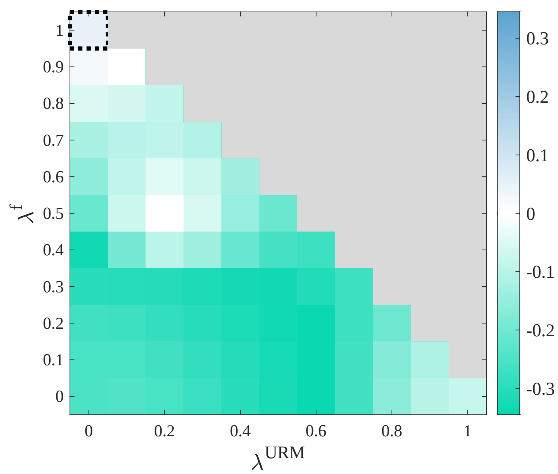
(a) Program-Level Segregation



(b) Achievement Gaps



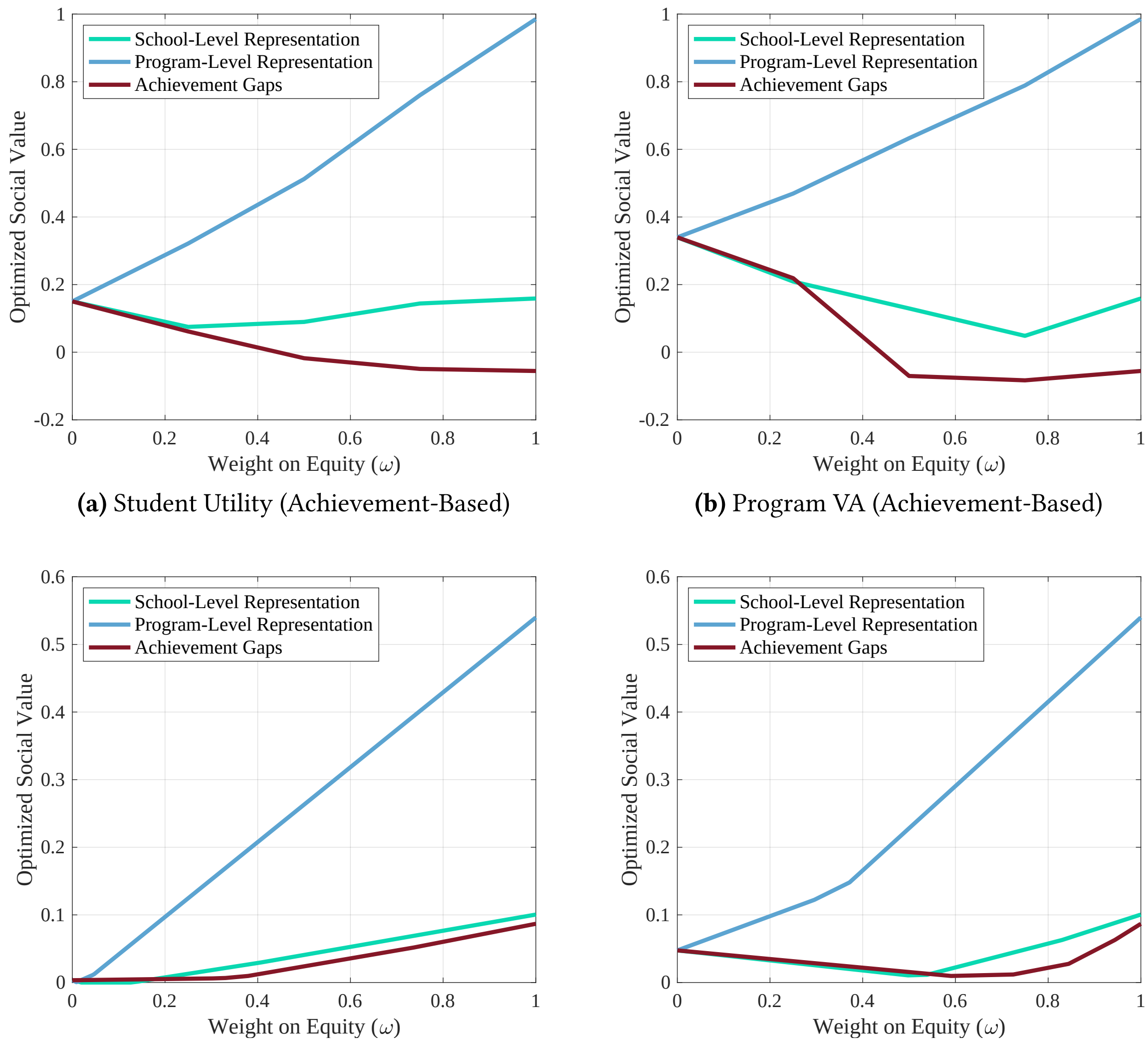
(c) Student Utility



(d) Program VA

NOTES: The above heat maps display the percent change in the stated outcome, relative to the simulated outcome with a URM quota of 0.2 and a female STEM quota of 0.5. Each cell displays the percent change for a given combination of quotas. We plot *positive* percent changes in blue and *negative* percent changes in green. The solution to the district's problem is the cell contained by the black dotted line.

