Sorting or Supporting? The Effect of Gifted Education on Achievement and Access*

Jimmy Chin† Geoffrey Kocks‡

July 12, 2026

Abstract

We evaluate the impacts of New York City kindergarten gifted and talented programs on academic trajectories using regression discontinuity and lottery designs. Gifted enrollment markedly changes the classroom environment but does not cause achievement gains. It significantly increases elite middle school enrollment, especially among low-income and higher-ability students. Using a structural model, we study a recent reform eliminating the gifted admissions exam. The reform increases representation of low-income and lower-ability students, decreasing the treatment effect on the treated for elite school access. Exam-based admissions lies closer to a possibilities frontier that trades off diversity and causal effects on school access.

*We thank Parag Pathak, Joshua Angrist, Amy Finkelstein, and Chris Walters for their help and encouragement throughout this project. We are grateful to the Office of Student Enrollment at New York City Public Schools, especially Daniel Hildreth and Josh Smith, for providing data and support. We further thank the NYC G&T teachers whom we met and who shared valuable insights with us. Thanks also to Katharina Adler, Claudia Allende, Isaiah Andrews, David Autor, Nano Barahona, Laura Caron, Christopher Cleveland, Sharada Dharmasankar, Laura Giuliano, Hilary Hoynes, Dean Li, Clemence Idoux, Scott Imberman, Pat Kline, Tomas Larroucau, Enrico Moretti, Ashesh Rambachan, Jesse Rothstein, Nina Roussille, Yang Song, Victor Quintas-Martínez, Connie Xu, Danny Yagan, and Yuci Zhou, as well as participants at the MIT Labor and Public Finance Lunches, 2024 AEFP Annual Conference, 2024 AEA CSQIEP Conference, 2025 ACLEC, the Boston University PREREQ Seminar, and the 2025 CESifo/ifo Junior Workshop on the Economics of Education for helpful comments. We are grateful to Eryn Heying, Jennifer Jackson, Niamh McLoughlin, and Jim Shen for invaluable administrative support. Finally, we thank Vivien Huang, Lauren Nieto, Renzl Obispo, and Edha Singh for excellent research assistance. This paper reports on research conducted under data-use agreements between MIT, the project's principal investigator, and New York City Public Schools. This paper reflects the views of the authors alone. We acknowledge generous funding from the MIT Integrated Learning Initiative, the NSF GRFP under grant 2141064 (Kocks), the NBER Global Math Talent Fellowship (Kocks), and Berkeley Economists for Equity.

†UC Berkeley Department of Economics. Email: jimmychin@berkeley.edu

‡MIT Department of Economics. Email: gkocks@mit.edu

1 Introduction

In school districts around the country, a perennial debate centers on whether students should be grouped into different classrooms based on ability. Gifted education, which provides specialized instruction to students with high cognitive ability, is a cornerstone of this debate. Supporters argue that gifted and talented (G&T) programs promote learning by tailoring instruction to student skills. This belief underpins the popularity of G&T and makes it a common tool for policymakers seeking to retain families in public schools (Loveless, 2013). Critics counter that G&T yields limited academic benefits while increasing segregation by race and class. Both sides also see G&T as shaping *access* to future education opportunities, which underscores its importance for long-run outcomes.[1] Reflecting this divide, school districts have adopted sharply different approaches to G&T (Goldstein, 2018). Seattle, for instance, began phasing out its gifted programs in 2024, while New York City expanded them (Bryan, 2024; Fadulu, 2022).

In this paper, we ask two research questions. We first ask if G&T increases academic achievement and access to elite middle and high schools. We then ask how a reform that replaces an entrance exam with teacher recommendation-based admissions affects later access to elite middle schools. Our setting is New York City (NYC), where G&T programs, until recently, grouped kindergarten students into separate classrooms and schools on the basis of an entrance exam.

