High School Admissions Policies in Equilibrium*

Sherrie Cheng
University of Pennsylvania

Zach Weingarten
University of Pennsylvania

Job Market Paper

Preliminary, please do not circulate.

February 27, 2025

Abstract

Over the last decade, many public school districts across the world have eliminated merit-based admissions criteria at highly competitive programs in an effort to improve access for disadvantaged students. This paper quantifies the equity and efficiency trade-offs associated with such policies in equilibrium. We focus on Toronto, where screened admissions to high school magnet programs were abruptly replaced with a centralized lottery that explicitly prioritized underrepresented minority (URM) and female students. Using novel administrative data on the universe of Toronto 9th graders, we leverage randomized lottery offers to recover causal estimates of school quality. We then decompose the overall effects into peer and non-peer components in order to trace out changes in the distribution of school quality in response to large shifts in student sorting. We find that the policy acted as a school quality "equalizer:" average magnet program quality declined by 0.17σ while traditional high school quality increased by 0.08σ. To determine the policy's effect on student welfare, we develop and estimate an equilibrium model of school choice under both policy regimes. Our results show that the policy was welfare-improving for URM students and low ability students but welfare-decreasing for non-URM students and high ability students. We use the estimated model to design optimal admissions policies under alternative definitions of equity and efficiency. We show that the socially optimal design depends on the measure of equity and efficiency used and that both policies studied have scope for improvement. Finally, we provide evidence that school districts can maintain a merit-blind initiative and jointly improve equity and efficiency by adopting a centralized, strategy-proof admissions mechanism.

JEL Classification: I20, I24, I31

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

*We are thankful to Petra Todd, Francesco Agostinelli, and Margaux Luflade for invaluable feedback and support. We also thank Juan Pablo Atal, Juan Camilo Castillo, Hanming Fang, Chao Fu, Olivier De Groote, Viviana Rodriguez, and Holger Sieg for helpful comments and suggestions. We finally thank the Toronto District School Board for providing us with data. Contact information: Cheng: scheng1@upenn.edu. Weingarten: zachwein@upenn.edu.

1 Introduction

Recent concerns over the equitable provision of high-quality education have prompted many school districts around the world to eliminate merit-based admissions policies in favor of merit-blind alternatives. These reforms aim to reduce historical and cultural barriers for students from disadvantaged backgrounds.[1] Despite the growing prevalence of these policies, they remain controversial. Proponents argue that the elimination of screening mechanisms promotes fair access to highly sought-after schools (Shapiro, 2021), while opponents insist that such policies could lead to an inefficient allocation of students to schools, with high-achieving students losing access to the most rigorous programs and low-achieving students enrolling in programs for which they may be unprepared (Elsen-Rooney, 2024). In the United States, this debate culminated in the landmark Supreme Court case *Students for Fair Admissions v. Harvard*, in which merit-blind initiatives were declared unlawful.[2]

Reconciling this debate is challenging *ex ante* for several reasons. First, changes in admissions criteria can cause unpredictable changes in students' application and enrollment decisions. Students may adjust their expectations about their chances of receiving an offer and the characteristics of other students who will receive offers alongside them. Moreover, the distribution of school quality is likely to be affected by changes in peer composition (Allende, 2019), which will further impact students' demand for schools. Finally, predicting which students gain the most from a high-quality school is theoretically ambiguous. If schools exhibit a high degree of complementarity with high-achieving students, then merit-based screening may be optimal. However, if these programs exhibit marginally decreasing returns to ability, then admitting lower-achieving students should be prioritized in order to mitigate educational achievement gaps.

In this paper, we address the following research question: What mechanisms drive the trade-off between equity and efficiency in education admissions, and how can policy interventions optimize this trade-off? To address this question, we analyze a recent policy change in Toronto that replaced achievement-based admissions to high school magnet programs with a centralized, "achievement-blind" lottery system. The new policy additionally sought to improve student representation in magnet programs by introducing

[1] For example, Shanghai abolished a policy in 2018 that gave high school admissions premiums to high socioeconomic status (SES) students (Zhu, Zhang and Wang, 2023) and Chile replaced the use of selective criteria with a centralized assignment mechanism in 2015 (Kutscher, Nath and Urzúa, 2023). In the United States, large school districts in urban areas like Boston (Barry, 2021), Chicago (Pathak and Sönmez, 2013), New Orleans (Abdulkadiroğlu et al., 2016), and Philadelphia (Mezzacappa, 2021) introduced similar changes to their admissions processes, each with a stated goal of heightened accessibility and increased equity for marginalized students.

[2] Although this decision explicitly impacted post-secondary admissions guidelines in the United States, it has further impacted K-12 criteria. On February 14, 2025, Acting Secretary for Civil Rights Craig Trainor released a memo explicitly detailing that "[it] would, for instance, be unlawful for an educational institution to eliminate standardized testing to achieve a desired racial balance or to increase racial diversity." The full memo can be found here.

a quota system that reserved 20% of all seats for underrepresented minority (URM) students and 50% of all seats at STEM programs for female students.

We implement a two-part analysis to address our research question. In the first part, we leverage the randomization in lottery-based admissions offers to construct causal measures of school quality. We decompose these quality measures to determine the relative contribution from peers and other education inputs, which allows us to trace out changes in school quality levels in response to changes in student enrollment decisions. This empirical analysis provides insight into how equilibrium school quality changes in response to changes to admissions.

To understand the mechanisms that govern student sorting across the district, we next develop and estimate a static equilibrium model of high school admissions under both policies, which provides us with students' preferences for programs and programs' preferences for students. We use the estimated model to measure equilibrium changes in student welfare, student achievement, and school quality. In counterfactual policy experiments, we consider several different policy objectives that speak to equity (either maximizing district-wide student representation or minimizing between-group achievement gaps) and efficiency (maximizing either student achievement or student welfare). We provide policy recommendations for how districts can achieve these goals and how to optimize the trade-offs associated with each. In our final counterfactual exercise, we situate these findings against centralized, strategy-proof mechanisms to determine the extent to which the studied policies can be made more efficient.

We rely on novel student-level administrative data provided to us by the Toronto District School Board (TDSB) to conduct our analysis. These data include three years of 9th grade students' demographic information, including race, gender, the high school or magnet program that each student attends, and the middle and high school that each student is zoned to attend based on their home residency. The administrative data also include standardized test scores in math (grades 3 and 9) and reading (grade 3). For the first year that Toronto utilized the achievement-blind admissions policy, we have detailed application data that includes students' rank-ordered lists, their initial offers, and their lottery waitlist numbers. We supplement these administrative records with publicly available information on school locations and amenities.

Descriptive results from analyzing the data show that the policy change led to drastic changes in the distribution of students across Toronto. URM student representation in magnet programs increased by 88% (14 percentage points) after the implementation of the achievement-blind admissions policy. In STEM programs, female representation increased by 5% (2.2 percentage points). However, these gains in representation were met with losses in student preparedness. In particular, incoming student achievement declined by 10% of a standard deviation (σ) on average, as measured by lagged reading and math scores.

These changes in peer characteristics translate to lower achievement in 9th grade. Across all magnet programs, students' 9th grade math scores decline by 0.15σ between policies. This decline is nearly twice as large in STEM programs.

Motivated by these findings, we determine the causal effect that the policy change had on the distribution of school quality. We follow Angrist et al. (2024) to estimate school-level value-added models that control for students' *ex ante* likelihood of program admission. These measures leverage both the randomization in offers induced by lottery admissions and students' stated preferences as measured by their submitted rank-ordered lists. This empirical framework addresses concerns of selection into programs, which allows us to recover credible estimates of the effect of a particular school on 9th grade achievement. Building off of the intuition in Allende (2019), we then decompose these overall effects into three separate channels: isolated peer effects, isolated non-peer effects, and a third channel that allows for complementarity between the two. We find that peer effects contribute significantly to a school's overall effect, but that the influence of high ability peers is greatest at traditional high schools. As a result, the policy caused traditional school quality to improve by 0.08σ and magnet program quality to decline by 0.17σ on average. In this way, the adoption of the achievement-blind admissions scheme acted as a school quality "equalizer."

Although this reduced-form analysis informs us of the policy's effect on school quality, it cannot address the possible effects that eliminating admissions criteria has on student welfare, nor can it provide insight to determine which students should be prioritized or how to design an optimal admissions policy for a given district objective. To address this, we develop and estimate a static equilibrium model of school choice under each admissions scheme. The demand side of the market is characterized by students who make magnet program application decisions and, after receiving offers, make school enrollment decisions. Students are strategic and thus consider their *ex ante* likelihood of admissions when applying to programs. Additionally, students' demand for schools depends on endogenous peer characteristics. These considerations are important in order to capture general equilibrium effects that occur in response to counterfactual policies.

The supply side of the market features magnet programs that admit students on the basis of program-specific preferences for student characteristics, subject to their capacity constraints. Under the original admissions design, programs choose which students to admit on the basis of perceived quality, which depends on incoming achievement measures and demographic characteristics. When these criteria are eliminated, programs lose their discretion to admit students and offers are instead allocated randomly according to the district's lottery mechanism.

We estimate the model in two parts. First, we recover students' preferences using simulated method

of moments (SMM) on the achievement-blind data alone. The benefit of this approach is that the policy removes programs' ability to select students, simplifying the identification of demand-side parameters by eliminating the supply side entirely. With students' preferences in hand, we then estimate programs' preferences for students via SMM on the achievement-based data. Estimating the model in this way provides us with greater flexibility to construct counterfactual policy experiments that consider *either* admissions regime as the baseline.

We uncover significant preference heterogeneity across student demographics. Students exhibit strong preferences for school value-added and each type of magnet program. Their preferences also point to strong homophily bias, with students deriving large marginal utility from increases in observably similar peers. On the supply side, our estimates suggest that magnet programs favored non-URM students in the original policy. Art programs place large admissions priority toward students with high lagged reading scores, whereas STEM and IB programs value prior math achievement most.

Our results indicate that the policy change in Toronto led to a 38% increase in average student utility. We estimate that approximately one-third of these gains are explained by reductions in application costs associated with the adoption of the achievement-blind regime, with the remaining two-thirds attributable to changes in access to magnet programs and subsequent student sorting patterns. However, these average improvements to welfare mask underlying heterogeneity in gains. While URM students and low ability students experience substantial gains driven by their newfound access to magnet programs, non-URM students and high ability students face significant losses.

We use the estimated model to conduct a continuum of counterfactual policy experiments. Our ultimate goal is to design optimal admissions policies that balance equity and efficiency trade-offs. For this, we require explicit definitions of both "equity" and "efficiency." We consider two flexible definitions for each in an effort to provide thorough and holistic guidance to districts. In particular, a policy that targets equity in *opportunity* seeks to maximize district-wide student representation across schools, whereas a policy that targets equity in *outcomes* minimizes between-group achievement gaps. On efficiency, we consider a policy-relevant definition (maximizing student achievement) and an economic definition (maximizing student welfare). We are able to trace out the frontier of the equity-efficiency trade-off using any of the aforementioned definitions and under either regime implemented by TDSB. Under the original policy, we estimate which collection of admissions weights achieves the stated goal, while under the centralized framework, we consider different combinations of lottery quotas.

The results of our counterfactual exercises show that the optimal policy design depends entirely on the district's objectives. We show that whenever welfare is used to measure efficiency, the equity-efficiency

trade-off is optimized by placing no weight on equity and all weight on efficiency. Alternatively, if standardized test score performance is used to measure efficiency, then an optimal admissions policy places all weight on equity, regardless of how equity is measured. These findings do not depend on whether or not we use the achievement-based or the achievement-blind regime as our baseline. However, the magnitude of the trade-off between equity and efficiency does depend on the method in which programs admit students.

Finally, we investigate whether equity and efficiency could be improved under a different class of admissions mechanisms. We design two different Deferred Acceptance (DA) algorithms based on the two Toronto policies. First, we consider an "achievement-based" DA mechanism that fixes the supply-side preferences to the achievement-based ones recovered in estimation. Next, we design an "achievement-blind" version of DA that replaces the supply side with program-specific lotteries. Both DA mechanisms generate considerable improvements to welfare. The achievement-blind DA yields gains in both measures of equity as well, suggesting that students would benefit holistically if magnet program admissions were conducted via a strategy-proof mechanism.

This paper relates to several strands of the literature. We first contribute to 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; Arteaga et al., 2022). Our setting features an abrupt change from a decentralized assignment process to a fully centralized one. The centralized process also uses a unique assignment mechanism that is not strategy-proof. In our paper, we are able to situate this assignment mechanism against a DA algorithm to determine the extent to which this process is inefficient.

We make additional modeling contributions by combining two disparate strands of work in the school choice literature. In particular, structural models of school choice often omit peers entirely due to concerns of multiple equilibria (Agarwal and Somaini, 2018). Those that do include endogenous peer effects as a component of demand are either estimated on data from districts that use strategy-proof centralized assignment mechanisms or make explicit assumptions that students do not behave strategically (Epple and Romano, 1998; Rothstein, 2006; Allende, 2019; Barseghyan, Clark and Coate, 2019; Abdulkadiroğlu et al., 2020). Conversely, recent work that allows for strategic behavior in students' decision-making shuts off peer effects entirely (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 includes both

endogenous peer effects and strategic behavior together. This distinction is important as it allows students' expectations about their own likelihood of admissions *and* the types of students who will attend alongside them to depend on the policy in place. A model that omits either of these attributes would understate possible general equilibrium effects and therefore yield inaccurate conclusions about student welfare.

Given the nature of the policy change, our paper also relates to work on affirmative action policies in large education markets. Previous studies that have examined the effects of policies that give admissions priority to underrepresented demographic groups have focused primarily on higher education settings (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). Much less attention has been given to primary or secondary education environments, in part due to the limited viability of enacting such policies in public school districts. Furthermore, most work that studies the impact of policies designed to increase K-12 access are limited to settings in the United States (Abdulkadiroğlu, Pathak and Walters, 2018; Barrow, Sartain and De La Torre, 2020; Dur, Pathak and Sönmez, 2020; Ellison and Pathak, 2021; Idoux, 2022). We build on this literature by focusing on a new setting (Toronto, Canada) that implemented explicit race and gender quotas to improve access to specialized secondary schools. Our empirical design additionally allows us to analyze the effects of a continuum of counterfactual admissions policies in equilibrium. Importantly, we are able to quantify the trade-off between equity and efficiency that occurs in response to each policy change. By doing so, we are able to optimize this trade-off from the perspective of a social planner or a district superintendent, as well as provide specific policy recommendations with an eye toward improving both equity and efficiency in simultaneity.

Finally, we provide methodological contributions for the estimation of school-level VA. A small but robust literature has been developed to recover credible measures of quality, or VA, at the school level (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 combine the strategies developed in Abdulkadiroğlu et al. (2017) and Angrist et al. (2024) with the VA analysis implemented in Allende (2019) to decompose school VA into peer effects and non-peer effects. We further introduce input complementarity into the production function of VA, which flexibly allows different types of peers to be heterogeneously productive at separate types of schools. This distinction is important, as a model that ignores peer-school complementarity would treat district-wide student reshuffling as a zero-sum game.