Figure B16: Additional Trade-Off Frontiers



(a) Student Utility (Achievement-Based)

(b) Program VA (Achievement-Based)

(c) Student Utility (Achievement-Blind)

(d) Program VA (Achievement-Blind)

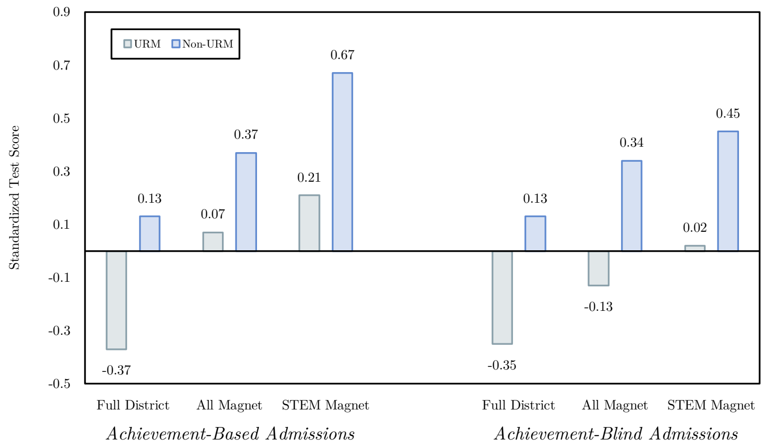
Notes: The above figures display the best response value for the stated admissions scheme and for a given social weight assigned to the given equity-relevant objective.

C Additional Descriptive Analyses

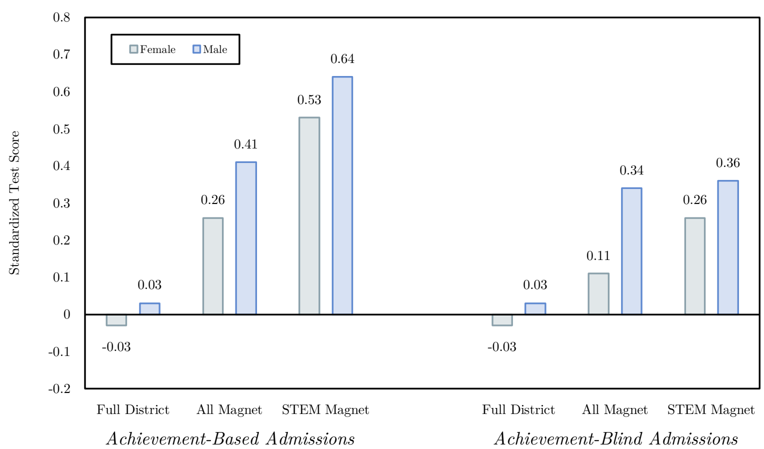
In another descriptive exercise, we investigate how the distribution of student ability in magnet programs changed between policies. Consistent with the literature on value-added, we use lagged, standardized test scores as a sufficient statistic for prior ability. We standardize these scores to have mean 0 and standard deviation 1 across-district and within-year. Figure C1 displays the average lagged test score (in units of standard deviation, σ) of all students in the district, followed by students that enroll in magnet programs and then students that enroll in STEM programs. We show these statistics for both admissions policies. Figure C1a (Figure C1b) shows these changes at the margin of race (gender).

In comparing the statistics in each panel for the full district, we find that the overall distribution of ability between groups did not change between policies. The gap in lagged test scores between URM students and non-URM students under either policy is equivalent to about 0.5σ on average. Among all magnet programs, the average ability of non-URM students does not change between policies; however, URM students admitted under the achievement-blind admissions policy have 0.23σ lower 3rd grade math exam scores on average. Among STEM programs, prior ability declines by 0.2σ among both groups. A similar pattern holds between male and female students, where the decline in ability for admitted female students is greater than that for male students.

Figure C1: Differences in 3rd Grade Test Performance Across Policies



(a) URM vs. Non-URM



(b) Female vs. Male

Notes: The above figures plot the average 3rd grade test scores for each listed demographic group. Math scores are standardized across district each year. Figure C1a displays changes between policies on the basis of race and Figure C1b does so on the basis of gender. "Full District" includes all 9th graders in the corresponding year, while "All Magnet" restricts to those that enroll in a magnet program and "STEM Magnet" further restricts to those enrolled in a STEM magnet program.