The NYC setting is unique for several reasons. First, NYC G&T begins in kindergarten. Prior work mostly focuses on students identified as gifted in late elementary school, and little is known about gifted effects at such early ages. Early identification could yield larger gains if early childhood is a sensitive period for human capital development (Cunha and Heckman, 2007). Second, our setting provides variation to identify treatment effects for both marginal students at a qualifying cutoff and inframarginal students who score above it, distinguishing our approach from the typical focus on marginal students. Finally, debates surrounding G&T are especially salient in NYC. A prominent concern centers on G&T and school access. For example, the lawsuit IntegrateNYC v. New York (2024) alleges that G&T forms part of a "racialized pipeline" that excludes minority students from "prime educational opportunities," including the city's high-performing secondary schools (Closson, 2024).[2]

[1] Recent work also emphasizes social capital as a mechanism by which programs like G&T could impact long-run outcomes (Chetty et al., 2022).

[2] Qualitative research also suggests that parents perceive G&T as a "feeder for the better middle schools, which then seems to be a feeder for the better high schools" (Roda, 2015). In recent years, NYC published school quality reports that provide information on frequently attended middle schools for students at each elementary school, which we interpret as additional evidence that families care about school access. See for

To answer our first research question, we identify causal effects by exploiting two features of the exam-based G&T assignment system used prior to 2021. The first arises from a discontinuity in gifted eligibility. Students who scored above a cutoff on the entrance exam were eligible to apply for G&T, and we use a fuzzy regression discontinuity (RD) design to estimate the causal effect of G&T for marginal students at the cutoff. The second source of variation arises from the centralized G&T assignment algorithm, which used random lottery numbers to ration seats at oversubscribed programs. The lottery research design relies on the instrumental variables framework introduced by Abdulkadiroğlu et al. (2017). Estimates obtained using variation from the assignment process pertain to students across the full support of the exam score distribution above the eligibility cutoff. We use the lottery design to estimate how effects vary for students with different baseline scores. The lottery design also provides a unique test of RD external validity, for which extrapolation away from the cutoff requires additional assumptions (Angrist and Rokkanen, 2015).

Our causal effect estimates reveal that G&T enrollment does not boost test scores through 8th grade, despite large changes in the kindergarten classroom environment. G&T dramatically changes peer exposure and significantly reduces class size for some students. For example, in the RD design, G&T lowers class size by over 4 students and decreases the share of low-income classmates by 26 percentage points (as proxied for by free- and reduced-price lunch eligibility; FRPL). The null effect on achievement holds across both RD and lottery designs, as well as for subsamples defined by demographics and baseline exam scores (though effects on White and Asian students are possibly negative). In our most precise specification, which pools the RD and lottery samples across grades, the largest G&T causal effect estimate (in absolute value) for either math or English Language Arts (ELA) is -0.05σ and confidence intervals exclude effects larger in magnitude than roughly $\pm 0.1\sigma$. The findings on achievement are consistent with prior work on G&T (Bui et al., 2014; Card and Giuliano, 2014; Cleveland, 2023), though Card and Giuliano (2016a) find benefits for high-achieving non-gifted students who nonetheless enroll in G&T classes. Our confidence intervals exclude the $+0.5\sigma$ effect on minorities reported by Card and Giuliano (2016a).

We then estimate the causal effects of G&T on middle school access. We focus on two characteristics of the middle schools to which students apply and enroll. The first outcome is the share of a middle school's students that later enroll in the city's elite specialized high schools.[3] This outcome, which we refer to as the *specialized high school share*,

example: `https://tools.nycenet.edu/snapshot/2023/02M033/EMS/`.

[3]NYC has nine specialized high schools that require an exam for admission. The schools enroll ex-

reflects the extent to which a middle school is a high-performing "feeder school." Our second outcome is the middle school's test score *value-added*, which reflects its causal effects on student learning. For the marginal student, we find little impact of G&T on the types of middle schools where students apply or enroll. In the lottery design, we find large positive effects on the specialized high school share and smaller but positive effects on value-added. Enrolling in G&T increases the specialized high school share of a student's first-choice middle school by 7.1 percentage points and that of their enrolled middle school by 9.0 percentage points. The impacts on application behavior are consistent with a boost to the aspirations of G&T students (emphasized in Carlana et al., 2022), though changes in information and other factors might also mediate the effect. We find that effects are larger for students who are FRPL-eligible, White or Asian, and male. Effects also increase in the entrance exam score.