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 school quality. 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 and describes the effects of the policy change on student welfare. Section 7 provides an overview of our counterfactual policy analysis 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, Canada, one of the largest school districts 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 27% as either Black, Hispanic, or Middle Eastern. Nearly 15% of students identify as LGBTQ+ and 8% self-report as having a disability.[3] 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.[4]

Toronto offers approximately 85 public secondary schools.[5] Each student in the district is guaranteed a seat at their default secondary school 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 groups: art, STEM, and International Baccalaureate (IB).[6] 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

[3] These figures come from a 2023 survey administered by the district to all families (60% response rate).
[4] These statistics come from the 2021 Census conducted by the city of Toronto.
[5] Our analysis includes the 67 schools that have a 9th grade student body of at least 10 students.
[6] 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.

admissions process, which depended on the program's focus.[7] Although this process was decentralized, TDSB limited students to apply to a maximum of two magnet programs in any given admissions cycle.[8]

Citing continued inequity in access to these programs, TDSB eliminated the achievement-based admissions process entirely beginning 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:

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

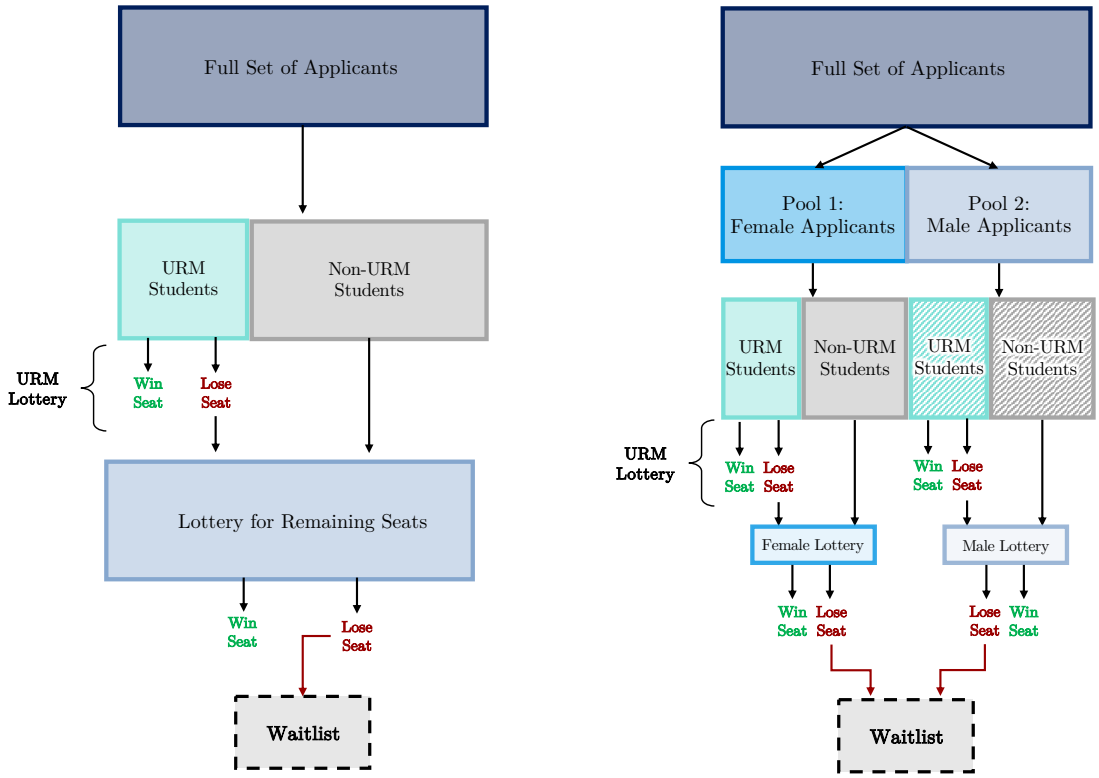
2. A program's applicant pool is the set of students that rank that program *either* first or second. STEM programs split this pool into two groups—one for female applicants and one for male applicants—while non-STEM programs consider the pool in its entirety. All lotteries in the district are conducted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

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.

[7] For example, a visual arts magnet programs may have required students to submit a portfolio of their past artistic work.
[8] 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).

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 have constructed three years of newly available administrative data on 9th graders. For the 2021-2022, 2022-2023, and 2023-2024 academic years, we have detailed information on the universe of 9th graders' demographics, achievement, and enrollment. The demographic data include

students' race, gender, and nationality. These data further include students' standardized test scores in reading, math, and writing for 3rd grade, as well as math test scores for 9th grade.[9]

The district also provided us with detailed enrollment information for all K-12 students. For each academic year, we observe students' enrollment decisions, as well as their assigned default school. The panel structure of these data allow us to recover each 9th grade student's default middle school. Approximately 50% of families participated in a student census survey conducted by the district, which provides us with additional demographic information related to parents' marital status and education level.

We also have application data for the universe of 8th grade students that applied to either an art, STEM, or IB magnet program during the first year of the achievement-blind admissions scheme. These data include students' program rank-ordered lists, lottery numbers, offer outcomes, and waitlist outcomes. We use unique student identifiers to merge these data with data on enrollment decisions.

Finally, we have a three-year panel of the approximately 2,400 secondary school teachers in Toronto that worked at a magnet program for at least one year during our panel. Although we cannot determine any teachers' demographics or experience level, these data provide us with a sense of district-wide teacher mobility each year. In particular, we use this information to test for behavioral responses from teachers after the policy change. We display descriptive statistics from these data in Appendix 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.[10] We collect publicly available catchment area boundaries for the nearly 500 elementary, middle, and high schools in the district, which also provide 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 residence. We establish a student's home address as the centroid of their corresponding zone of residency.[11]

This measure of home residency is used to compute travel times to each magnet program in the district, as well as the travel time between a student's home and default high school. To do so, we rely on Google API travel time calculations for local public transit, which is the most common mode of transportation for high school students in Toronto. We set the time required to travel between a school and home as the simple average of those found 45 minutes before school starts and 15 minutes after schools ends.

Finally, the administrative data are missing information on race for 23% to 33% of 9th graders, depend-

[9] Although students take these tests once more in 6th grade, the Covid-19 pandemic paused these exams for the students in both years of our sample.
[10] We also observe the designated elementary school for the subset of students that submit applications to magnet programs.
[11] See Appendix Figure B2 for an illustrative example of this imputation procedure.

ing on the year. Similarly, 3rd grade math test scores are missing for about one-third of all students each year. We 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 displays the summary statistics for our two main analysis samples. The first three columns show information on 9th graders in the 2022-2023 academic year, the last year to use the achievement-based admissions policy. The final three columns display the same information for the 2023-2024 cohort, the first to experience the new achievement-blind admissions policy.

In both years, we remove from the analysis sample students with severe learning disabilities due to unobserved accommodations they may receive in their traditional public schools, as well as the fact that many of these students do not take standardized exams. In the 2023-2024 cohort, we further restrict the magnet sample to those with whom we have linked application data. We remove a small fraction of students that enroll in magnet programs but have no ROL data available, resulting in a pre-reform magnet sample that is 5% larger than the post-reform sample. Finally, we remove all students that either rank or enroll in athletic magnet programs, which were not impacted by the policy change and did not screen on the basis of 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 either art, STEM, or IB magnet programs. Columns (3) and (6) restrict only to those who enroll in a STEM program. Changes in demographic composition capture changes in the way that students sorted into the magnet application system *and* changes in the way that programs accepted 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 school quality across the district.

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

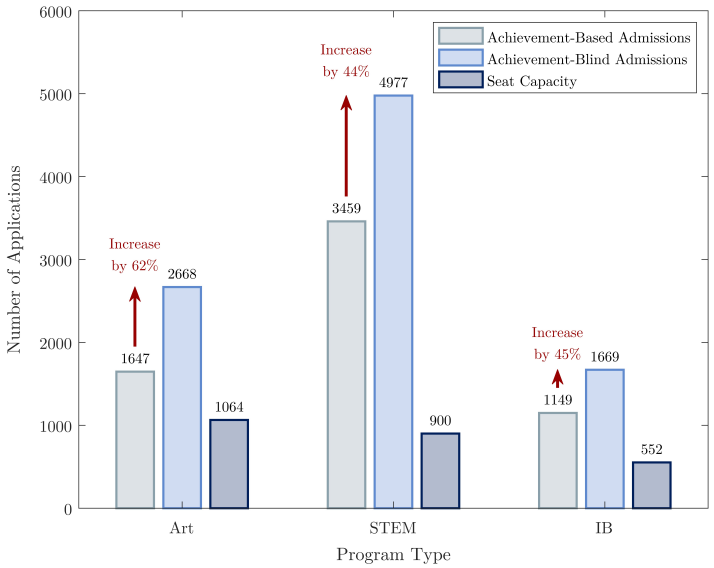
Notes: The above table displays descriptive statistics for the universe of 9th grade students attending Toronto public schools during the achievement-based admissions policy (2022-2023) and the achievement-blind admissions policy (2023-2024). Test scores are standardized to have mean 0 and standard deviation 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 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.

2.6 Descriptive Analysis

As shown in Figure 2, demand for magnet programs in Toronto greatly outpaces supply under either admissions regime. During the final year of the achievement-based admissions policy, magnet programs that had an art, STEM, and IB focus saw respective application-to-seat ratios of 1.5:1, 4:1, and 2:1. 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 from all students in the district. In fact, demand for each type of program increased by 62%, 44%, and 45%, 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 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 40% more likely to receive an offer from their favorite program and 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.

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.

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.

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. Among the full sample of applicants, approximately 40% enroll in a magnet program and 43% enroll in their default high school in the district. The remaining fraction of students attend either private schools in the district or other schools outside of the district (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.

Table 2: Application Outcomes Under Achievement-Blind Admissions

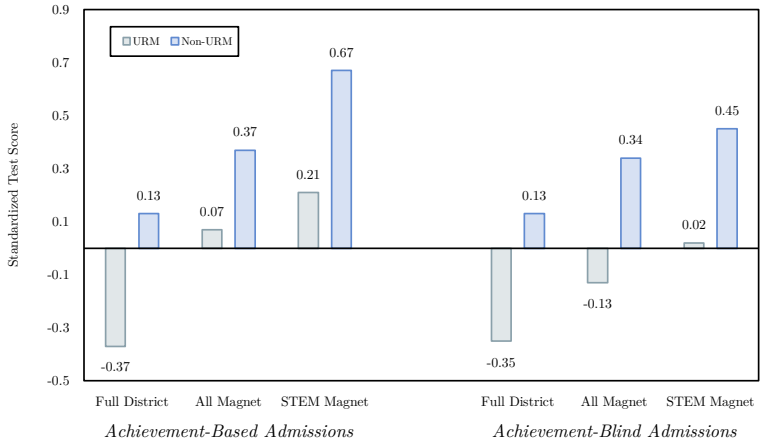
	All Rank-Ordered Lists			STEM-Only Rank-Ordered Lists		
	All Applicants (1)	URM Only (2)	Female Only (3)	All Applicants (4)	URM Only (5)	Female Only (6)
Panel A: Initial Lottery Outcomes						
Win Offer at First Choice	0.322	0.452	0.345	0.236	0.375	0.260
Win Offer at Second Choice	0.121	0.143	0.128	0.083	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 summary statistics for those same students in the beginning of the 2023-2024 academic year. Students that select the outside option refer to students that apply to the district's magnet programs from outside of the district. For these students, we only observe their lottery outcome and not their subsequent enrollment decision. Panel C 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.

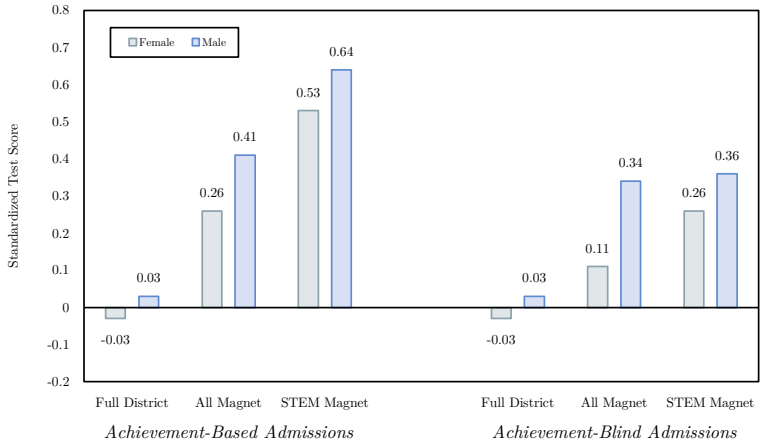
One possible explanation for why students may reject offers from programs that they choose to rank is that new information about programs is revealed to students between each stage of the application-enrollment cycle, as in Walters (2018). Similarly, students' idiosyncratic tastes may change between 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.

In our final 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 3 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

Figure 3: 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 3a displays changes between policies on the basis of race and Figure 3b 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.

policies. Figure 3a (Figure 3b) 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.

The descriptive findings in this section point to large changes in the composition of students in magnet programs due to the policy change. 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 identification strategy and findings.

3 The Causal Effects of Magnet Program Admissions

In order to understand the policy's impact on the district-wide distribution of school quality, we must first construct a credible measure of school quality. We follow the economics literature and consider school-level value-added (VA) as our measure of school quality. In this context, "value-added" refers to the gains in 9th grade test score performance a particular school provides to its students relative to the average school in the district. Let $i = 1, ..., N$ index the set of 9th graders and $j = 1, ..., J$ index the set of magnet programs and traditional high schools.

The conventional VA 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 $\sum_{j=1}^{J} d_{ijt} = 1$, $\widetilde{\theta}_{jt}$ is the conventional VA for school j in year t, and ε_{it} is an idiosyncratic error with mean zero. This model provides a simple framework to measure the causal effect of attending school j relative to the average school in the district.

In general, VA models for teachers and schools have been shown to exhibit near-unbiasedness for underlying measures of quality, especially after controlling for a sufficient statistic of students' incoming ability levels (Chetty, Friedman and Rockoff, 2014; Angrist et al., 2017, 2024). However, a causal interpretation of the estimate for θ_{jt} in Equation 1 requires (quasi-)experimental variation that eliminates selection bias unexplained by the controls. In school choice settings with centralized assignment, researchers often exploit random tie-breaking rules to instrument for selective enrollment.[12] We follow this strategy and

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

leverage the random assignment induced by magnet program admissions lotteries in the post-period.[13]

Our setting also features *unordered treatment* from the fact that there does not exist a global ranking of alternatives shared by all families in the district. Kirkebøen, Leuven and Mogstad (2016) show that in these types of settings, instrumental variables analyses that do not control for idiosyncratic preferences yield estimates that carry no meaningful interpretation.[14] Therefore, we further adjust for student-level preferences by leveraging our rich application-level data. One way to make this adjustment is to control for the submitted rank-ordered lists (ROLs) directly; however, this high-dimensional approach is infeasible in our setting given that we observe 385 unique application types across 5,739 students. Instead, we follow Abdulkadiroğlu et al. (2017) and construct magnet program assignment propensity scores that account for both the randomization induced by lottery admissions and the preferences that students have over programs. This strategy drastically reduces the conditioning set in our estimation.