D An Illustrative Example of Unordered Treatment

Consider a district with five students (Krista, Lee, Penny, Rachel, and Sarah) that all share the same default school (h). In addition, the district offers two magnet programs, one with a STEM focus (m) and another with an art focus (a). Students may submit rank-ordered lists (ROLs) of at most length two. Students are entered into the lottery at j whenever they rank j as their first or second choice. Let $z_{ij} \in \{0, 1\}$ denote the lottery outcome (lose and win, respectively) and let $d_{ij} \in \{0, 1\}$ denote the subsequent enrollment decision.[1]

Table D1 displays the students' ROLs (columns 1 and 2), their lottery outcomes (column 3), and their subsequent enrollment decisions (column 4). As shown in the ROLs, Krista and Penny have preference ordering $m \succ a \succ h$, Lee and Rachel have preference ordering $a \succ m \succ h$, and Sarah has preference ordering $a \succ h \succ m$. This highlights the fact that treatment in this setting is *unordered*, meaning there does not exist a global ranking of alternatives across the district.

Column 3 shows that Krista and Rachel win their respective first-round lotteries while Lee and Penny lose theirs and therefore enter into the lotteries at their second-choice programs. Sarah, having only ranked one program, loses the lottery and can only attend her default school. The enrollment decisions in column 4 demonstrate that each student attends their most preferred school among those in their offer set.

Without loss of generality, suppose an economist seeks to identify the causal effect of attending the STEM magnet program, θ_m. Under a standard IV design, they would compare the outcomes of those who randomly won the lottery for m (Krista and Lee) to those who randomly lost (Penny). Denote this recovered estimate by $\hat{\theta}_m$. A key issue arises from the fact that heterogeneity in individuals' preference ordering generates undetectable differences in treatment effects. To see this, note that Krista would have attended a had she counterfactually lost the m lottery, while Lee would have attended h instead. As a result, $\hat{\theta}_m$ will be some unknown linear combination of the VA contribution from attending m over a and the contribution from attending m over h. In this example, the economist directly observes these weights (here, 0.5 and 0.5, because there is one student belonging to each margin), but this will not be true in general.

Even if the economist has data on students' ROLs, Kirkebøen, Leuven and Mogstad (2016) show that $\hat{\theta}_m$ does not have a meaningful interpretation under standard IV assumptions.[2] In particular, the estimated VA for m is an unknown combination of three partial effects: (1) the gains induced by attending m over h, (2) those induced by attending m over a, and (3) those induced by attending a over h. In other words,

[1] Toronto's lottery design guarantees that $\sum_j z_{ij} = 1$.

[2] This result is stated in Proposition 1 of Kirkebøen, Leuven and Mogstad (2016).

Table D1: Example of Unordered Treatments in School Choice

	Rank 1 School (1)	Rank 2 School (2)	Lottery Outcome (z_{ij}) (3)	Enrollment Decision (d_{ij}) (4)
Krista	m	a	$z_m = 1$	$d_m = 1$
Lee	a	m	$z_a = 0; \; z_m = 1$	$d_m = 1$
Penny	m	a	$z_m = 0; \; z_a = 1$	$d_a = 1$
Rachel	a	m	$z_a = 1$	$d_a = 1$
Sarah	a	–	$z_a = 0$	$d_h = 1$

Notes: The above table displays an example of five students and their corresponding rank-ordered list (columns 1 and 2), their lottery outcome (column 3), and their subsequent enrollment decision (column 4). Omitted rankings capture the possibility that students submit incomplete lists that do not maximize the district's total ROL capacity.

$\hat{\theta}_m$ is a function of $\hat{\theta}_a$, and vice versa. This means that Sarah's test scores will partially contribute to $\hat{\theta}_m$, despite not ranking m at all.

These issues can be resolved by conditioning on students' observed preferences. Kirkebøen, Leuven and Mogstad (2016) estimate two-stage least squares (2SLS) models at the j-k margin for $j \succ k$ marginal relevant preferences. This procedure is highly data-intensive and requires a substantial number of students with j-k relevant preferences. We lack sufficient power to conduct this analysis due to the fact that we only have lottery data for one year. An alternative approach could pool all magnet programs of each specialty to increase power. We avoid this strategy given our goal of estimating school-level VA and because this approach would eliminate much of the variation contained in the ROLs. Instead, we follow Abdulkadiroğlu et al. (2017) and Angrist et al. (2024) to simulate propensity scores that condition on each student's preferences and program-specific priorities induced by the quota system. In doing so, we control for preference heterogeneity in a lower-dimensional way than under the methods proposed by Kirkebøen, Leuven and Mogstad (2016). We include these controls in an otherwise straightforward IV design that leverages magnet program lotteries and traditional high school assignment.

E Derivation of Probit Formulation for Admissions Probabilities

Here, we formally derive the probit formulation of the probability displayed in Equation 20.