With reduced form estimates in hand, we turn towards our second research question. We ask how an admissions reform that replaces exam-based admissions with admissions based on teacher recommendations and interviews changes the average treatment effect on the treated (ATT) for measures of middle school access. Since the reform was implemented in 2021, we do not yet observe outcomes for the post-reform cohorts. Furthermore, the COVID-19 pandemic may have caused changes in student outcomes independent of the reform. We therefore aim to simulate the impacts of the policy by reweighting conditional average treatment effects from the pre-reform period to match the post-reform demographic distribution. The reweighting exercise, however, is complicated by the fact that entrance exam scores do not exist for the post-reform cohorts, since the exam was eliminated. As noted above, the entrance exam score is a key dimension of effect heterogeneity, which necessitates an alternative strategy for estimating the distribution of (latent) exam scores for the post-reform cohort.

We overcome this challenge via a structural model of G&T demand, which we use to simulate the composition of post-reform G&T students. In the model, families apply to G&T programs by solving an optimal portfolio problem, receive offers from the assignment mechanism, and decide whether to enroll in G&T or an outside option. We estimate and simulate the model on data during the exam admissions period, leveraging a sample for which we observe baseline exam scores. The scores, which are missing post-reform, are a necessary input for predicting the ATT. We reweight subgroup-specific pre-reform treatment effects to the simulated post-reform distribution of students. The exercise reveals that the reform likely reduces the ATT on elite middle school access. In our baseline specification, the ATT on the specialized high school share of the middle

tremely high-achieving students and admissions are intensely competitive.

schools that students rank as their first choice decreases from $+5$ percentage points to -5 percentage points when moving from test- to recommendation-based admissions. While the reform increased the share of G&T FRPL students, who have larger treatment effects, it also substantially increased the share of students who would have had low entrance exam scores prior to the reform. Effects on the latter group are small or negative, which greatly dampens the ATT.

We conclude by tracing a possibilities frontier for what hypothetical policies could achieve. Both exam and recommendation-based policies are suboptimal such that there are alternatives that could simultaneously improve both diversity and treatment effects on middle school access. Nevertheless, G&T enrollment under exam-based admissions is closer to the frontier than enrollment under recommendation-based admissions. Finally, the shape of the frontier suggests that there need not be a tradeoff between program diversity and effectiveness, as moving along the frontier initially improves both dimensions.

Our study relates most directly to prior work on the causal effects of gifted education (Bui et al., 2014; Card and Giuliano, 2014; Card et al., 2024; Cleveland, 2023; Goldstein et al., 2026; Thompson, 2021), as well as to a closely related literature on the effects of tracking (e.g., Abdulkadiroğlu et al., 2014; Card and Giuliano, 2016a; Carlana et al., 2022; Dobbie and Fryer, 2014; Duflo et al., 2011). We contribute by identifying effects on inframarginal students across the ability distribution, studying very early-age G&T, and simulating the impacts of an admissions reform. The studies most closely related to ours are Cleveland (2023) and Strickland et al. (2024), both of which study NYC G&T. We build on their work by using a research design that exploits all of the random variation embedded in the G&T lotteries, estimating effects on a broader population of students and for longer-run outcomes, and by studying a recent admissions reform.

Our paper also relates to the literature on aspirations and preference formation in school settings (Ajayi et al., 2017; Andrabi et al., 2017; Corcoran et al., 2018; Corradini, 2024; Hastings and Weinstein, 2008). Past research focuses primarily on information interventions (Hahm and Park (2024) and Idoux and Corradini (2023) being exceptions). We study the effects of changes in the primary school environment on subsequent schooling decisions. We use rich preference data generated by centralized assignment to do so.