3.1 The Propensity-Adjusted Value-Added Model

Unlike centralized assignment settings where students must submit ROLs to enter into 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 address this channel of self-selection into the magnet program lottery pool.

Students who enter the pool have stated 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 students belonging to marginalized groups through mandatory lottery quotas. As a reminder, all magnet programs, when possible, were required to allocate at least 20% of their seats to URM students. In addition, STEM programs were required to allocate 50% of their seats to female students. Let x_i^{URM} (x_i^f) indicate whether or not student i self-identifies as URM (female) and denote by z_{ij} an indicator for whether student i receives an offer from magnet program j through the random lottery procedure. We define a student's *applicant type* as $\tau_i \equiv (\succ_i, x_i^f, x_i^{\text{URM}})$, which captures both preferences from students and priority groupings from the district.

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

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

[13] In the district we study, lotteries are conducted exclusively on the basis of total demand, measured by the collection of ROLs that rank a program either first or second.

[14] We provide an illustrative example in Appendix C.

which captures the *ex ante* probability that student i will receive an offer from j when ranking it r^{th} on their ROL. For any j that i does not rank, $p_{ij,r} = 0$. Notably, all magnet programs in our setting are oversubscribed, which eliminates any possible degenerate propensity scores that take $p_{ij,r} = 1$ for all i and r. In practice, we obtain the propensity scores for each student that applies to magnet programs by simulating the district-wide lottery procedure one million times, holding fixed students' preferences, school capacities, and the district's lottery quotas.

Because we are interested in the possible distributional effects of the policy across the district, it is imperative that we recover credible estimates of VA for *both* traditional high schools and magnet programs. Therefore, we cannot rely on an IV design that leverages random lottery offers because traditional high schools do not admit students via lottery. In other words, an IV approach that leverages randomization in offers would recover causal estimates for magnet program quality alone. Furthermore, the recovered values of VA are relative to the average, meaning the coefficients obtained in an IV model would be relative to the average magnet program and could not be extrapolated to traditional high school quality. We therefore estimate "propensity-adjusted" VA models in the style of Angrist et al. (2024) that control for the *ex ante* likelihood of offer assignment and students' self-selection into the lottery pool.[15]

Our estimating equation thus takes the following form,

$$y_{it} = \alpha y_{it-6} + \boldsymbol{\gamma}' \boldsymbol{x}_{it} + \sum_{j=1}^{J} \theta_{jt} d_{ijt} + g(p_i) + \varepsilon_{it}, \tag{3}$$

where $g(\cdot)$ is a flexible polynomial and p_i is the stacked vector of program-specific assignment propensity scores and a set of indicators for $p_{ij} = 0$ at each j. The inclusion of the propensity scores controls for both the *ex ante* assignment probability and students' preferences for programs, while the set of indicators mitigates additional bias that influences enrollment effects unexplained by the included propensity scores. The parameter of interest then becomes θ_{jt}, the propensity-adjusted VA of program j. Note that where Equation 1 included a one-year lagged math test score to proxy for ability, we include in Equation 3 a six-year lagged test score due to data availability.[16] In our preferred specification, the controls in $\boldsymbol{x}_{it}$ include race, gender, and lagged reading scores. Importantly, estimating Equation 3 yields J estimates of VA, one for every school in the district.

[15] Angrist et al. (2024) call this the "risk-adjusted" model.

[16] Students in Toronto typically take standardized test scores in grades 3, 6, and 9, but sixth grade exams were canceled for our analysis cohort due to the Covid-19 Pandemic.

3.1.1 The Decomposition of Value-Added

Ultimately, we would like to understand the role of endogenous peer effects on program quality, especially as we look toward conducting counterfactual policy experiments that flexibly allow for general equilibrium effects. To isolate the contribution from peers, we decompose VA in a manner similar to Allende (2019). Let $\boldsymbol{m}_{jt}$ denote the vector of demeaned school-level average characteristics of 9th grade peers, including the fraction of URM students, female students, and high ability students, defined as having an above-average lagged math score. Let $\boldsymbol{x}_j$ denote the vector of observed non-peer education inputs, including indicators for magnet program type and school building conditions, proxied for by the demeaned number of health infractions recorded in the most recent Toronto School Health & Safety workplace inspection.

After estimating Equation 3, we treat the recovered VA as data and estimate regressions of the form

$$\theta_{jt} = \boldsymbol{\theta}_{\boldsymbol{m}}' \boldsymbol{m}_{jt} + \boldsymbol{\theta}_{\boldsymbol{x}}' \boldsymbol{x}_j + \sum_{k=1}^{3} \sum_{\ell=1}^{4} \theta_{k,\ell} (m_{jkt} \cdot x_{j\ell}) + \xi_j, \tag{4}$$

where ξ_j is a residual quality that is unobserved by the econometrician but observed by students. By definition, ξ_j is independent of $(\boldsymbol{m}_{jt}, \boldsymbol{x}_j)$ and has mean zero. The coefficients $\{\boldsymbol{\theta}_{\boldsymbol{m}}, \boldsymbol{\theta}_{\boldsymbol{x}}\}$ capture the relative contribution to VA from peer effects and non-peer effects in isolation, while $\{\theta_{k,\ell}\}_{k,\ell}$ captures the additional influence brought on by possible complementarity between peer characteristic m_k and non-peer characteristic x_ℓ.

We make two assumptions about this VA production function in order to trace out changes in school quality between policies. First, we assume that the coefficients obtained by estimating Equation 4, as well as the corresponding functional form, are policy-invariant. In other words, changes to school quality 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 are indexed by t. Due to the fact that our analysis compares school quality between one year and the next, this assumption allows us to directly relate changes in school quality 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.

Under these assumptions, we can write the change in school quality ($\Delta\theta_j$) as a function of the change in peer characteristics ($\Delta\boldsymbol{m}_j$),

$$\Delta\theta_j = \boldsymbol{\theta}_{\boldsymbol{m}}' \Delta\boldsymbol{m}_j + \sum_{k=1}^{3} \sum_{\ell=1}^{4} \theta_{k,\ell} (\Delta m_{jkt} \cdot x_{j\ell}). \tag{5}$$

We highlight that this specification of the VA production function allows the policy change to have an effect beyond "reshuffling" students according to a zero-sum game. In particular, by allowing different types of students to have distinct productivity levels at different types of schools, we allow for the possibility that changes to student sorting lead to net gains or losses across the district. A model without input complementarity such that $\theta_{k,\ell} = 0$ for all k, ℓ combinations will, by construction, generate net zero impacts on school quality, holding fixed the district-wide distribution of student characteristics.

3.2 Results

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

	Student Demographics			Non-Peer Characteristics			
	Frac. URM (1)	Frac. Female (2)	Frac. High Ability (3)	Art (4)	STEM (5)	IB (6)	Building Condition (7)
Baseline Effect	−0.004	0.005	0.020	0.289	0.174	0.482	−0.001
	(0.002)	(0.002)	(0.003)	(0.088)	(0.203)	(0.128)	(0.000)
× Art	−0.007	−0.008	−0.017	—	—	—	—
	(0.007)	(0.005)	(0.004)				
× STEM	−0.016	0.008	−0.017	—	—	—	—
	(0.012)	(0.004)	(0.010)				
× IB	−0.001	−0.028	−0.028	—	—	—	—
	(0.018)	(0.018)	(0.008)				
× Building Condition	−0.000	−0.000	0.000	—	—	—	—
	(0.000)	(0.000)	(0.000)				

Notes: Student demographic characteristics are demeaned across district. URM refers to "underrepresented minority." Building condition refer 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. Regression estimates are presented with robust standard errors below in parentheses.

We present the condensed estimation results of both Equation 1 and Equation 3 in Appendix Table A1. In general, we find that VA estimates are robust to our specification, with the correlation between the conventional and propensity-adjusted model estimates found to be 0.83 for the subset of magnet programs. Appendix Figure B5 further shows that VA is an important component of students' demand for magnet programs. An increase in school-level VA by one standard deviation is associated with an increase in the rate of oversubscription by 120%.[17]

Table 3 presents the results of the VA decomposition (Equation 4).[18] These results reveal two important insights about the production of school quality. First, peer effects capture a large and significant portion of school quality. An increase in the fraction of high ability students at any given traditional high school by one percentage point is associated with an increase in VA by 0.02σ on 9th grade math test scores.

[17] This is estimated from a univariate regression with estimated constant 3.04 and coefficient 3.80.
[18] Results from this exercise under alternative specifications of the VA production function are reported in Appendix Table A2.

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

			Magnet Program Type		
	All Schools	Traditional HS	Art	STEM	IB
	(1)	(2)	(3)	(4)	(5)
Panel A: Input Complementarity					
Propensity-Adjusted VA	0.006	0.079	−0.178	−0.278	0.073
	(0.022)	(0.018)	(0.041)	(0.065)	(0.086)
Conventional VA	0.005	0.075	−0.235	−0.342	0.343
	(0.028)	(0.017)	(0.051)	(0.107)	(0.171)
Panel B: No Input Complementarity					
Propensity-Adjusted VA	−0.015	0.060	−0.178	−0.235	−0.136
	(0.018)	(0.014)	(0.060)	(0.054)	(0.067)
Conventional VA	−0.015	0.069	−0.196	−0.258	−0.155
	(0.021)	(0.016)	(0.073)	(0.063)	(0.077)

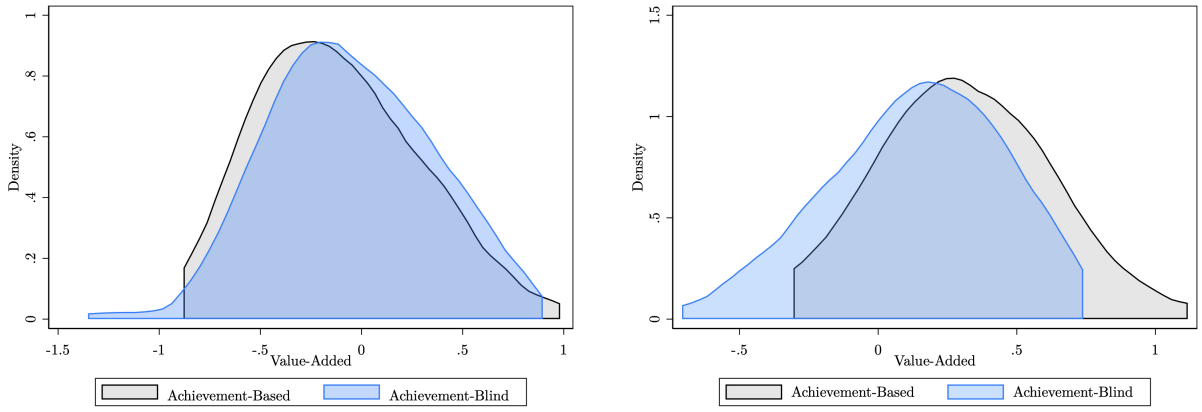
NOTES: The above table displays the relative change in value-added (VA) from the achievement-based policy to the achievement-blind policy, calculated as $\theta_{2024} - \theta_{2023}$ for θ_t the estimated VA. The first row compares estimates using the propensity-adjusted specification for our VA model, while the second row presents the same differences using estimates obtained by the conventional VA 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.

The same increase among the fraction of female students is associated with an increase in VA by 0.005σ, while this increase in URM students is associated with a decrease in VA by 0.004σ. Our large, albeit somewhat imprecise, estimates of non-peer effects reveal important levels of baseline influence for each type of magnet program independent of peers, as well as sizable negative effects induced by increases in school-wide health infractions.

Second, we find evidence of input complementarity between peer components and non-peer components. In particular, we find that high ability students are most effective at traditional high schools and less effective at each type of magnet program. For example, a one percentage point increase in high ability students at an art program is associated with only a 0.005σ increase in VA (a 75% decline relative to the baseline). This finding suggests that high ability students offer sizable spillovers for lower ability peers. We do not find any evidence to suggest that URM students are more or less productive at magnet programs relative to traditional high schools, but we do find that female students are more productive for their peers at STEM programs.

With the VA production function in hand, we can estimate the effect of the policy change on overall school quality. To do so, we substitute into Equation 4 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 VA production function with education input complementarity, whereas Panel B relies on a model without any interactions between peer and non-peer

Figure 4: Change in the Distribution of Value-Added



(a) Traditional High Schools

(b) Magnet Programs

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. Panel (a) displays this change among the set of default high schools in Toronto while panel (b) displays the same for the set of magnet programs. Value-added is measured in standard deviations of standardized 9th grade math test scores (σ).

characteristics. In each panel, we present findings under two different specifications for estimating VA: the preferred model that addresses preference heterogeneity and selection bias, as well as the "conventional" model that addresses neither of these concerns.

Estimates using our preferred model show that the policy acted as a school quality equalizer. Traditional public high schools saw a rise in quality that aligned with a decline in STEM and art magnet program quality. We find no detectable effect on IB programs. The results are qualitatively similar if we omit propensity score controls from the model, except for the fact that IB programs are found to improve. A model without input interactions produces similar findings, but with large declines across all magnet program types and comparably muted gains among traditional high schools.

Figure 4 graphically displays these shifts in the distribution of VA among traditional high schools (Panel A) and magnet programs (Panel B). On average, the elimination of admissions screening and the introduction of the lottery quotas led to an increase in VA at traditional high schools by 0.08σ. At the same time, magnet program VA declined by 0.17σ on average.

While these effects point to the impact of abrupt changes in student sorting on the distribution of school quality, it remains to be understood which students gained or lost the most as a result of the policy.

4 An Equilibrium Model of School Choice

In order to measure student welfare and estimate the effects of counterfactual policies, we develop a static equilibrium model that considers students' strategic application choices for secondary schooling and magnet programs' strategic admissions decisions. We develop the model under both the achievement-based admissions policy and the achievement-blind admissions policy. This equilibrium analysis, in conjunction with the reduced-form analyses of Section 3, is necessary for several reasons, which we detail below.

First, the structural model can measure changes in student welfare that the reduced-form estimates alone cannot address. To see this, note that even if the average magnet program experiences a decline in VA, it may be the case that average student welfare improves because students value VA less than other characteristics of a school, such as the fraction of URM students in a program. Furthermore, the reduced-form framework does not allow us to generalize to a wider class of admissions policies. The equilibrium model we develop provides flexibility to test for the causal effects of a continuum of different counterfactual policies using *either* admissions regime as the baseline. Finally, and perhaps most importantly, the initial analysis provides only partial insight into our main research question, namely, how should an admissions system be designed in order to mitigate the trade-off between equity and efficiency? We address this question directly using the estimated structural model.