The existence of a cut-off quality. Fix a magnet program j with capacity c_j. Let I_j^A be the set of all students who apply to j. Under achievement-based admissions, the program solves

$$I_j^* \in \underset{I_j \subseteq I_j^A}{\operatorname{argmax}} \sum_{i \in I_j} q_{ij} \quad \text{s.t.} \quad |I_j| \leq c_j.$$

Let $q_j^{(1)} \geq q_j^{(2)} \geq \cdots \geq q_j^{(|I_j^A|)}$ denote the order statistics of $\{q_{ij}\}_{i \in I_j^A}$. We can then write the set of admitted students as

$$I_j^* = \{\, i \in I_j^A : q_{ij} \geq q_j^{(c_j)} \,\},$$

so the cutoff $\bar{q}_j$ is equivalent to $q_j^{(c_j)}$. In other words, the program's problem implies a cutoff rule: if $|I_j^A| > c_j$, the optimal I_j^* consists of the c_j applicants with the largest q_{ij} values. Hence, there exists a threshold $\bar{q}_j$ such that i is admitted iff $q_{ij} \geq \bar{q}_j$.

Probit formulation. Student quality is observed by programs and specified as

$$q_{ij} = \boldsymbol{\gamma}_j' \boldsymbol{X}_i + v_{is}, \qquad v_{is} \sim \mathcal{N}(0,1),$$

where the $\boldsymbol{\gamma}_j$ terms are program specific and $\boldsymbol{X}_i$ stacks the observed applicant characteristics considered in the admissions process (demographics and expected achievement), while v_{is} is an idiosyncratic quality component distributed standard normal. Denote by $\overline{\boldsymbol{X}}_j$ the characteristics of the marginally admitted student at program j and write $\bar{q}_j = \boldsymbol{\gamma}_j' \overline{\boldsymbol{X}}_j + \bar{v}_{js}$ with $\bar{v}_{js} \sim \mathcal{N}(0,1)$, independent of v_{is}.

The *ex ante* admission probability for applicant i is

$$\Pr(q_{ij} \geq \bar{q}_j) = \Pr(q_{ij} - \bar{q}_j \geq 0) = \Pr(\boldsymbol{\gamma}_j'(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_j) + v_{is} - \bar{v}_{js} \geq 0) = 1 - \Phi\left(\frac{\boldsymbol{\gamma}_j'(\overline{\boldsymbol{X}}_j - \boldsymbol{X}_i)}{\sqrt{2}}\right),$$

which follows from the fact that the difference of two independent $\mathcal{N}(0,1)$ random variables is itself normally distributed, specifically

$$v_{is} - \bar{v}_{js} \;\sim\; \mathcal{N}(0,2).$$

This form is exactly the one given in Equation 20.

F Generating Equilibrium-Consistent Propensity Scores

We simulate B samples of data and use the true lottery mechanism to assign simulated magnet program seats. For each draw $b \in B$ of data, we generate a simulated economy of N_b applying students with demographic characteristics ($\boldsymbol{x}_{ib}$) and submitted ROLs ($\{r_{1b}, r_{2b}\}$). Holding each program's capacity constraint (c_j) fixed to the actual capacities, we then run the lottery algorithm S times to recover b-specific propensity scores for each program.

This provides us with B "bootstrapped" simulations of data that map district-wide student decisions into program offers. By averaging across simulations, we can estimate program-specific propensity scores. In particular, we estimate linear probability models of the form

$$\begin{aligned}
p_{ij,1} &= \rho_0 + \rho_1 x_i^f + \rho_2 x_i^f \cdot x_j^{\text{STEM}} + \rho_3 x_i^{\text{URM}} + \rho_4 x_j^{\text{STEM}} + \rho_5 x_j^{\text{art}} + \rho_6 \left(\frac{N_{1,j}}{c_j}\right) + \rho_7 \left(\frac{N_{2,j}}{c_j}\right) + \epsilon_{ij,1} \\
p_{ij,2} &= \varrho_0 + \varrho_1 x_i^f + \varrho_2 x_i^f \cdot x_j^{\text{STEM}} + \varrho_3 x_i^{\text{URM}} + \varrho_4 x_j^{\text{STEM}} + \varrho_5 x_j^{\text{art}} + \varrho_6 \left(\frac{N_{1,j}}{c_j}\right) + \varrho_7 \left(\frac{N_{2,j}}{c_j}\right) + \epsilon_{ij,2}
\end{aligned} \tag{F1}$$

In the above, $p_{ij,k}$ is the probability of receiving an offer from the k-ranked program, $N_{k,j}$ is the total number of students that rank program j as k^{th} on their ROL, c_j is the program's capacity constraint, x_j^{STEM} (x_j^{art}) indicates whether j has a STEM (art) designation, and x_i^f (x_i^{URM}) indicates whether i is female (URM). We estimate these probabilities using simulated lottery offers in each of the B data samples.