Finally, our study builds on a large literature on segregation and achievement gaps, particularly in elite education (e.g., Reardon, 2018; Hoxby and Avery, 2013). Our exercise simulating the effects of an equity-motivated admissions reform relates to a parallel literature evaluating the consequences of other reforms designed to broaden access to elite education. This body of work includes studies on affirmative action and other diversity-based admissions policies (Ellison and Pathak, 2021; Idoux, 2021), top percent policies

(Bleemer, 2020; Kapor, 2024), and universal screening (Card and Giuliano, 2016b). In contrast to this research, we study whether G&T in early childhood increases access to elite secondary schools, and whether reforms to G&T admissions change access.

2 Setting and Data

2.1 Gifted and Talented Programs in NYC

G&T programs are extensive in NYC. In the 2022-23 school year, approximately three percent of kindergarten students were enrolled across 85 G&T programs. Schools may direct distinct programs, each of which could vary in admissions criteria. Gifted students are sorted into separate classrooms and schools, distinguishing it from G&T in other districts that provide enrichment to gifted students in otherwise integrated classrooms.

Underrepresentation of minority students in G&T classes makes the program controversial, and in recent years, the program has been subject to prominent policy changes. For example, outgoing mayor Bill de Blasio eliminated G&T in 2022, only for successor Eric Adams to reinstate and expand it (Shapiro, 2021). Both reforms were rationalized, in part, on the basis of a perceived pipeline from G&T programs to selective middle and high schools.

The district offers G&T programs for students enrolled in kindergarten through fifth grade. Kindergarten is the most common entry point. G&T programs are intended to "offer accelerated education," but there is no standardized curriculum beyond alignment with the Common Core and implementation is flexible. There are two types of G&T programs: *district* and *citywide*. District programs are offered by schools that also offer non-gifted programs, co-located in the same building. They offer assignment priority to those who reside in the same school district and, in some cases, a neighboring district.[4] Citywide programs are likewise open to all students but do not offer district priority. Five schools host citywide programs and do not offer other district or non-G&T programs. Four of the five citywide schools enroll students from grades K-8. The remaining school enrolls students from K-12. District G&T programs conclude in 5th grade at the latest.

Offers are made to applicants via a centralized assignment mechanism. Prior to fall 2020, an exam was required for eligibility. Beginning in fall 2021, the exam was replaced with an admissions system based on teacher recommendations and interviews. Variation in eligibility and assignment during the exam admissions period forms the basis of our empirical strategies.

[4]There are 32 school districts that comprise the NYC public school system.

2.2 Exam Admissions

During the period with exam-based admissions, families initiate the application process by requesting the G&T entrance exam.[5] Two tests comprise the examination. For most of our sample period, the tests were the Naglieri Nonverbal Ability Test (NNAT) and the Otis-Lennon School Ability Test (OLSAT). Prior to 2013, NYC used the Bracken School Readiness Assessment (BSRA) instead of the NNAT. For each student i, percentile ranks achieved on the two tests were combined to form a composite entrance exam score $R_i \in \{0, 1, ..., 99\}$. Students needed a score of at least 90 to qualify for a district program and at least 97 to qualify for citywide. See Appendix A for more details about the exams and composite score formula.

Families rank up to 12 G&T programs on their application, with assignments determined by the single-offer student-proposing deferred acceptance (DA) algorithm, which we refer to as the "match." Each student has an assignment priority at each program. Let $\varrho_{is} \in \{1, ..., K, \infty\}$ define the priority of student i at school s, where $\varrho_{is} < \varrho_{js}$ indicates that student i has a higher priority than student j at program s. Ineligible applicants have priority $\varrho_{is} = \infty$. Priorities are determined by unique combinations of sibling status, in-district status, and the entrance exam score R_i.[6]

All district programs used the same priority ordering. The highest priorities were awarded to siblings living in-district in descending order of the entrance exam score R_i. Next were siblings living out-of-district sorted by R_i, then in-district non-siblings sorted by R_i, and finally out-of-district non-siblings again sorted by R_i. Put differently, priority orderings are lexicographic: within each cell defined by sibling-district status, applicants with higher exam scores have a higher priority. The priority ordering at citywide programs was identical but without the district component. Beginning in 2016, some programs offered additional priority to students from underrepresented groups (e.g., low-income and English language learners) as part of the city's Diversity in Admissions (DIA) initiative. Under the DIA scheme, underrepresented students are prioritized for a fraction of a program's capacity. If the reserved seats are unfilled, non-underrepresented students could be assigned to the reserved seats. Appendix A provides an example of how priorities are constructed.