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 j. Although magnet programs are often 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 two or fewer magnet programs of their choosing. Let A_{ij} be an indicator for whether student i submits an application to program j. Then $A_i = (A_{i1}, ..., A_{iJ^M})$ is the portfolio of application decisions that satisfies the district's application capacity, $\sum_{j \in J^M} A_{ij} \leq 2$.

Schools and programs are defined by their observed and unobserved characteristics. In particular, j

is characterized by its value-added, θ_{jt}, its location ℓ_j, a vector summarizing the average demographic attributes of its student body ($\boldsymbol{m}_{jt}$), and indicators for magnet program type ($\boldsymbol{x}_j$). We decompose school quality according to Equation 4, i.e., $\theta_{jt} = \theta(\boldsymbol{m}_{jt}, \boldsymbol{x}_j, \xi_j)$. This specification accounts for endogenous student sorting by mapping changes in peer composition directly to changes in VA. Consistent with our reduced-form framework, we assume that only peer inputs are variable in the short-run. Finally, each school faces a capacity constraint c_j that denotes the total number of available seats.

Given this framework, we next detail the strategies employed by students and schools under both the original (achievement-based) and the new (achievement-blind) admissions design. In what follows, we drop time subscripts (t) for ease of readability.

4.2 The Student's Admissions Problem

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

$$u_{ij} = \alpha_i \theta(\boldsymbol{m}_j, \boldsymbol{x}_j, \xi_j) + \boldsymbol{\beta}_i' \boldsymbol{m}_j + \boldsymbol{\zeta}_i' \boldsymbol{x}_j - \psi D_{ij} + \varepsilon_{ij}^{\mathcal{O}} + \varepsilon_{ij}, \tag{6}$$

where $\theta(\boldsymbol{m}_j, \boldsymbol{x}_j, \xi_j)$ is school-level VA, $\boldsymbol{m}_j$ is a vector of average student characteristics (the fraction of female students, URM students, and high ability students, as measured by having above-average lagged math test scores), $\boldsymbol{x}_j$ 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 style of Walters (2018), 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 to allow for heterogeneity at the margin of several different student-level characteristics. In practice, we parameterize α_i as

$$\alpha_i = \alpha_0 + \alpha_1 x_i^f + \alpha_2 x_i^{\text{URM}} + \alpha_3 x_i^h. \tag{7}$$

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 is assumed 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 students to accept offers, we parameterize the distribution of $\varepsilon_{ij}^{\mathcal{O}}$ as normal with mean zero and standard deviation $\sigma_{\mathcal{O},i}$, where $\sigma_{\mathcal{O},i} = \exp(\sigma_{\mathcal{O},0} + \sigma_{\mathcal{O},1} x_i^f + \sigma_{\mathcal{O},2} x_i^{\text{URM}} + \sigma_{\mathcal{O},3} x_i^h)$.

For completeness, we normalize the utility received from the outside option (leaving the district's public school system, attending a private school in the district, attending a Catholic school in the district, etc.) 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 6.

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

$$v_i \coloneqq \max\left\{\widetilde{u}_{ih}, \varepsilon_{i0}\right\}, \tag{8}$$

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.

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

$$\max_{A_i \in \mathbf{P}(J^M;2)} \Pi(A_i, \mathbf{\Omega}_i), \tag{9}$$

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.

The portfolio problem, as written, does not take into account students' strategic decision-making behavior. 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.

With this in mind, we can rewrite Equation 9 recursively, which greatly eases the difficulty in solving

the model and explicitly accounts for students' strategic behavior. Following the notation in Section 3.1, let $\{p_{ij,1}, p_{ij,2}\}$ be the corresponding propensity scores for receiving an admissions offer from program j in the first or second round, respectively. Program j_2^* solves the student's second-choice application decision,

$$V_2(\mathbf{\Omega}_i) = \max_{j_2 \in \{J^M \setminus j_1\} \cup \{\varnothing\}} \Big\{ p_{ij,2}\widetilde{u}_{ij} - \kappa_{2,i}\mathbf{1}\{j_2 \in J^M \setminus j_1\} + (1 - p_{ij,2})v_i \Big\}. \tag{10}$$

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, $\kappa_{2,i}$ captures the marginal cost of submitting a second application under the decentralized system, with parameterization $\kappa_{2,i} = \kappa_{2,0} + \kappa_{2,1}x_i^f + \kappa_{2,2}x_i^{\text{URM}} + \kappa_{2,3}x_i^h$.

We iterate backwards one period to the initial stage and recast Equation 9 as

$$V_1(\mathbf{\Omega}_i) = \max_{j_1 \in J^M \cup \{\varnothing\}} \Big\{ p_{ij,1}\widetilde{u}_{ij} - \kappa_i\mathbf{1}\{j \in J^M\} + (1 - p_{ij,1})V_2(\mathbf{\Omega}_i) \Big\}. \tag{11}$$

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 "continuation value" of not receiving an offer (Equation 10). In the above, κ_i is the cost to entering the application pool and ranking the first program, which we parameterize as before, with $\kappa_i = \kappa_0 + \kappa_1 x_i^f + \kappa_2 x_i^{\text{URM}} + \kappa_3 x_i^h$. Similar to before, we say that choosing $\{\varnothing\}$ means choosing the expected reservation value and substitute v_i into both instances of $\widetilde{u}_{ij}$. To reduce our parameter space, we impose that κ_i and $\kappa_{2,i}$ are related by the scalar $\delta \in [0, 1]$ such that $\kappa_{2,i} = \delta \cdot \kappa_i$.

Achievement-Based Admissions. Although we do not observe application behavior under this regime, we do observe the set of students who attend each magnet program. During this period, magnet programs accepted students in a one-shot manner after receiving the final set of applications (see Equation 14). We assume that programs rank the set of applying students according to observed student quality. Unlike previously studied settings like Boston or New York City, these admissions decisions do not depend on a standardized entrance exam and instead encompass a more holistic array of student characteristics that predict achievement. Thus, we construct program type-specific quality measures for each student accord-

ing to

$$q_{ik} = \gamma_{0,k}\bar{y}_i^M + \gamma_{1,k}\bar{y}_i^R + \gamma_{2,k}x_i^{\text{URM}} + \gamma_{3,k}x_i^f + \upsilon_i \tag{12}$$

where $k \in \{\text{Art, STEM, IB}\}$, the γ_k terms capture the preference that a magnet program of type k has for a given component of a student's application, and υ_i is a residual quality component observed by programs but unobserved to the econometrician. We assume that this component follows a standard normal distribution. The $\bar{y}_i$ terms denote student i's lagged standardized test scores in math and reading, respectively.[19] We also include race and gender as characteristics possibly considered in the admissions process.

Let $\bar{q}_j$ be the endogenous marginal quality level that magnet program j accepts. Students receive an admissions offer whenever $q_{ij} \geq \bar{q}_j$ and are rejected otherwise. Under this regime, the student's application problem is written as

$$\max_{j \in J^M \cup \{\varnothing\}} \Pr(q_{ij} \geq \bar{q}_j)\widetilde{u}_{ij} - \kappa_{ij}\mathbf{1}\{j \in J^M\}. \tag{13}$$

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 13 then represents a discrete choice problem with J^M+1 alternatives.[20] Here, κ_{ij} is the marginal cost of applying to some program j, which is allowed to depend on the program itself.[21] We allow students to submit up to two applications.

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 their attending students. Thus, programs solve

$$\begin{aligned} \max_{I_j \subseteq I_j^A} \quad & \sum_{i \in I_j} q_{ij} \\ \text{s.t.} \quad & |I_j| \leq c_j \end{aligned} \tag{14}$$

where I_j^A is the set of students applying to program j. The assumption that programs maximize their *total* student quality guarantees that the constraint holds with equality 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 results in a cutoff rule for admitting students. If $|I_j^A| \leq c_j$, program j admits all of its applicants

[19] To maintain a comparable scale, we transform lagged scores into district-wide percentiles.

[20] Although students have the ability to apply to two programs under the original policy, we omit this due to data limitations and due to the fact that approximately 75% of applicants submitted one application during this cycle, according to the district. Our lack of application data during this period further prohibits us from identifying an application cost parameter.

[21] 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.

and $I_j^A = I_j$. If $|I_j^A| > c_j$, program j admits its top c_j applicants.

Achievement-Blind Admissions. Under the achievement-blind admissions scheme, the magnet program's problem (Equation 14) is replaced with the quota-based lottery procedure outlined in Section 2.2. All magnet 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 magnet programs must allocate a fraction $\lambda^f \in [0,1]$ of their seats to female students. Therefore, all magnet programs face the constraint

$$\sum_{i \in I_j} x_i^{\text{URM}} \geq \lambda^{\text{URM}} c_j, \tag{15}$$

where x_i^{URM} takes the value 1 if i is an underrepresented minority, and STEM programs face the constraint

$$\sum_{i \in I_j} x_i^f \geq \lambda^f c_j, \tag{16}$$

with x_i^f a female student indicator. In the baseline policy imposed by the district, $\lambda^{\text{URM}} = 0.2$ and $\lambda^f = 0.5$.[22] Under this scheme, programs have no ability to screen or exert discretion in admissions decisions. With one or both of Equation 15 and Equation 16 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 send offers out to students. Students that receive an offer from a magnet program also receive an updated admissions shock, $\varepsilon_{ij}^{\mathcal{O}}$. As in Walters (2018), this offer shock rationalizes the fact that some students reject admissions offers from programs that they rank. Among students that enter the application pool, this shock is realized at the program that a student receives an offer from, i.e., the j satisfying $z_{ij} = 1$.

Denote by $\mathcal{O}_i$ the offer set of student i. With their updated information set, students choose the optimal school that satisfies

$$j^* = \underset{j \in \mathcal{O}_i \cup \{h_i, 0\}}{\text{argmax}} \left\{ \alpha_i \theta(\widehat{\boldsymbol{m}}_j^E, \boldsymbol{x}_j, \xi_j) + \boldsymbol{\beta}_i' \widehat{\boldsymbol{m}}_j^E + \boldsymbol{\zeta}_i' \boldsymbol{x}_j - \psi D_{ij} + \varepsilon_{ij} + \varepsilon_{ij}^{\mathcal{O}} \right\}. \tag{17}$$

We allow for students' expectations over $\boldsymbol{m}_j$, denoted by $\widehat{\boldsymbol{m}}_j^E$, to differ from their application stage ex-

[22] 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.

pectations. Students that lose at all of their ranked programs' lotteries, as well as those who do not enter the magnet program application pool, choose between $\widetilde{u}_{ih}$ and the outside option.

4.5 The Formulation of Rational Expectations

Students formulate rational expectations over the endogenous components that govern their decision-making at each stage of the admissions process. In our model, these components are students' admissions probabilities and the peer characteristics relevant to utility (Equation 6). The admissions probabilities depend directly on the policy in place. During the achievement-blind policy, students' beliefs about their odds of admissions must be consistent with the propensity scores generated from the district-wide set of applications, $\{p_{ij,1}, p_{ij,2} \mid \mathcal{A}\}_j$.

Under the achievement-based admissions, the probability of admissions depends on the quality of each student in the set of students that apply, I_j^A. We assume that students have rational expectations about each program's marginally admitted student. Denote the quality of this marginal student at program j by $\bar{q}_j$ and rewrite the parametrization of quality (Equation 12) as $q_{ij} = \boldsymbol{\gamma}_j' \boldsymbol{x}_i + \upsilon_i$. Given our distributional assumption on υ_i, we can write the expected probability that i earns admission to j under the achievement-based policy as

$$\Pr(q_{ij} \geq \bar{q}_j) = 1 - \Phi\left(\frac{\boldsymbol{\gamma}_j'(\bar{\boldsymbol{x}}_j - \boldsymbol{x}_i)}{\sqrt{2}}\right). \tag{18}$$

In a similar manner, we can impose that students have rational expectations over the composition of students at each program when making admissions and enrollment decisions. We make the same equilibrium argument about peer characteristics when estimating preference parameters during the achievement-blind scheme.

4.6 Definition of Equilibrium

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

Definition 1 (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, $\{\boldsymbol{m}_j^A, \boldsymbol{m}_j^E\}_j$, 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 11) 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 17) and are induced by rational expectations over peer characteristics at the enrollment stage.*

Definition 2 (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,* $\{\boldsymbol{m}_j\}_j$*, 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 14);*
- *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 17).*

5 Estimation and Identification

Denote by $\boldsymbol{\Theta} \coloneqq \{\alpha, \boldsymbol{\beta}, \boldsymbol{\zeta}, \sigma_\varepsilon, \kappa, \delta\}$ the collection of 38 parameters that govern utility. We estimate $\boldsymbol{\Theta}$ and the offer shock standard deviation parameters ($\sigma_{\mathcal{O}}$) by solving the student's problem during the achievement-blind admissions policy, using the two-step procedure described in Section 5.1. We then estimate the supply side of the model using data from the achievement-based admissions policy (Section 5.2). This provides us with data-driven estimates for magnet programs' preferences for students ($\boldsymbol{\gamma}$).

5.1 Estimation for Achievement-Blind Admissions

In general, solving the student's problem involves a nested fixed point algorithm, with the outer nest encompassing the fixed point for students' beliefs over peer characteristics at each school and the inner nest encompassing the *ex ante* probability of admissions. We alleviate this computational burden by imposing the equilibrium observed in data. By doing so, we enforce that students' beliefs over peer characteristics and offer probabilities are consistent with the application-level student demographics and the propensity scores obtained in Section 3. In counterfactual analyses, we cannot impose these expectations and must

instead recover rational expectations via fixed point algorithms.[23]

5.1.1 Simulated Method of Moments – Application Stage

We can recover consistent estimates of $\boldsymbol{\Theta}$ by focusing on the student's application problem alone. The student's application problem also embodies the decision to enter the application pool or not, which allows us to include the full sample of students in the estimation procedure. Let M denote the set of moments generated from the application data and $M(\boldsymbol{\Theta})$ the corresponding set of moments generated by preference parameters $\boldsymbol{\Theta}$. We estimate $\boldsymbol{\Theta}$ via simulated method-of-moments (SMM) by solving

$$\widehat{\boldsymbol{\Theta}} = \underset{\boldsymbol{\Theta}}{\operatorname{argmin}} \Big(M - M(\boldsymbol{\Theta})\Big)' W \Big(M - M(\boldsymbol{\Theta})\Big), \tag{19}$$

with W a weighting matrix. Table A3 provides an overview of the 121 moments used to identify the 38 utility parameters. We estimate the corresponding standard errors via the delta method.

5.1.2 Simulated Method of Moments – Enrollment Stage

The above process gives us estimates for all preference parameters and application costs, but not the terms that govern the offer shock ($\sigma_{\mathcal{O},i}$). To obtain this, we solve the students' enrollment problem using a second SMM procedure, taking the preference parameters as known. Using students' preferences, we simulate an equilibrium of applications and conduct the district's lottery algorithm once. This provides us with simulated offers, which we hold fixed in estimation. We then match four moments to identify each term that composes $\sigma_{\mathcal{O}_i}$. These moments include the average fraction of students who accept their offer (the "compliance rate"), as well as the compliance rate for the subset of female, URM, and high ability students.