Note that all estimates of the regression coefficients ($\widehat{\boldsymbol{\rho}}, \widehat{\boldsymbol{\varrho}}$) are *policy-variant* because they are constructed from the actual lottery mechanism. In other words, counterfactual changes in lottery quotas necessitate the re-estimation of Equation F1. Simulating the data B times in this manner is necessary because estimating the equations from the true data would not allow $\frac{N_{1,j}}{c_j}$ or $\frac{N_{2,j}}{c_j}$ to vary, nor would a simple bootstrap of the original data.

Table F2 presents summary statistics for the first-choice propensity score simulations. Panel A displays simple conditional probabilities of receiving an offer for URM and non-URM students, while Panel B displays estimates of Equation F1. Column 1 displays this information using the actual data. In Panel B, the dependent variable in column 1 is an indicator for whether student i received an offer from program j. Column 2 "bootstraps" the original data in a way that allows N to vary but fixes the URM quota to the true value. We reproduce the summary statistics and regression coefficients well. Columns 3 and 4 increase the URM quota and display the corresponding estimates. Holding fixed the original ROL distribution, the quota increase leads to a lower likelihood that a non-URM student receives a first-choice offer. The coefficient on URM increases, but other coefficients remain stable.

Table F2: Results for the Determinants of Simulated Offers

	Original Data	Simulations: URM Quotas			
		$\lambda^{\text{URM}} = 0.2$	$\lambda^{\text{URM}} = 0.4$	$\lambda^{\text{URM}} = 0.6$	$\lambda^{\text{URM}} = 0.8$
	(1)	(2)	(3)	(4)	(5)
Panel A: Offer Probabilities					
URM Students (First Choice)	0.452	0.480	0.698	0.701	0.772
Non-URM Students (First Choice)	0.283	0.272	0.247	0.227	0.213
URM Students (Second Choice)	0.143	0.212	0.204	0.190	0.167
Non-URM Students (Second Choice)	0.114	0.194	0.199	0.189	0.181
Panel B: Regression Results					
First Choice					
Female	−0.045	0.002	−0.005	−0.004	−0.004
	(0.027)	(0.002)	(0.002)	(0.002)	(0.002)
Female × STEM	0.099	0.056	0.057	0.053	0.050
	(0.032)	(0.002)	(0.002)	(0.002)	(0.002)
URM	0.174	0.176	0.320	0.444	0.534
	(0.019)	(0.001)	(0.001)	(0.001)	(0.001)
STEM	−0.028	0.175	0.167	0.156	0.145
	(0.033)	(0.002)	(0.002)	(0.002)	(0.002)
Art	0.038	0.030	0.030	0.029	0.026
	(0.026)	(0.002)	(0.002)	(0.002)	(0.002)
Choice 1 Oversubscription Rate	−0.051	−0.074	−0.069	−0.063	−0.059
	(0.004)	(0.000)	(0.000)	(0.000)	(0.000)
Choice 2 Oversubscription Rate	0.009	−0.036	−0.035	−0.036	−0.038
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Second Choice					
Female	−0.024	−0.012	−0.008	−0.009	−0.010
	(0.023)	(0.002)	(0.002)	(0.002)	(0.002)
Female × STEM	0.064	0.047	0.043	0.041	0.038
	(0.027)	(0.003)	(0.003)	(0.003)	(0.003)
URM	0.027	0.014	0.021	0.003	−0.026
	(0.016)	(0.002)	(0.002)	(0.002)	(0.002)
STEM	0.015	0.105	0.104	0.105	0.106
	(0.016)	(0.003)	(0.003)	(0.002)	(0.002)
Art	0.033	0.002	0.006	−0.007	−0.013
	(0.021)	(0.002)	(0.002)	(0.002)	(0.002)
Choice 1 Oversubscription Rate	0.011	−0.054	−0.049	−0.046	−0.045
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Choice 2 Oversubscription Rate	−0.035	−0.014	−0.015	−0.017	−0.018
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)

NOTES: Panel A presents the average likelihood of receiving an offer for the listed demographic group at their kth choice program. Panel B presents the Equation F1 coefficient estimates obtained via OLS. Robust standard errors are presented in parentheses. In each simulated estimation (columns 2-5), we set B to 100. The female quota in the above regressions are fixed at the true level of 0.5.