[5] In the fall 2020 admissions cycle, November 2019 was the deadline to apply for the G&T entrance exam. The exam was administered in January 2020. Scores and G&T applications were received by May 2020, prior to the application deadline. Assignments were determined in June 2020. Prior to 2008, G&T assignment was decentralized.

[6] Some programs offered priority to applicants in neighboring districts when there were no G&T programs offered in that district. Appendix Figure A1 depicts the geographic distribution of kindergarten programs across districts.

At oversubscribed programs, ties among equally qualified applicants in the same priority group were broken via a lottery number. Formally, each applicant was assigned a tie-breaker $t_i \overset{iid}{\sim} U[0,1]$, fixed for each applicant across all programs. Define *applicant rank* as $\chi_{is} = \varrho_{is} + t_i$. Under DA, students are assigned to programs in order of χ_{is}. Define a program's *marginal priority* as the priority of the last student assigned to s: $\varrho_s = \max_i\{\varrho_{is} \mid i \text{ assigned to } s\}$. Random assignment arises because assignments for applicants in the marginal priority group, those for which $\varrho_{is} = \varrho_s$, are determined by a random tie-breaker t_i. We detail a strategy to exploit this randomness below.

During the study period, the G&T match was completely independent from the non-G&T match (which also uses the DA algorithm). Due to high demand, families were advised that G&T eligibility did not guarantee them a seat.

2.3 Admissions Reform

For fall 2021 admissions, the G&T entrance exams were replaced with recommendations from pre-K instructors and staff interviews. Interviews are reserved for kindergarten applicants who are not enrolled in pre-K and do not have teachers available to make recommendations. Under the new system, when families submit an application for G&T, pre-K teachers evaluate the student against a list of criteria associated with "giftedness." Appendix A provides details on the criteria.[7] For those who requested interviews, the interviews were conducted by staff at the New York City Public Schools' Division of Early Childhood Education. For fall 2022 admissions, 6% of kindergarten G&T applicants applied through an interview and 94% applied through teacher recommendations, with similar acceptance rates under each.

Upon receiving a recommendation, students are eligible for the match, which is run in a similar fashion as in the exam period. Families again rank up to 12 programs, but priorities no longer depend on an exam score. As a consequence, the only distinction between district and citywide admissions criteria is the absence of district priority for the latter. Citywide programs no longer have a higher standard for admission and no distinction is made between the programs when teachers recommend students. Random lottery numbers are once again used to break ties between applicants within the marginal priority group.

The admissions reform did not notably change the number of G&T applicants, but it substantially increased the number of applicants who qualified for G&T and therefore

[7]Beginning in the following year, for fall 2022 admissions, preschool teachers were instructed to universally screen all of their students using the same criteria. This change also coincided with an expansion of the number of G&T programs.

participated in the match. Appendix Figure A2 shows that, in fact, a smaller share of NYC kindergarten students applied to G&T programs in 2021 than in 2020 or 2019, continuing the downward trend in application rates since 2017. However, 27% of applicants were deemed eligible in 2020 under the exam system while 91% were in 2021 during the first recommendation year. As a result, the number of students who participated in the G&T match increased substantially from about 3,000 in 2020 to almost 10,000 in 2021.

The reform also generated differences in the composition of G&T students. The share of FRPL G&T students increased substantially, as depicted in Figure 1, despite few changes in the FRPL share of G&T applicants (see Appendix Figure A3). The increase in the FRPL share was much larger than changes for groups defined by race and ethnicity in percentage point terms. However, increases in Black and Hispanic representation were larger as a percentage of pre-reform levels.