5.2 Estimation for Achievement-Based Admissions

Under our assumption that preferences are policy-invariant, we then recover the magnet program admissions weights ($\boldsymbol{\gamma}$ in Equation 12). To do so, we estimate the two-sided model using enrollment data for the 2022-2023 school year (the last year of the achievement-based policy).

[23] The algorithm works as follows: for the current values of $\{p_{ij,1}, p_{ij,2}, \widehat{\boldsymbol{m}}_j^A\}$, students solve their application problem (Equation 11), which generates a district-wide set of optimal applications, $\mathcal{A}^*$, with corresponding peer characteristics $\{\boldsymbol{m}_j^A\}_j$. Updating admissions probabilities directly would require us to take the current district-wide set of applications and simulate the lottery procedure thousands of times, which is computationally intractable. Instead, we project the lottery procedure into a linear probability model that depends on student characteristics, program indicators, and programs' relative rates of oversubscription. We show in Appendix D that this projection replicates the true data well and can therefore be used as an alternative to the computationally burdensome simulation procedure. We repeat this process iteratively until the propensity scores and peer characteristics converge.

Recall that students anticipate an offer from magnet program j with *ex ante* probability $\Pr(q_{ij} \geq \bar{q}_j)$. We lack application data for this cohort of students, but having the demand-side parameters in hand aids our ability to model the decision-making process of students and programs in tandem. In estimation, we allow students to submit up to two applications. Programs then give offers to students after solving their admissions problem. We repeat this iteratively until either each program's capacity is filled or no students remain in the application pool.

As before, we solve for $\boldsymbol{\gamma}$ via SMM. Let $\widetilde{M}$ denote the collection of moments generated by administrative enrollment data during the achievement-based year, and let $\widetilde{M}(\boldsymbol{\gamma})$ denote the same moments generated in simulations for a set of admissions weights $\boldsymbol{\gamma}$. We recover estimates that solve

$$\widehat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma}}{\operatorname{argmin}} \left(\widetilde{M} - \widetilde{M}(\boldsymbol{\gamma})\right)' \widetilde{W} \left(\widetilde{M} - \widetilde{M}(\boldsymbol{\gamma})\right), \tag{20}$$

and again recover standard errors for $\widehat{\boldsymbol{\gamma}}$ via the delta method.

5.3 Identification

We next describe the identification assumptions required for each portion of the estimation procedure. When estimating the preference parameters, we assume that the post-application shock does not impact the students' application decisions. Although students are forward-looking when ranking programs, they are not adjusting for anticipated discrepancies between the application stage and the enrollment stage. This assumption rules out the possibility that, e.g., students rank programs in part because there is greater certainty about which peers will accept offers.

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 value-added. It is further identified by the tendency of students at low VA default schools to apply for magnet programs. Heterogeneous preference parameters can similarly be identified from 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 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. 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 test scores, and how heavily they value reading scores versus math scores.[24]

The lack of application data during this period requires us to make an additional assumption about 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 be -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 B8 and its subsequent discussion, we show the robustness of our findings to this elasticity assumption.

We end 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 general 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.

[24] Identification of programs' preferences requires us to normalize the variance of students' unobserved quality to 1.

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 Figure B5, we find that students value a program's VA. At baseline, students would be willing to travel an additional 6 minutes to attend a school with one additional standard deviation of VA.[25] Female students value VA most of all and are willing to travel an additional 31 minutes for a school with one standard deviation higher VA. We also find strong evidence of homophily bias along every demographic margin that we examine. For example, the baseline student is willing to travel an additional 2.6 minutes for a school with a 1 percentage point higher fraction of female students, whereas female students are willing to travel one minute more for the same increase.[26] Similarly, students exhibit strong distaste for increases in URM peers, except for URM students, who value similar peers greatly.

Students of every demographic group exhibit strong preferences for magnet programs. At baseline, students are willing to travel at least 22 minutes to attend any type of magnet program. We find that female students have large preferences for STEM programs and that URM students exhibit strong preferences for all magnet programs. Finally, our baseline estimates suggest that it costs students 60 minutes on average to submit the first application during the achievement-blind regime and an additional 33 minutes for the second. We uncover significant heterogeneity in costs, with high ability students facing lower costs to submission and URM students facing the highest costs.

On the supply side, we find that programs have heterogeneous preferences for student attributes. Art programs have preferences for skills associated with students' lagged reading scores but distaste for skills associated with students' lagged math scores. STEM and IB programs both view these skills in the opposite direction, placing high value on math ability. All programs place a negative weight on URM status. We find no evidence that gender is a large consideration in decentralized admissions.

[25] School-level VA has mean 0 and standard deviation 0.35. We additionally transform VA to belong to the positive reals for the sake of measurement in utility. Our VA measure therefore approximately belongs to the range [0,1].

[26] The numbers displayed in Table 5 refer to going from a school with 0% of the student body having demographic characteristic x to one with 100% of the student body having that characteristic. This is due to the fact that each demographic characteristics m_j is measured continuously along [0,1].

Table 5: Parameter Estimates

Panel A: Utility Parameters				
	Baseline	× Female	× URM	× High Ability
	(1)	(2)	(3)	(4)
Value-Added (α)	17.78	88.81	−0.542	23.16
	(8.238)	(6.173)	(7.341)	(11.79)
Fraction Female (β^f)	264.29	97.83	49.60	−54.83
	(14.72)	(15.31)	(14.68)	(9.823)
Fraction URM (β^{URM})	−182.20	−59.76	297.36	−54.74
	(15.08)	(17.66)	(28.22)	(10.47)
Fraction High Ability (β^h)	189.90	−181.81	253.66	127.26
	(15.16)	(21.86)	(18.13)	(17.69)
Art (ζ^{Art})	21.70	26.86	95.69	17.21
	(3.738)	(7.076)	(2.789)	(4.494)
STEM (ζ^{STEM})	41.30	50.14	119.55	19.59
	(3.184)	(4.951)	(1.814)	(2.584)
IB (ζ^{IB})	72.48	45.02	50.86	5.418
	(2.001)	(5.358)	(6.503)	(2.152)
Building Condition (ζ^b)	−68.69	28.02	19.92	22.37
	(4.989)	(10.82)	(12.62)	(8.824)
Taste Shock Std. Dev. (σ_ε)	89.25			
	(1.887)			
Offer Shock Std. Dev. ($\sigma_\mathcal{O}$)	3.698	0.491	1.607	2.787
	(0.398)	(0.225)	(0.220)	(0.198)
First-Rank Application Cost (κ)	59.99	29.14	82.47	−8.024
	(1.435)	(1.875)	(2.751)	(0.870)
Second-Rank Cost Depreciator (δ)	0.554			
	(0.007)			
Panel B: Program Preferences				
	Art	STEM	IB	
	(1)	(2)	(3)	
Lagged Math (γ_0)	−10.09	11.61	18.84	
	(1.614)	(0.734)	(3.570)	
Lagged Reading (γ_1)	16.49	−14.99	−13.55	
	(2.778)	(0.993)	(2.560)	
URM (γ_2)	−10.21	−12.86	−5.782	
	(3.166)	(1.256)	(1.479)	
Female (γ_3)	0.205	−0.044	0.365	
	(0.627)	(0.633)	(0.347)	

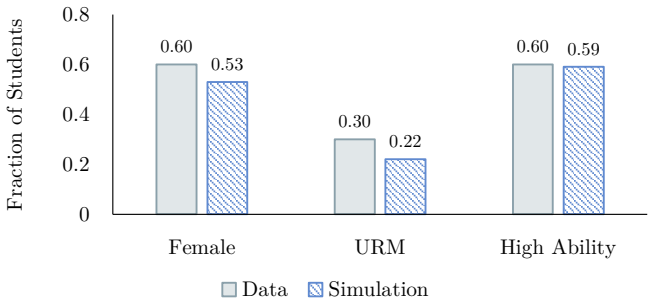
Notes: Parameter estimates come from estimating the corresponding equilibrium model. Preference parameters and application costs (demand-side parameters, Panel A) are estimated using application data on 16,810 students from the achievement-blind admissions scheme. This includes students outside of Toronto who apply but are not accepted. The offer shock variance is estimated from a subset of students in this data who apply to at least one magnet program (5,605 students). Supply-side program preference parameters (Panel B) are estimated using enrollment data on 15,795 students from the achievement-based admissions scheme. Standard errors are presented below each point estimate in parentheses.

6.2 Model Validity

The model replicates the data well, including targeted moments used in the SMM procedures and untargeted moments analyzed after estimation. Figure 8 presents a selection of untargeted comparison moments related to enrollment decisions during the achievement-blind regime. We simulate the full application,

lottery, and aftermarket enrollment using our estimated preference parameters. We then compare the demographic characteristics enrolled in magnet programs and find that the model matches the data. The administrative data shows that 30% of magnet program students are URM, whereas 22% of magnet students in our simulation are URM. We also underestimate female students' propensity to enroll in magnet programs, with 53% of magnet students in our simulation being female compared to 60% in data. However, the model matches high ability students' representation in magnet programs almost exactly. Appendix Figure B7 presents these same fractions for art, STEM, and IB programs in isolation. Overall, the model captures patterns in the data well.

Figure 5: Comparison of Untargeted Moments in Data and Simulation



NOTES: The above figure plots the realized representation of each demographic group in magnet programs during the achievement-blind admissions regime. Gray bars plot the fractions calculated in administrative data, whereas the slashed blue bars plot the simulated values recovered from 100 simulations.

6.3 Welfare Effects of the Policy Change

We next seek to determine the welfare impacts of Toronto's policy change. Ideally, we would use enrollment data from each year and our estimated preference parameters to recover the realized utility under each admissions regime. However, our data limitations make this an infeasible exercise, as we would systematically overstate welfare levels during the achievement-based policy because we cannot determine which students apply to programs and are rejected, or which students enroll in a program after submitting multiple applications. Instead, we use only the achievement-blind data and simulate outcomes under both policies.[27] This method has the added benefit of holding fixed the demographic distribution of students across the district, which pins down all of the resulting change in welfare directly to the policy change.

Table 6 presents the welfare changes caused by the adoption of the achievement-blind admissions policy. The first row displays the average change in utility level (across 100 simulations) for the listed

[27] Doing so requires us to implement a fixed point algorithm to maintain rational expectations of program-specific demographic shares during the "out-of-sample" achievement-based regime.

Table 6: Estimated Effects of Policy Change on Student Welfare

	All	URM	Non-URM	Female	Male	High Ability	Low Ability
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Δ Welfare	67.11	308.03	−18.00	92.80	46.55	−32.59	173.29
	(4.815)	(14.22)	(4.961)	(8.575)	(5.358)	(9.412)	(2.931)
95% CI	[57.68, 76.55]	[280.14, 335.92]	[−27.72, −8.275]	[75.99, 109.61]	[36.03, 57.04]	[−51.04, −14.14]	[167.54, 179.04]

NOTES: The above table displays the estimated changes in welfare using the achievement-blind administrative data and the set of estimated preference parameters. Standard errors are presented in parentheses, calculated directly as $SE = \sqrt{N^{-1}(\mathbb{V}[u(0)] + \mathbb{V}[u(1)])}$, with $u(0)$ the average utility under the original policy (simulated 100 times) and $u(1)$ that under the new one. 95% confidence intervals are derived from two-sample t-tests on the distribution of average utility. We mitigate the influence of outliers by dropping the highest and lowest 0.1% of utility values in each policy.

demographic group between the new policy and the original policy. We report calculated standard errors and 95% confidence intervals for the difference in means, derived from t-tests comparing the two distributions. We find the achievement-blind policy to be welfare-improving on average. Across all students, average utility increases by 67 travel minutes (38% increase) after eliminating admissions criteria. However, this overall effect masks significant treatment effect heterogeneity. URM students see a gain in welfare of 308 minutes (188% increase) while non-URM students see a decrease in welfare of 18 minutes (-10%). Furthermore, high ability students' average utility declines by 33 minutes (-10%) between policies.

To understand the extent to which these gains in welfare are driven by the changes in application costs, we repeat this exercise in a counterfactual that features uniform costs between policies. For this, we establish the decentralized costs to the estimated achievement-blind costs. These results are presented in Table A4. We find that the reduction in costs explains about 37% of the overall gains in utility. Under this specification, non-URM students see a reduction in welfare that is two times the decrease found with the adjusted costs. Similarly, welfare for high ability students is 22% lower between policies if cost changes are eliminated.

Figure B8 shows how the main results change under different values of assumed own-price elasticity. Recall that in the main results we impose students to have an elasticity of -2, meaning that art, STEM, and IB costs were 45%, 29%, and 30% higher under the achievement-based admissions scheme, respectively. If students are *more* sensitive to the change in application costs (i.e., more elastic demand) then welfare gains diminish on average and plateau around 30 travel minutes. Alternatively, if students are *less* sensitive to application costs, then we recover even greater welfare gains. However, we rule out unit elasticity (and therefore inelastic demand) due to the fact that unit elastic demand generates an equilibrium that is inconsistent with that found in data. Specifically, unit elasticity generates costs to art programs that are 163% higher compared to the achievement-blind policy, leading to fewer than 1% of students enrolling in

art programs. For these reasons, we argue that our results are characteristic of those generated under the feasible set of underlying application cost changes.

7 Counterfactual Analyses

7.1 Overview

The question of how to optimally design a district-wide admissions scheme for oversubscribed programs depends on what that district hopes to achieve and to what extent it prioritizes the possible equity-efficiency trade-offs that may arise as a result of the policy design. In what follows, we consider a set of competing objectives that a distract may have. For each, we allow the district to operate under *either* the original achievement-based scheme or the new achievement-blind scheme. In the former, the district chooses the admissions weights, $\boldsymbol{\gamma}$, to achieve some objective of interest. In the latter, they choose the value of the race and gender quotas, $\lambda \coloneqq \{\lambda^{\text{URM}}, \lambda^{f}\}$, for the lottery procedure.[28]

In an effort to avoid writing each problem twice, we shorthand these choice variables to $\boldsymbol{\Gamma}(\tau)$, where $\tau = 1$ denotes the achievement-blind regime and $\tau = 0$ the other. We lastly situate each of these findings against variations of the Gale-Shapley algorithm (deferred acceptance, or DA) to compare these mechanisms against an efficient, strategy-proof one.

7.2 The District's Possible Objectives

In order to quantify the equity-efficiency trade-off of a given policy, we must first define "equity" and "efficiency" in the context of public education. We consider two distinct definitions of each to encompass a possible misalignment that may exist between economists and policymakers. In particular, we allow "equity" to refer to equity in *opportunity* and equity in *outcomes*, and "efficiency" to mean maximizing either students' welfare or their achievement.