G Formulation of Standard Errors

G.1 Preferences

Here, we describe the procedure for computing standard errors for our estimated structural parameters using the simulated method-of-moments (SMM) approach. Without loss of generality, denote the parameters to estimate by Θ, the vector of moments used in SMM by M, the corresponding moments based on the current guess of parameter values by $M(\Theta_0)$, and the weighting matrix by W. Recall that we estimate Θ by solving

$$\widehat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \Big(M - M(\Theta)\Big)' W \Big(M - M(\Theta)\Big). \tag{G1}$$

The Central Limit Theorem provides us with the following limiting distribution for the SMM estimator,

$$\sqrt{N}(\widehat{\Theta} - \Theta) \xrightarrow{d} \mathcal{N}\Big(0, (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1}\Big), \tag{G2}$$

where $G = \nabla_{\Theta} M(\Theta)$ is the Jacobian of the moment vector with respect to the parameters and Ω is the variance-covariance matrix of the simulated moments. We numerically approximate G by perturbing each parameter by a small amount, h_{Θ}. Each entry for the approximated G is established as the central difference of the simulated moment vector, holding fixed all unperturbed parameters.

H Overview of Possible Social Objectives

Program-level segregation (EQ II). The previous policy attempts to make each of the J^M magnet programs in the district a microcosm of the district-wide distribution of demographics. This policy ignores program-specific preference heterogeneity at the margin of different demographic levels and instead enforces all programs to be balanced in these observed characteristics. In a more flexible objective, a policymaker may seek to balance demographics at the margin of overall magnet enrollment. In other words, parity may be measured at the *program* level instead of the school level. Similar to before, let $SGE_g(J^M)$ denote the same group exposure of group g across all magnet programs and $SGE_g(J^D)$ that across all traditional high schools. To minimize program-level segregation, a social planner solves

$$\min_{\mathbf{\Gamma}(\tau)} \sum_g |SGE_g(J^M) - SGE_g(J^D)|. \tag{H1}$$

Group-specific achievement gaps (EQ III). The previous two policies focus on improving students' access to magnet programs. Alternatively, the social planner may wish to minimize achievement gaps between students of different demographic groups, using *outcomes* as the metric of interest. For example, a district may want to adjust admissions criteria to provide disadvantaged students with higher quality resources to compensate against lower incoming achievement levels. Let $\overline{Y}_g$ denote the district-wide average 9th grade math score for group g and $\overline{Y}_{\not g}$ the same for those who do not belong to group g. We can then write the district's problem as

$$\min_{\mathbf{\Gamma}(\tau)} \sum_g |\overline{Y}_g - \overline{Y}_{\not g}|. \tag{H2}$$

The implicit assumption in this form is that the district values these between-group achievement gaps identically. Another version of this problem could be written that allows for Pareto weights that are group specific. For simplicity, we assume all weights are identical.

Student utility (EF II). From an economic perspective, an optimal policy may be one that maximizes the average level of student utility across a district. This would account for differences in students' preferences, which the previous social objectives ignore. We can define the problem of the district that seeks to maximize student utility as

$$\max_{\mathbf{\Gamma}(\tau)} \frac{1}{N} \sum_{i=1}^{N} u_{ij}. \tag{H3}$$

Magnet program VA (EF III). The previous policy focused exclusively on students' well-being. The district may instead design an admissions scheme in order to maximize the overall value-added of its magnet programs. This social objective can be stated as

$$\max_{\boldsymbol{\Gamma}(\tau)} \frac{1}{J^M} \sum_{j=1}^{J^M} \theta_L(\boldsymbol{m}_j, \boldsymbol{x}_j, \xi_j) + \theta_H(\boldsymbol{m}_j, \boldsymbol{x}_j, \xi_j). \tag{H4}$$

The above form calculates a program's *overall VA* as the sum of returns for low and high ability students. We focus on maximizing magnet program VA rather than all schools' VA because reshuffling students to improve standardized test scores is a zero-sum game in our model.[3]

[3] To see this, note that the decomposition of value-added considers peer effects to be linear. A version of this decomposition that includes complementarities between peer effects and non-peer effects yields negligible differences from our preferred version.

I The Deferred Acceptance Algorithm

I.1 Details

As a separate class of counterfactuals, we benchmark each of our objectives against a scenario in which the district implements a DA admissions mechanism. Here, we detail the first two versions considered in the manuscript. First, in the "achievement-based" DA, we establish students' preferences for schools as the estimated $\widehat{\boldsymbol{\Theta}}$ and similarly establish schools' preferences for students as the estimated $\widehat{\boldsymbol{\gamma}}$. In the second, which we call the "achievement-blind" DA, we eliminate programs' preferences and replace the offer-generating process with a simple, random lottery. DA is strategy-proof when ROLs are unconstrained. We therefore allow students to rank all 28 programs (instead of Toronto's limit of two). The strategy-proofness of the mechanism means that students' dominant strategy is to report their preferences truthfully, i.e., on the basis of their utilities undistorted by the likelihood of admissions.

The "achievement-based" DA mechanism works in the following way:

1. Students submit an ROL ($\mathcal{A}_i$) of length L_i, where

$$L_i \;=\; \big|\{\, j \in J^M : u_{ij} \;\geq\; v_i \}\big|, \qquad \mathcal{A}_i \;=\; \big(j_{i,1},\, j_{i,2},\, \ldots,\, j_{i,L_i}\big),$$

   with

$$u_{i\,j_{i,1}} \;\geq\; u_{i\,j_{i,2}} \;\geq\; \cdots \;\geq\; u_{i\,j_{i,L_i}} \;\geq\; v_i.$$

   Note: If $L_i = 0$, then $\mathcal{A}_i = \varnothing$.