2.4 Data and Outcomes

Our analysis uses data covering all kindergarten G&T applicants from school years 2011-12 through 2020-21, obtained from New York City Public Schools. We use information generated by the G&T match, which includes student rank-ordered preferences over programs, priorities at those programs, and the program to which they were assigned.[8] We merge the application data to files containing student demographics, including gender, race, FRPL, disability, English language learner status, residential location, as well as all public school enrollment by grade. Finally, we link these data to New York State standardized exams and middle and high school applications. G&T enrollment is defined as enrolling in a class where at least 90% of the students have been identified as gifted.[9] See Appendix B for more details on data sources.

Our analysis begins by estimating G&T effects on the elementary school classroom environment and achievement. Outcomes for the former are class size and the share of classmates who are FRPL-eligible or White or Asian.[10] Our achievement outcomes are New York State standardized assessments in math and ELA, administered in grades 3-8, and Algebra I Regents scores, typically administered in grades 8-9. All scores are standardized to be mean zero and unit variance by test-grade-year. Tests are available

[8]Our files omit the lottery tie-breakers used for student assignment. We describe in Section 3.2 how assignment risk can be computed with simulated lottery numbers instead. The application files are also missing sibling priority information for a subset of years. We discuss further below.

[9]For a subset of years, we have data on the official class codes that the district uses to classify G&T programs. Our measure lines up very closely with theirs.

[10]We winsorize measured class size at the 1st and 99th percentiles to reduce the influence of a few outliers.

in all years except 2020 and 2021, when exams were suspended during the COVID-19 pandemic.

Our school access outcomes are the characteristics of middle schools that students rank as their first choice and enroll.[11] We estimate impacts on two characteristics. The first is a middle school's *specialized high school share*, which is the share of sixth grade students from that school who later enrolled in an elite specialized high school. This outcome measures prestige. Much attention is paid to such "feeder" middle schools, and the specialized high school share is a publicized metric that conveys elite school status (Lee, 2018).[12]

The second school access outcome is *value-added* on sixth grade achievement. This outcome reflects a school's causal effectiveness. We estimate value-added from student-level regressions of sixth grade achievement on school enrollment dummies and controls for demographics and functions of lagged test scores, following Angrist et al. (2017, 2024a). Angrist et al. (2024a) show that value-added estimates for math obtained from models of this form are nearly unbiased for sixth grade math in NYC middle schools. Accordingly, we use math value-added to test whether G&T enrollment has meaningful impacts on the causal effectiveness of the middle schools where students apply and enroll.

Finally, we construct outcomes for access to elite high schools. The first outcome is an indicator for whether a student took the Specialized High School Admissions Test (SHSAT), which is required for admission to specialized high schools. The second outcome is whether the student received an offer to any specialized high school.[13]

2.5 Summary Statistics

Table 1 reports summary statistics for all kindergarten students, G&T applicants, and those in our analysis samples. Applicants for kindergarten G&T are more likely to be White or Asian and less likely to be Black or Hispanic than non-applicants. They are also much less likely to be FRPL-eligible, have a disability, or be an English language learner. The estimation samples are more selected than the sample of all applicants. Among the estimation samples, the lottery sample with higher scoring applicants is even more se-

[11] A small number of students enroll in the same school for fifth and sixth grade, and do not appear in the assignment file if they opt to remain in the same school instead of participating in the centralized match. We treat their fifth and sixth grade school as their first choice.

[12] In principle, the deferred acceptance is strategy-proof. However, even under strategic reporting, our outcome captures realized application behavior, which is what matters for the pipeline regardless of underlying preferences.

[13] In supplementary analyses, we look directly at impacts on high school enrollment, though results are less precise as we have less data for this outcome.

lected than the RD sample. Roughly 22% of all kindergartners applied for G&T. Among applicants, 26% qualified for district programs and 11% qualified for citywide programs.