Equity in opportunity. A policy that seeks to maximize students' equity in opportunity is one that strives for parity in representation at each magnet program. To formalize this problem, we define a measure of segregation based on same group exposure, which is the probability that a member of a group meets another member of the same group, in a manner similar to Idoux (2022) and Margolis, Dench and Hashim (2020). To measure segregation within a program, we compare same group exposure in a program to the marginal distribution of that group in the district at large. We define the level of segregation at program j for group g as $S_{g,j} = SGE_{g,j} - P_g$, where $SGE_{g,j}$ is the same group exposure of group g at school j

[28] We continue to assume that the gender quota will only apply to STEM programs, while the race quota applies to all programs.

and P_g is the marginal distribution of group g across the school district. From this, we can construct the segregation index, $|S_{g,j}|$.

If groups are identically distributed across schools, same group exposure is equal to the marginal distribution of that group within the district. Therefore, the segregation index at a school can be interpreted as a measure of distance between the current student demographic composition and the demographic composition under perfect integration. Maximizing equity in opportunity is equivalent to minimizing the segregation index across magnet programs, which we write as

$$\min_{\mathbf{\Gamma}(\tau)} \sum_{j=1}^{J^M} \sum_{g} |S_{g,j}|. \tag{21}$$

Equity in outcomes. Alternatively, a policy that seeks to maximize students' equity in outcomes is one that seeks to minimize achievement gaps between students of different demographic groups. 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

$$\min_{\mathbf{\Gamma}(\tau)} \sum_{g} |\overline{Y}_g - \overline{Y}_{\not g}|. \tag{22}$$

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.

Efficiency (welfare). From an economic perspective, the most efficient allocation of students to programs may be such that student welfare is maximized. This would account for differences in students' preferences, although two-sided matching efficiency is unlikely to be achieved under the achievement-blind mechanism because programs have no discretion to admit students according to their own preferences. In either case, we can define the problem of the district that seeks to maximize student welfare as

$$\max_{\mathbf{\Gamma}(\tau)} \frac{1}{N} \sum_{i=1}^{N} u_{ij}. \tag{23}$$

Efficiency (achievement). The drawback to the problem stated in Equation 23 is its limited policy-relevance. In particular, a district may seek to maximize student welfare, but the immeasurability of utility

makes achieving this policy directly infeasible. A policy-relevant alternative is to use measurable student outcomes as a proxy for student welfare. There are many choices for these outcomes, but we focus on 9th grade test scores given our data availability. Thus, the problem can be restated as

$$\max_{\mathbf{\Gamma}(\tau)} \frac{1}{N} \sum_{i=1}^{N} y_{ij}. \tag{24}$$

Again, this assumes that all students are weighted identically. An extension of this could include Pareto weights, but we omit this for simplicity.

Tracing out the equity-efficiency frontier. In reality, a social planner may consider the trade-off between both equity and efficiency jointly. Denote by $\omega \in [0,1]$ the weight assigned to equity and $(1-\omega)$ that assigned to efficiency. Suppose that the district seeks to balance some measure of equity (E) and some other measure of efficiency (F). Then the planner's problem can be written as

$$\max_{\mathbf{\Gamma}(\tau)} \left\{ \max_{\omega \in [0,1]} -\omega E + (1-\omega) F \right\}. \tag{25}$$

The joint solution $\{\mathbf{\Gamma}^*(\tau), \omega^*\}$ is the optimal admissions scheme and equity priority for the specified definitions of equity and efficiency.[29]

As written, Equation 25 does not account for scale differences in our measures of equity and efficiency.[30] To adjust for this, we benchmark each counterfactual outcome against the baseline measure observed in data to compute relative percent changes according to

$$\widetilde{E}_{\mathbf{\Gamma}(\tau)} = \frac{E_{\mathbf{\Gamma}(\tau)} - E_0}{E_0}; \quad \widetilde{F}_{\mathbf{\Gamma}(\tau)} = \frac{F_{\mathbf{\Gamma}(\tau)} - F_0}{F_0}, \tag{26}$$

where E_0 and F_0 are some equity and efficiency measures in baseline, respectively, and $E_{\mathbf{\Gamma}(\tau)}$ is the counterfactual equity level with policy parameters $\mathbf{\Gamma}(\tau)$. We can then rewrite Equation 25 as

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

In practice, we solve this eight times: once for each combination of equity and efficiency and once for each admissions regime.

Deferred Acceptance (DA) algorithm. To situate these counterfactuals against a strategy-proof mech-

[29] We approximate the solution to Equation 25 using a discretized version of ω, denoted by $\widetilde{\omega} \in \{0, 0.1, ..., 0.9, 1\}$. We find the solution for each selected weight, $\mathbf{\Gamma}^*(\tau; \widetilde{\omega})$, and then conduct a simple grid search to find which choice of $\widetilde{\omega}$ maximizes the objective function following the best response admissions scheme.

[30] Unadjusted differences in the magnitude of either measure will bias the optimality results in favor of the "larger" measure.

anism, we also consider the four goals under a scenario in which the district continues to use centralized admissions, but adopts the DA algorithm. We consider an "achievement-based" and "achievement-blind" version of the DA algorithm. Appendix E provides generalized details on this mechanism.

7.3 Results

We first focus on counterfactuals that use the achievement-blind regime as a baseline. For a given social weight assigned to the measure of equity, the district optimizes the equity-efficiency trade-off by setting quotas for URM and female students.

Figure 6 displays the optimal quota system for each possible objective in isolation. Panel (a) shows how each combination of magnet program quotas impacts equity in opportunity (Equation 21). We convert each cell into a rate of change relative to the baseline ($\lambda^f = 0.5$, $\lambda^{\text{URM}} = 0.2$). The optimum is highlighted by the dotted black line and is shown to be the policy the allocates 10% of seats to URM students and 80% of STEM seats to female students.

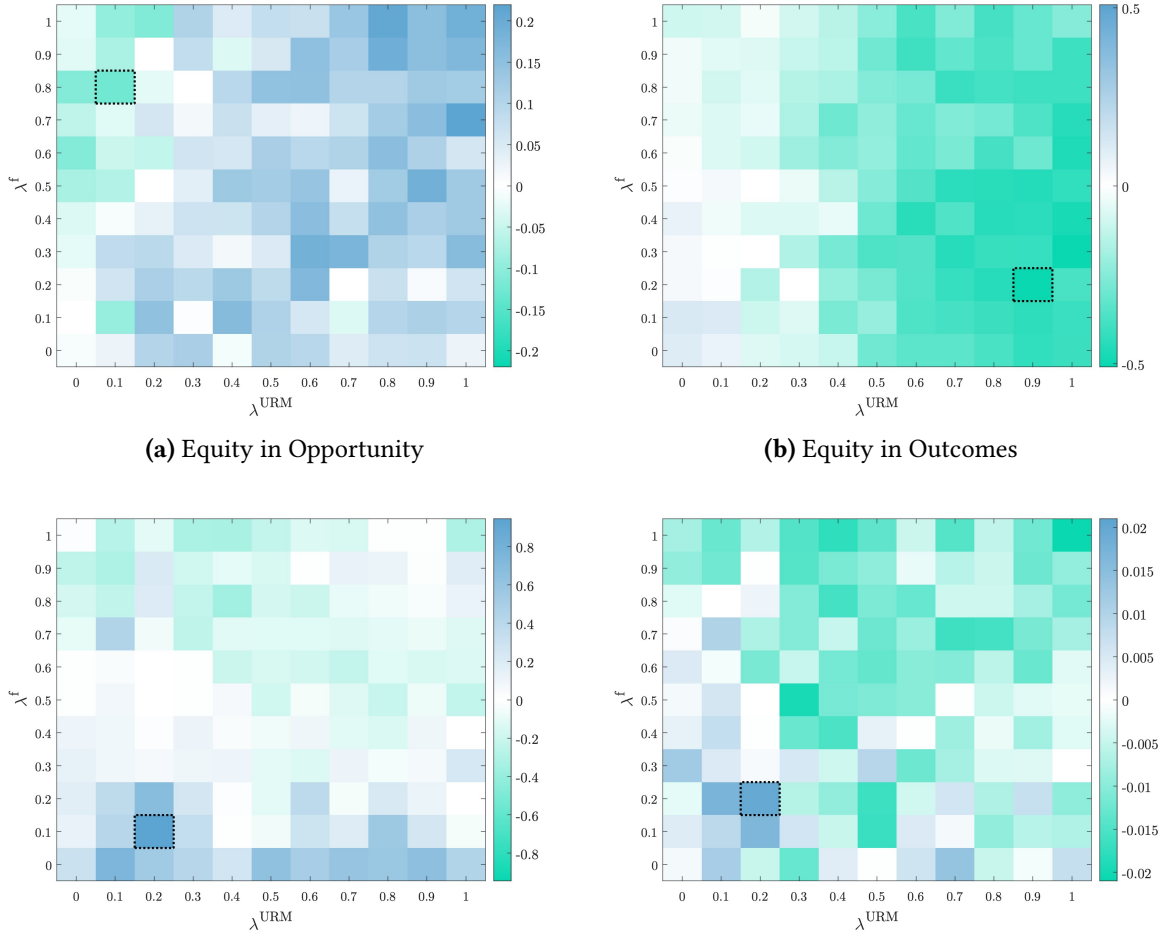
Panel (b) presents the effect of each quota on district-wide change in equity in outcomes (Equation 22). Here, equity is maximized by giving URM students nearly total priority in admissions (90% of seats) and allocating 20% of STEM seats to female students. This finding highlights the implications of using competing definitions of social parity. It also points to the magnitude of a policy intervention required to mitigate racial and gender gaps in 9th grade test performance.

Panels (c) and (d) provide heat maps for the two definitions of efficiency. The first considers the goal of maximizing average utility (Equation 23) and the second considers the goal of maximizing average student achievement (Equation 24). Again, the optimal policy depends on the district's stated goal. In order to maximize welfare under a lottery admissions scheme, the district would need to allocate 20% of all seats to URM students and 10% of STEM seats to female students. To maximize district-wide achievement, the planner would need to increase the female quota at STEM programs to 20%. It is worth highlighting that, on average, there is little scope for improving test score performance across district.

Under the original regime, programs admit students on the basis of observed achievement. For this, we hold fixed the value of ω and solve the district's problem (Equation 27) four times, one for each possible combination of equity and efficiency. The solution to the district's problem is the corresponding twelve-dimensional vector of program-specific admissions weights.

Figure 7 plots the trade-off between equity and efficiency for the four possible combinations of equity-efficiency definitions. Under each specification, we plot the social value achieved under a fixed social weight assigned to equity (ω) and the corresponding best-response quota design. In Panel (a), the maximum

Figure 6: Optimal Lottery Quotas



(a) Equity in Opportunity

(b) Equity in Outcomes

(c) Efficiency (Welfare)

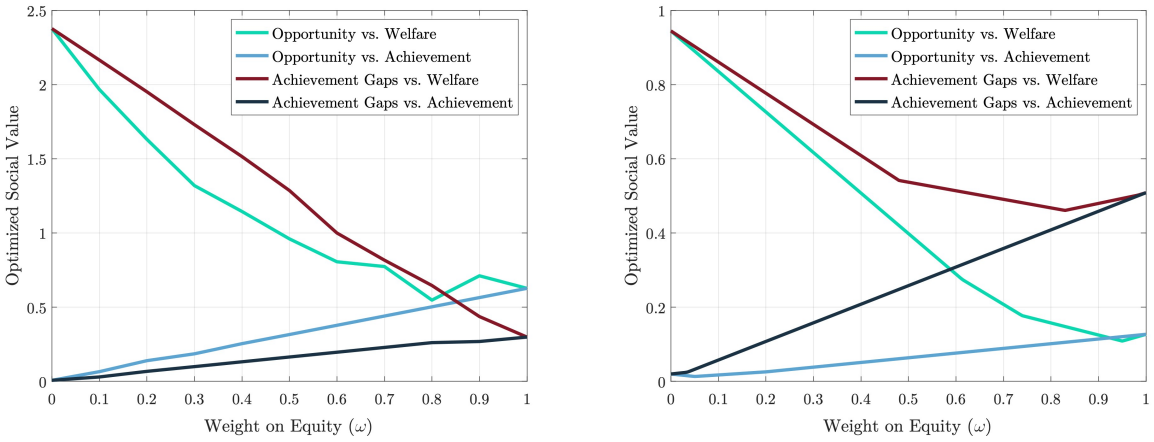
(d) Efficiency (Achievement)

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 *larger* percent changes in blue and *lower* percent changes in green. The solution to the district's problem is the cell contained by the black dotted line.

point along each line is generated by the optimum $\{\gamma^*, \omega^*\}$, whereas in Panel (b) the maximum is achieved by $\{\lambda^*, \omega^*\}$. Our results show striking similarities regardless of the policy structure in place. We find that optimizing the trade-off between equity in opportunity and welfare occurs when no weight is given to equity. We find that the trade-off between equity in outcomes (minimizing achievement gaps) and welfare is also maximized when all weight is given to welfare.

For the other two combinations, we find that the trade-off is optimized whenever all weight is given to the measure of equity (i.e., $\omega^* = 1$). There are therefore three possible optimal admissions designs for both the achievement-based and achievement-blind structure, depending on which definitions of equity and efficiency are used: (1) the $\mathbf{\Gamma}(\tau)$ that maximizes welfare when $\omega^* = 0$, (2) the $\mathbf{\Gamma}(\tau)$ that maximizes

Figure 7: Equity and Efficiency Trade-Off Frontier



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

NOTES: The above figure plots the social value of the equity-efficiency trade-off for a given combination of definitions of equity and efficiency and a given weight assigned to equity. Panel (a) plots the best response for program-specific admissions weights conditional on the stated weight toward equity. Panel (b) plots the best response for lottery quotas, again conditional on the stated equity weight. The maximum point along each line represents the maximum social value achievable for the given measures of equity and efficiency.

equity in opportunity when $\omega^* = 1$, and (3) the $\mathbf{\Gamma}(\tau)$ that maximizes equity in outcomes when $\omega^* = 1$. The optimal policies for each baseline regime are displayed in Appendix Table A5.

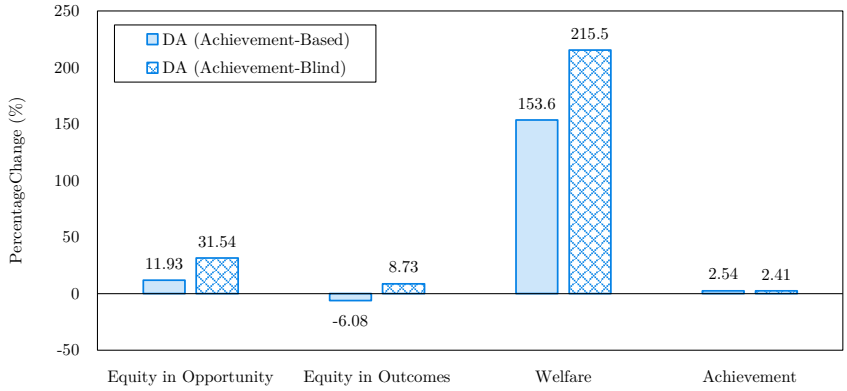
7.4 Improvements from Deferred Acceptance

We have shown that either policy implemented in Toronto has scope for improvement on the basis of either equity or efficiency. However, optimal versions of these admissions policies may still be suboptimal compared to other possible mechanisms. We investigate this by considering two separate versions of the Deferred Acceptance (DA) algorithm.[31] We first build off of the decentralized admissions scheme by holding fixed programs' preferences for student characteristics and embedding them into a DA algorithm. We then consider an "achievement-blind" version which shuts down programs' preferences and instead allocates offers via random, unweighted lotteries.