2. In round $t = 1$, each student i with $L_i \geq 1$ applies to their first choice $j_{i,1}$. Define

$$I_{j,1} \;=\; \{\, i : j_{i,1} = j \}, \qquad j \in J^M.$$

   Program j has capacity c_j. It computes each q_{ij} for $i \in I_{j,1}$, orders that set in descending order of q_{ij}. Then,

$$\begin{cases} \text{if } |I_{j,1}| \;\leq\; c_j, & j \text{ tentatively admits all of } I_{j,1}, \\ \text{if } |I_{j,1}| \;>\; c_j, & j \text{ admits only the top } c_j \text{ students by } q_{ij}. \end{cases}$$

   Every other student in $I_{j,1}$ is tentatively rejected.

3. In general, for round $t \geq 2$, let

$$U_{t-1} \;=\; \{\, i : i \text{ is unmatched after round } t-1 \text{ and } L_i \geq t \}.$$

Each $i \in U_{t-1}$ applies to their t-th choice $j_{i,t}$. Define

$$I_{j,t} \;=\; \{\, i \in U_{t-1} : j_{i,t} = j \}.$$

Furthermore, define $I^*_{j,t-1}$ the set of students that j tentatively accepted in round $t-1$. Program j then considers

$$\widetilde{I}_{j,t} \;=\; I^*_{j,t-1} \,\cup\, I_{j,t},$$

ranks every $i \in \widetilde{I}_{j,t}$ in descending order of q_{ij}, and admits up to c_j:

$$\begin{cases} \text{if } |\widetilde{I}_{j,t}| \;\leq\; c_j, & j \text{ admits all of } \widetilde{I}_{j,t}, \\ \text{if } |\widetilde{I}_{j,t}| \;>\; c_j, & j \text{ admits only the top } c_j \text{ by } q_{ij}. \end{cases}$$

Any student who had been tentatively admitted in a prior round but now falls outside the top c_j is rejected and returns to the unmatched pool.

4. The algorithm stops when either

   - every student is either tentatively admitted to some $j \in J^M$ or has exhausted their ROL (i.e. $L_i < t$), or
   - every program j already holds c_j students and no unmatched student has any remaining choice left.

5. At termination ($t = T$), each magnet j has a final set of at most c_j students, namely

$$\{\, i : i \text{ was ever admitted to } j \text{ in rounds } 1, \ldots, T \text{ and is among the top } c_j \text{ by } q_{ij} \}.$$

Those students become matched to magnet j. Any student who was never tentatively admitted chooses between their default school and the outside option.

The "achievement-blind" version follows the same admission steps, except that programs ignore $\widehat{\gamma}$ and instead assign seats by random lottery. Specifically:

1. In round 1, each program j conducts a lottery among all students who ranked j first. If this fills j's capacity c_j, then j's DA algorithm terminates. If seats remain, the process moves on to round 2.

2. In round 2, program j lotteries over those students who are still unmatched after round 1 and rank j second. Winners fill any remaining seats; if j then becomes full, it stops, and if not, the next round of unmatched students are considered.

3. The process repeats until either all seats in j are filled or no unmatched student has j at any remaining rank.

Note that each q_{ij} includes an unobserved continuous error term. Therefore, ties occur in our simulation with probability zero.

I.2 DA Properties with Peer Effects

When peer effects enter utility, preferences are assignment dependent, so the global guarantees of stability and strategy-proofness need not hold. We therefore interpret our DA counterfactual as the outcome of a fixed-point algorithm over peer aggregates and recover the usual DA properties at that fixed point for the induced lists.

Let $\boldsymbol{m} = (\boldsymbol{m}_j)_{j \in J^M}$ denote the vector of peer aggregates across programs. For any fixed $\boldsymbol{m}$, each student i forms utilities $\{u_{ij}(\boldsymbol{m})\}_{j \in J^M}$ and submits a rank-ordered list $\mathcal{A}_i(\boldsymbol{m})$ consistent with

$$u_{i\,j_{i,1}}(\boldsymbol{m}) \;\geq\; u_{i\,j_{i,2}}(\boldsymbol{m}) \;\geq\; \cdots \;\geq\; u_{i\,j_{i,L_i}}(\boldsymbol{m}) \;\geq\; v_i.$$