3 Empirical Strategies

3.1 RD at the Qualifying Cutoff

The cutoff from the G&T entrance exam first lends itself to an RD design where we estimate effects for students on the margin of gifted qualification. Qualification does not imply enrollment, so we instrument for enrollment in a fuzzy RD setup. We estimate the following via two-stage least squares (2SLS):

$$\begin{aligned} D_i &= \pi 1(R_i \geq 90) + \lambda_1 R_i + \gamma_1 R_i 1(R_i \geq 90) + X_i' \Psi_1 + \nu_i \\ Y_i &= \beta D_i + \lambda_2 R_i + \gamma_2 R_i 1(R_i \geq 90) + X_i' \Psi_2 + \varepsilon_i, \end{aligned} \tag{1}$$

where D_i indicates if student i enrolled in kindergarten G&T, R_i is the entrance exam score, and X_i is a vector of baseline controls included to increase precision. The control vector includes indicators for demographic variables, year dummies, and interactions between them. The parameter of interest is β, which is the local average treatment effect (LATE) of kindergarten G&T enrollment at the qualifying cutoff. Applicants may re-take the entrance exam in later years if they fail to qualify. There remains a substantial first stage on ever enrolling in G&T, as depicted in Appendix Figure C1.

Our baseline RD estimates impose linear fits and a bandwidth that includes five points on either side of the qualifying cutoff for both the first stage and reduced form. We forgo data-driven bandwidth selection (Calonico et al., 2014; Imbens and Kalyanaraman, 2012) as there are few reasonable choices for a symmetric bandwidth given the integer-valued running variable and truncation induced by the citywide cutoff at 97. Appendix Table E2 shows that our findings are robust to alternative specifications.

The validity of our RD design relies on standard assumptions on instrument relevance and continuity of potential outcomes and treatments (Hahn et al., 2001; Dong, 2018). Conditional expectations of the pre-treatment covariates are also assumed to be smooth at the cutoff (Calonico et al., 2019). The instrument has a strong first stage, as we discuss in Section 3.4. We provide support for the smoothness assumptions in two ways. First, Appendix Figure C2 shows there is no evidence of bunching around the district cutoff, suggesting that exam scores are not substantially manipulated. Second, Appendix Table D1 shows that G&T qualification largely fails to predict baseline covariates. The corresponding RD plots are depicted in Appendix Figure D1. Differential attrition could also generate selection bias even when the RD assumptions are satisfied. We find little evidence for attrition in the RD sample, a finding we discuss further in Section 3.3.

3.2 Lottery IV

Our second strategy relies on the lottery-based rationing of G&T seats at oversubscribed programs. As described above, G&T assignments are determined by the DA algorithm. Formally, families submit a rank-ordered list (ROL) of their preferences over G&T programs, $\succ_i$, and have priorities at programs, $\varrho_i = (\varrho_{i1}, \ldots, \varrho_{iS})$. Define student type as a student's combination of preference rankings and priorities: $\theta_i \equiv (\succ_i, \varrho_i)$. Under DA, same-type students share the same probabilities of assignment. That is, conditional on type, variation in assignments is solely a function of random lottery numbers. This feature of DA substantiates the key conditional independence assumption underlying our IV strategy: gifted assignment, denoted Z_i, is independent of potential outcomes conditional on student type.[14]

In practice, there are relatively few students who share the same type, which makes θ_i a high-dimensional object. This makes full-type conditioning (controlling for indicators for each value of θ_i) impractical. Therefore, following Abdulkadiroğlu et al. (2017), our strategy replaces θ_i with the assignment propensity score, $p_i \equiv \Pr(Z_i = 1|\theta_i)$. Conditioning on the propensity score provides dimension reduction by pooling different-type students who share the same propensity score, while still eliminating selection bias as if we had controlled for student type directly (Rosenbaum and Rubin, 1983). This strategy extracts all of the random variation embedded in the assignment algorithm, which improves precision.