Overall, we find that the DA algorithm improves equity and efficiency, but that the achievement-blind iteration of the mechanism yields the most dramatic gains. District-wide welfare improves by more than two-fold compared to the original achievement-based regime. The DA algorithm also offers improvements to equity, with student representation increasing by 32% and achievement gaps decreasing by 9%. These findings point toward holistic gains for students that could be achieved if the district adopted a centralized,

[31] See Appendix E for generalized details. In these strategy-proof mechanisms, students consider their expected utility and not their likelihood of admissions. We continue to embed a fixed point algorithm over student attributes.

Figure 8: Equity and Efficiency Gains from Deferred Acceptance Algorithms



NOTES: The above figure plots the percentage change in each of the four equity and efficiency outcomes under the stated DA algorithm, relative to the simulated achievement-based scheme. "DA (Achievement-Based)" holds fixed the estimated supply-side parameters. "DA (Achievement-Blind)" shuts off the supply-side and instead allocates offers randomly.

strategy-proof admissions mechanism.

8 Conclusion

Public school districts across the world have begun eliminating screened admissions in an effort to increase representation among disadvantaged students. Despite their increased prevalence, these policies continue to draw controversy and remain understudied. This paper fills this gap by analyzing the equilibrium effects of a large change to admissions on students and schools. We focus on the setting of Toronto, where high school magnet program admissions abruptly changed from a decentralized, achievement-based design to a centralized, achievement-blind one. We show descriptive evidence that the policy dramatically improved the representation of URM students in magnet programs, but that the ability levels of incoming students declined as well.

We then estimate a value-added model to determine how the policy changed school quality across the district. We recover causal estimates of school-level quality by leveraging the randomization of lottery offers during the achievement-based policy. In a decomposition exercise, we find that peer effects explain a large portion of the overall effect of a given school. When students sort differently in response to the policy change, the distribution of school quality changes as a result. In particular, we find that magnet program quality declines on average while traditional high school quality improves on average.

To measure welfare and conduct policy-relevant counterfactual exercises, we develop and estimate a static equilibrium model of high school enrollment. Using our estimated model, we show that Toronto's policy was welfare-improving on average, but harmed non-URM students and high ability students. Ap-

proximately one-third of these gains can be explained by estimated reductions in costs associated with applying to programs. In counterfactual policy experiments, we show that both the original policy and the new one have scope for improvement, regardless of whether the district seeks to improve equity, efficiency, or both. We then quantify the equity-efficiency trade-off frontier under either admissions scheme and show how optimizing this trade-off depends entirely on which goals a district seeks to achieve. Lastly, we show that there are large efficiency gains to be made by adopting a deferred acceptance (DA) algorithm. We show that a district can maintain merit-blind initiatives and improve district-wide efficiency by adopting a DA mechanism.

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.

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.

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.

Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2014. "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates." *American economic review*, 104(9): 2593–2632.

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, 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. 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.

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.

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, Chris. 2013. "Targeted vouchers, competition among schools, and the academic achievement of poor students." *Documento de trabajo). Yale University. Recuperado de http://economics. sas. upenn. edu/system/files/event_papers/Neilson_2013_JMP_current. pdf.*

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.

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*.

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: VA Model Covariate Estimates

	Propensity-Adjusted VA Model (1)	Conventional VA Model (2)
Lagged Math	0.395	0.431
	(0.009)	(0.009)
Lagged Reading	−0.012	−0.015
	(0.009)	(0.009)
URM	−0.224	−0.233
	(0.017)	(0.016)
Female	−0.074	−0.076
	(0.013)	(0.013)
Observations	14,435	14,435

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. Column (1) displays results from the propensity-adjusted model that includes 28 propensity score controls and 28 zero-probability indicators. 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. Robust standard errors are reported in parentheses.

Table A2: Alternative Decomposition Specifications

	(1)	(2)	(3)
Art	0.221	0.171	0.289
	(0.073)	(0.117)	(0.088)
STEM	0.132	−0.089	0.174
	(0.101)	(0.085)	(0.203)
IB	0.262	−0.095	0.482
	(0.096)	(0.086)	(0.128)
Building Condition	−0.001	−0.001	−0.001
	(0.001)	(0.000)	(0.000)
Frac. URM Students		−0.008	−0.004
		(0.002)	(0.002)
Frac. Female Students		0.004	0.005
		(0.003)	(0.002)
Frac. High Ability Students		0.014	0.020
		(0.002)	(0.003)
URM × Art			−0.007
			(0.007)
URM × STEM			−0.016
			(0.012)
URM × IB			−0.001
			(0.018)
URM × Building Condition			−0.000
			(0.000)
Female × Art			−0.008
			(0.005)
Female × STEM			0.008
			(0.004)
Female × IB			−0.028
			(0.018)
Female × Building Condition			−0.000
			(0.000)
High Ability × Art			−0.017
			(0.004)
High Ability × STEM			−0.017
			(0.010)
High Ability × IB			−0.028
			(0.008)
High Ability × Building Condition			0.000
			(0.000)
Observations	95	95	95
R^2	0.091	0.695	0.768

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. Robust standard errors are reported in parentheses.

Table A3: Main Sources of Identification for Model Parameters

Model Parameter	Main Source of Identifying Variation
Preferences	
Value-Added (α)	Average VA of program $r = 1$ for demographic group X
Peer Characteristic (β)	Average peer characteristic of program $r = 1$ for demographic group X
Non-Peer Characteristic (ζ)	Average non-peer characteristic of program $r = 1$ for demographic group X
Taste Shock Std. Dev. (σ_ε)	Correlation in ranking preferences
Application Components	
First-Rank Cost (κ)	Fraction of demographic group X that enter application pool
Second-Rank Scalar (δ)	Correlations between first-rank choice and second-choice being blank
Offer Shock Std. Dev. ($\sigma_{\mathcal{O}}$)	Average propensity to accept an offer among demographic group X
Program Preferences	
Preference of Program Type k (γ_k)	Average accepted level of given attribute at program type k

NOTES: This table displays the generalized moments used to identify the corresponding parameters. We use 121 moments to identify 38 preference parameters and application costs, 4 moments to identify the offer shock variance, and 15 moments to identify the 12 admissions weights. In the above table, X is a placeholder that typically refers to some or all of the following: all students, URM students, female students, and high ability students. $r = 1$ refers to the collection of programs that are ranked first.

Table A4: Estimated Effects of Policy Change on Student Welfare with Fixed Costs

	All	URM	Non-URM	Female	Male	High Ability	Low Ability
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Δ Welfare	42.19	265.32	−36.38	61.61	27.55	−39.73	130.01
	(4.579)	(13.89)	(4.653)	(8.223)	(5.064)	(6.840)	(2.357)
95% CI	[33.21, 51.16]	[238.08, 292.55]	[−45.50, −27.26]	[45.49, 77.73]	[17.63, 37.48]	[−57.54, −21.91]	[124.92, 135.10]

NOTES: The above table displays the estimated changes in welfare using the achievement-blind administrative data and the set of estimated preference parameters, holding fixed the application costs to the achievement-blind ones. Standard errors are presented in parentheses, calculated directly as $SE = \sqrt{N^{-1}(\mathbb{V}[u(0)] + \mathbb{V}[u(1)])}$, with $u(0)$ the average utility under the original policy (simulated 100 times) and $u(1)$ that under the new one. 95% confidence intervals are derived from two-sample t-tests on the distribution of average utility. We mitigate the influence of outliers by dropping the highest and lowest 0.1% of utility values in each policy.

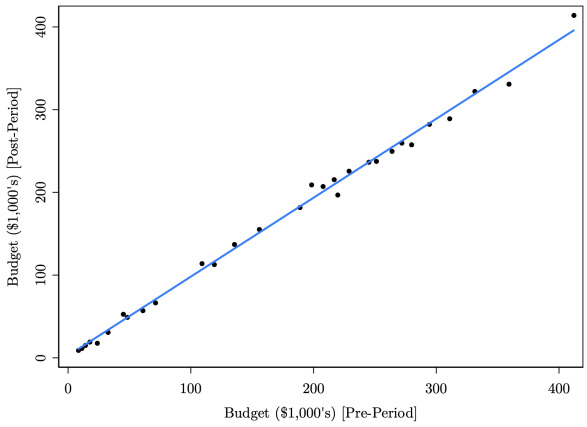
Table A5: Optimal Policy Design by Baseline Policy and Equity-Efficiency Combination

	Optimal Equity Weight, ω^*	Optimal Policy Design, $\mathbf{\Gamma}^*(\tau)$			
Panel A: Achievement-Based					
		Lagged Math	Lagged Reading	URM Status	Female Status
Opportunity vs. Welfare	0	{6, 4, 11}	{0, 1, 11}	{18, 13, 18}	{8, 8, −4}
Opportunity vs. Achievement	1	{18, 4, 11}	{14, 18, −8}	{6, 3, 6}	{6, 2, 2}
Achievement Gaps vs. Welfare	0	{6, 4, 11}	{0, 1, 11}	{18, 13, 18}	{8, 8, −4}
Achievement Gaps vs. Achievement	1	{−8, 12, 3}	{12, −7, 6}	{18, 19, 18}	{12, 0, 16}
Panel B: Achievement-Blind					
		URM Quota	Female Quota		
Opportunity vs. Welfare	0	0.1	0.8		
Opportunity vs. Achievement	1	0.9	0.2		
Achievement Gaps vs. Welfare	0	0.2	0.1		
Achievement Gaps vs. Achievement	1	0.2	0.2		

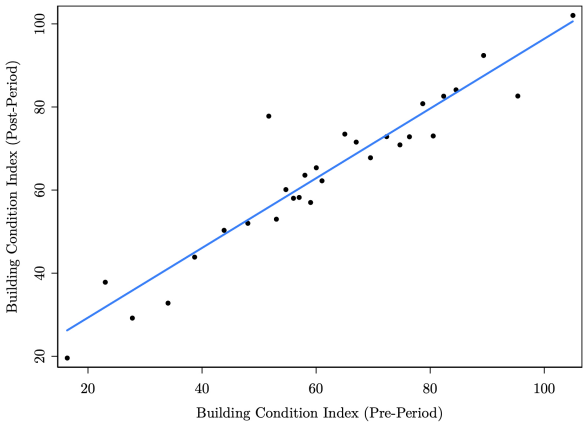
Notes: The above table displays the optimal admissions policy design for a given combination of equity and efficiency and for a given baseline policy. Panel A displays the decentralized, program-specific admissions weights that optimize the stated objective. The weights correspond to {Art, STEM, IB}. 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

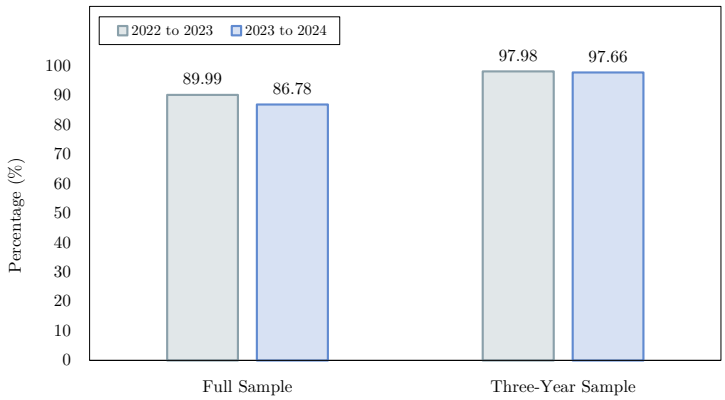
Figure B1: Relative Stability of Education Inputs



(a) Annual Preliminary Budget



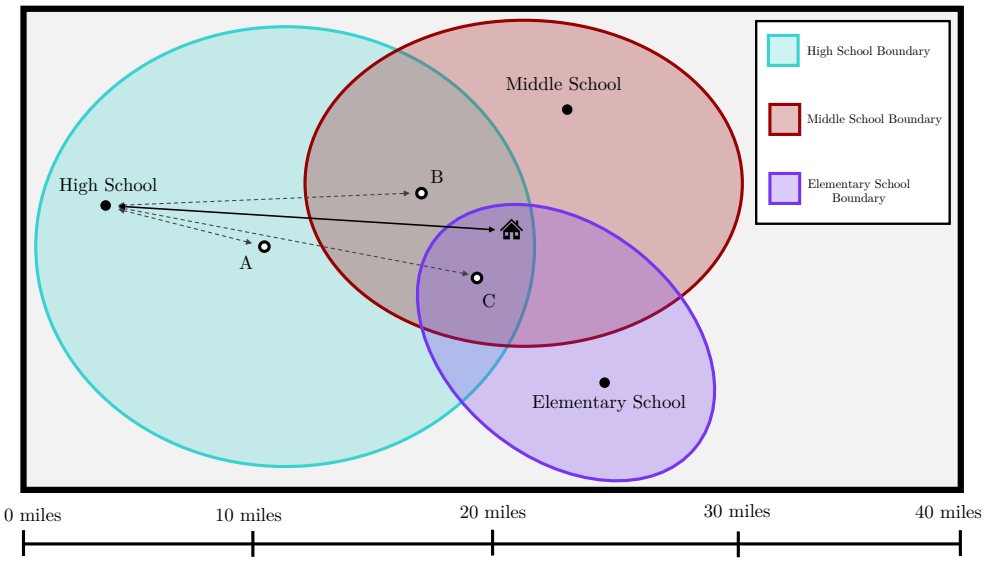
(b) Building Condition Index



(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 B7a displays these results for the annual school budget designated at the start of the academic year. Figure B7b displays the same for the district-assigned index for the school's building condition. The final panel (Figure B7c) 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