Let $\mathcal{A}(\boldsymbol{m}) \equiv \{\mathcal{A}_i(\boldsymbol{m})\}_{i=1}^{N}$ denote the profile of ROLs. Given program priorities implied by $\widehat{\boldsymbol{\gamma}}$ (or lotteries in the achievement-blind version), running DA with inputs $\big(\mathcal{A}(\boldsymbol{m}), \widehat{\gamma}\big)$ returns a matching

$$\mu(\boldsymbol{m}) \;=\; \mathrm{DA}\big(\mathcal{A}(\boldsymbol{m}), \widehat{\gamma}\big).$$

Let $\mathcal{M}(\mu)$ map an allocation into the implied peer aggregates. Define the fixed-point mapping

$$\Phi(\boldsymbol{m}) \;=\; \mathcal{M}(\mu(\boldsymbol{m})) \;=\; \mathcal{M}\Big(\mathrm{DA}\big(\mathcal{A}(\boldsymbol{m}), \widehat{\gamma}\big)\Big).$$

A *DA–peer fixed point* is any $(\boldsymbol{m}^*, \mu^*)$ with $\boldsymbol{m}^* = \Phi(\boldsymbol{m}^*)$ and $\mu^* = \mathrm{DA}\big(\mathcal{A}(\boldsymbol{m}^*), \widehat{\gamma}\big)$.

Recall that the peer component of u_{ij} is linear in the peer-share vector $\boldsymbol{m}_j$, and the portion of program value-added that depends on peers is also linear in $\boldsymbol{m}_j$ under our decomposition. Let $\widehat{\boldsymbol{\beta}}_i^{\text{peer}}$ denote the estimated coefficient vector on peer aggregates in u_{ij} for student i.

Stability at the fixed point. *Claim.* μ^* is Gale–Shapley stable with respect to the ROL profile $\mathcal{A}(\boldsymbol{m}^*)$ and the program priorities used in DA.

Proof. For fixed $\boldsymbol{m}$, the submitted lists $\mathcal{A}(\boldsymbol{m})$ are exogenous inputs to DA. The DA outcome $\mu(\boldsymbol{m})$ is stable relative to those lists and the stated priorities. At a fixed point, $\mathcal{M}(\mu^*) = \boldsymbol{m}^*$, so the peer aggregates used to form $\mathcal{A}(\boldsymbol{m}^*)$ coincide with those implied by μ^*. □

Strategy-proofness with peer effects. With unconstrained, costless ROLs and fixed priorities, truthful reporting is a dominant strategy in DA for fixed preferences. When preferences depend on peer aggregates, a student can alter those aggregates by misreporting, so dominant-strategy truthfulness need not hold globally. Let c_j denote program capacities and $n_{\min} = \min_j c_j$. At a DA-peer fixed point $(\boldsymbol{m}^*, \mu^*)$, moving student i from program a to b changes at most two peer means: for each component k, $|\Delta m_{a,k}| \leq 1/c_a$ and $|\Delta m_{b,k}| \leq 1/c_b$, hence

$$\|\Delta \boldsymbol{m}\|_\infty \;\leq\; \frac{1}{c_a} + \frac{1}{c_b} \;\leq\; \frac{2}{n_{\min}}.$$

We then apply this logic to our utility model. Recall that

$$u_{ij} = \alpha_i \theta_k(\boldsymbol{m}_j, \boldsymbol{x}_j, \xi_j) + \boldsymbol{\beta}_i' \boldsymbol{m}_j + \boldsymbol{\zeta}_i' \boldsymbol{x}_j - \psi D_{ij} + \varepsilon_{ij} + \varepsilon_{ij}^{\mathcal{O}}.$$

Substituting $\theta_k(\boldsymbol{m}_j, \boldsymbol{x}_j, \xi_j) = \boldsymbol{\theta}_{m,k}' \boldsymbol{m}_j + \boldsymbol{\theta}_{x,k}' \boldsymbol{x}_j + \xi_j^k$, the peer-aggregate deviation yields

$$\Delta u_{ij} \;=\; (\alpha_i\, \boldsymbol{\theta}_{m,k}' \;+\; \boldsymbol{\beta}_i') \Delta \boldsymbol{m}_j.$$

By Hölder's inequality and $\|\Delta \boldsymbol{m}_j\|_\infty \leq 2/n_{\min}$,

$$|\Delta u_{ij}| \;\leq\; \Big(|\alpha_i|\, \|\boldsymbol{\theta}_{m,k}\|_1 + \|\boldsymbol{\beta}_i\|_1 \Big)\, \|\Delta \boldsymbol{m}_j\|_\infty \;\leq\; \frac{2}{n_{\min}} \Big(|\alpha_i|\, \|\boldsymbol{\theta}_{m,k}\|_1 + \|\boldsymbol{\beta}_i\|_1 \Big) \equiv M.$$

Therefore truthful reporting is an ε–best response at $(\boldsymbol{m}^*, \mu^*)$ with $\varepsilon_i \;\leq\; M$. The same bound applies under lottery priorities when the seed is fixed.