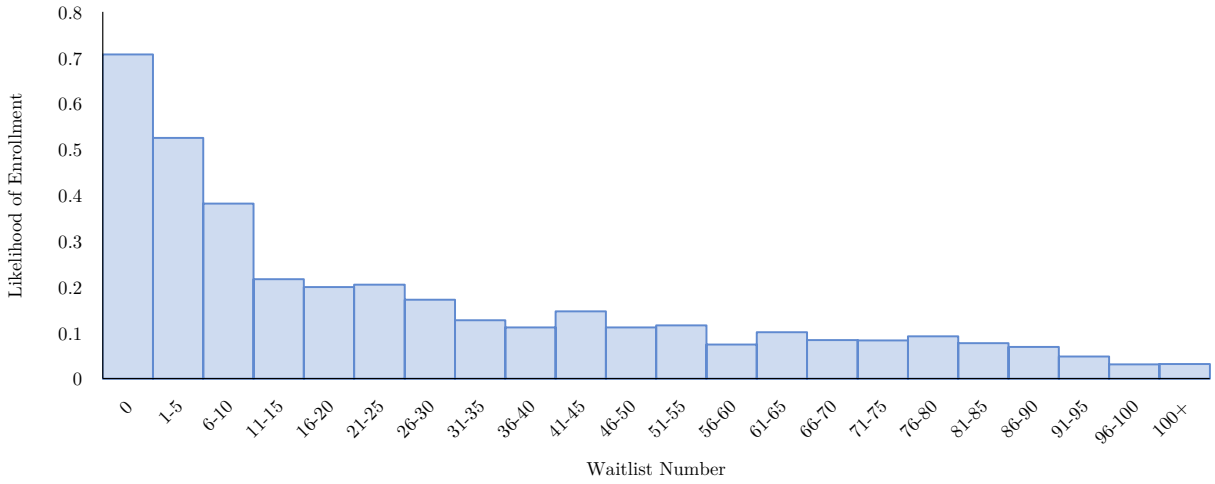


(a) 2022-2023 School Year

(b) 2023-2024 School Year

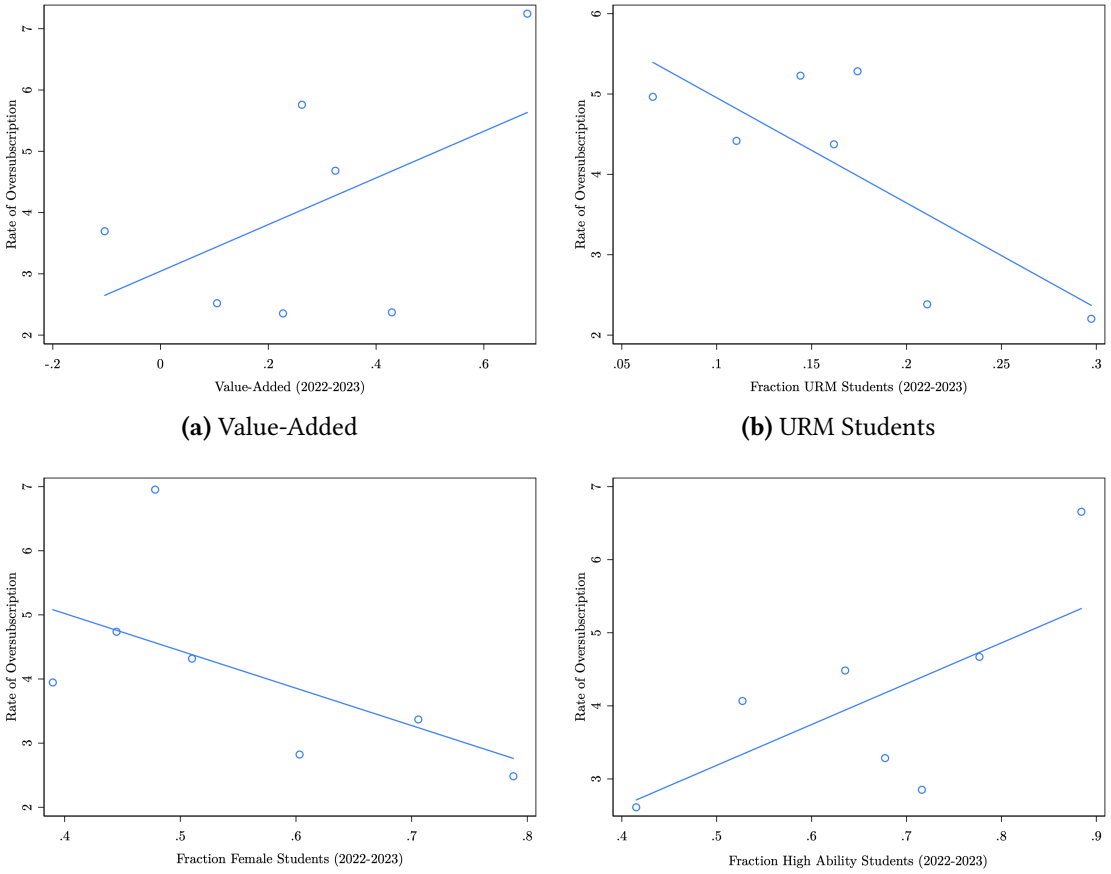
NOTES: The above figures display the home locations of 9th graders in the district, which we impute using information on default high school, middle school, and elementary school assignment, along with each school's corresponding catchment area. Out-of-district students are imputed as living on the border 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

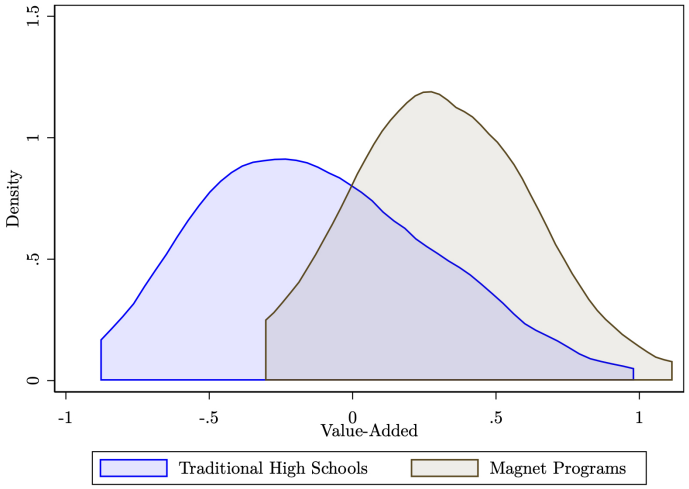


(a) Value-Added **(b)** URM Students

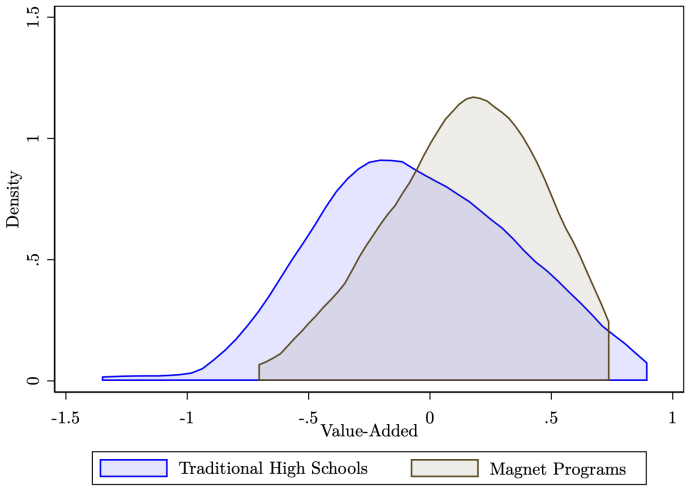
(c) Female Students **(d)** High Ability Students

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



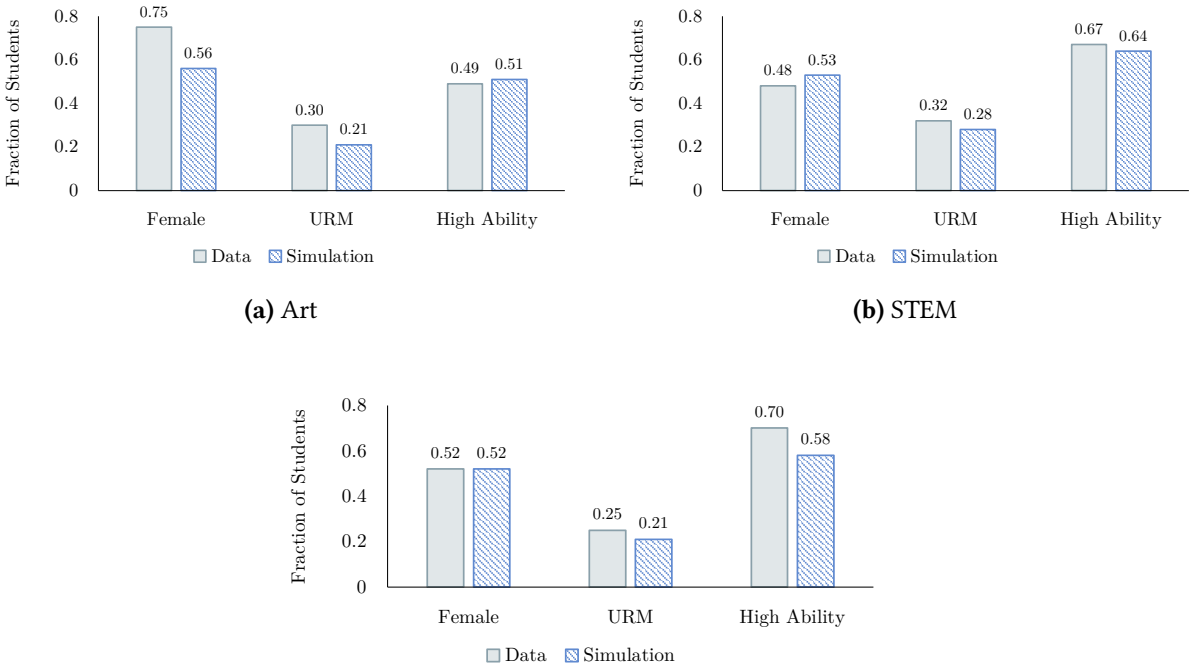
(a) Achievement-Based Admissions



(b) Achievement-Blind Admissions

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. Panel (a) displays the distribution of VA in the 2022-2023 academic year while (b) displays 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: Additional Untargeted Moments



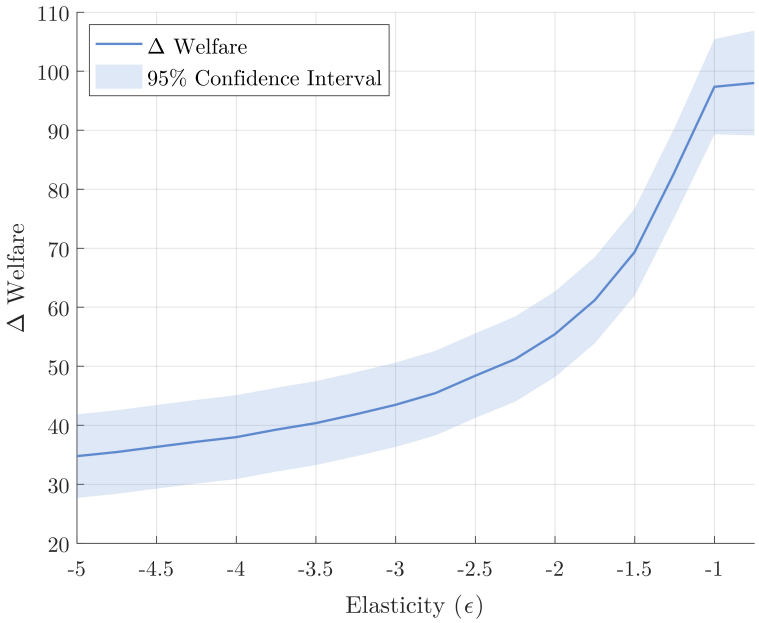
(a) Art

(b) STEM

(c) International Baccalaureate

Notes: The above figures display comparisons in demographic representation in each type of magnet program. Panel (a) presents the rate of female, URM, and high ability students in art programs, whereas Panels (b) and (c) present the same for STEM and IB programs. "Data" columns present raw means from the administrative data, whereas "Simulation" refers to the same statistics calculated in the model simulation.

Figure B8: Policy Impact on Average Welfare as a Function of Student Elasticity



NOTES: The above figure plots the average difference in utility between policies 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 B8

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.

C 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 C1 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] The district's design of the lottery guarantees that $\sum_j z_{ij} = 1$.

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

Table C1: 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).

D 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

$$\begin{aligned} p_{ij,1} &= \rho_0 + \rho_1 x_i^f + \rho_2 x_i^f \cdot x_j^{\text{STEM}} + \rho_3 x_i^{\text{URM}} + \rho_4 x_j^{\text{STEM}} + \rho_5 x_j^{\text{art}} + \rho_6 \left(\frac{N_{1,j}}{c_j}\right) + \rho_7 \left(\frac{N_{2,j}}{c_j}\right) + \epsilon_{ij,1} \\ p_{ij,2} &= \varrho_0 + \varrho_1 x_i^f + \varrho_2 x_i^f \cdot x_j^{\text{STEM}} + \varrho_3 x_i^{\text{URM}} + \varrho_4 x_j^{\text{STEM}} + \varrho_5 x_j^{\text{art}} + \varrho_6 \left(\frac{N_{1,j}}{c_j}\right) + \varrho_7 \left(\frac{N_{2,j}}{c_j}\right) + \epsilon_{ij,2} \end{aligned} \tag{D1}$$

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 D1. 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 D2 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 D1. 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.

Table D2: Results for the Determinants of Simulated Offers

		Simulations: URM Quotas			
	Original Data	$\lambda^{URM} = 0.2$	$\lambda^{URM} = 0.4$	$\lambda^{URM} = 0.6$	$\lambda^{URM} = 0.8$
	(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 D1 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.

The coefficient on URM increases, but other coefficients remain stable.

E Details of the Deferred Acceptance Algorithm

As a separate class of counterfactuals, we benchmark each of our objectives against a scenario in which the district implements a DA admissions mechanism. We consider two versions. 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, meaning that students do not consider their idiosyncratic probability of admissions when formulating their rank-ordered list (ROL). The "achievement-based" DA mechanism works in the following way:

1. Students submit an ROL comprised of their top two programs, where a student ranks program j first if both $u_{ij} - \kappa_{ij} > u_{ik} - \kappa_{ik}$ for all $k \in J^M \setminus j$ *and* $u_{ij} - \kappa_{ij} \geq v_i$, the student's reservation value.

2. Programs have a capacity c_j and consider the set of students that rank them first, denoted by $I_{j,1}$. Each student in this set has a corresponding program-specific quality q_{ij}. Programs order the set of first-choice students by quality and admit all students if $|I_{j,1}| \leq c_j$. If instead $|I_{j,1}| > c_j$, only the top c_j students are taken.

3. Students without a match are able to apply to their second-choice program. If a student only ranks one program, they are removed from the algorithm.

4. Programs receive the updated set of second-choice students, $I_{j,2}$. They then order the *total* set of applying students, $I_{j,1} \cup I_{j,2}$, by quality. The top c_j students are accepted. All students are accepted whenever $|I_{j,1} \cup I_{j,2}| \leq c_j$.

5. Students who were tentatively accepted as part of $I_{j,1}$ but lose their offer after the inclusion of $I_{j,2}$ students are then eligible to apply to *their* second-choice school. This forms the sub-sample of students $I_{j,3}$.

6. This process continues until either all students have an offer or no available seats remain.

It is worth noting that because we assume that quality is measured with an unobserved error, ties occur in our simulation with probability zero.