We compute the propensity score via simulation. The procedure is as follows. Hold preferences and priorities fixed. Then, for each student, draw a lottery number from a uniform distribution between 0 and 1. Next, run the assignment algorithm using the lottery numbers and record the assignments. The simulated propensity score is the probability that a student is assigned to a gifted program across many simulations of the assignment algorithm. This estimate converges to the true finite-market propensity score as the number of simulations increases (Abdulkadiroğlu et al., 2017). In our application, we simulate the propensity score by running the match 100,000 times.[15]

[14] The G&T assignment instrument is the sum over all program-specific assignments such that $Z_i = \sum_s Z_{is}$, where Z_{is} denotes assignment of i to program s. DA is a single offer mechanism, so $Z_{is} = 1$ for at most one program.

[15] Prior applications of this identification strategy use analytic formulas to compute a large-market ap-

With propensity scores in hand, we estimate the following system by 2SLS:

$$\begin{aligned} D_i &= \varphi Z_i + \omega_1 p_i + X_i' \kappa_1 + \zeta_i \\ Y_i &= \tau D_i + \omega_2 p_i + X_i' \kappa_2 + \eta_i. \end{aligned} \tag{2}$$

The parameter of interest is τ, which is a convex-weighted average of LATEs for strata defined by values of the propensity score. Baseline covariates X_i are the same as in our RD approach and serve to increase precision. In contrast to the RD strategy, this procedure estimates an average effect for students across the full support of the (qualifying) exam score distribution rather than just at cutoff. Appendix Figure C2 confirms that students with non-degenerate assignment risk in the lottery sample span the support of the eligible score distribution.

Our strategy relies on accurate computation of the propensity score p_i. We note that prior to 2015, data on sibling priority is missing, preventing us from perfectly replicating G&T assignments. For these years, we assume that applicants do not have sibling priority when simulating the match. In this case, the identification strategy is predicated on the ignorability of the instrument given preferences and non-sibling priorities. Without sibling priorities, offers are not literally random but may nevertheless be independent of potential outcomes. Since offers are a function of student type and lottery numbers, independence is preserved when there is negligible confounding of the instrument mediated by sibling status. Tests for covariate balance confirm that our propensity score control strategy largely eliminates selection bias. Appendix Table D1 shows that assignment strongly predicts pre-determined student characteristics, but there remains little to no correlation after conditioning on assignment risk.

3.3 Attrition and Retention in NYC Schools

Differential attrition threatens the validity of either research design if the attrition is non-random and correlated with potential outcomes. For example, if high-achieving students are more likely to enroll in private schools when losing out on a G&T offer, then our causal effect estimates may be negatively biased. We note that attrition is also an interesting outcome in its own right. A major rationale for G&T is its potential to attract and retain families in the public school system. Prior work suggests that there are G&T retention effects in NYC and elsewhere (Cleveland, 2023; Bui et al., 2014). At the same time, other

proximation to the propensity score (e.g., Abdulkadiroğlu et al., 2017; Angrist et al., 2024a). We opt for the simulated finite-market score because our data omit lottery numbers, prohibiting us from using the analytic formulas. Abdulkadiroğlu et al. (2017) find that scores computed by simulation and formula align closely.

work on tracking finds little evidence for attrition (Card and Giuliano, 2016a).

We test for attrition by estimating causal effects on enrolling in an NYC public school. In the RD sample, we apply 2SLS to estimate the effect of receiving a G&T offer, separately by grade. In the lottery sample, we estimate the retention effect using the reduced form of the same outcome on G&T assignment, again estimated separately by grade. See Appendix C.1 for details.

Figure 2 depicts the retention estimates. Panel A shows that in the RD sample, G&T assignment has little impact on enrolling in the district from kindergarten through ninth grade. Effects in kindergarten and 1st grade are small (approximately 2 percentage points) and marginally significant, then decrease and become insignificant by 2nd grade. On the margin, G&T does not appear to retain students in public schools, and we conclude that differential attrition is not a concern for the validity of our RD design. However, we find larger retention effects in the lottery sample. Panel B of Figure 2 shows that differential attrition is significant and grows from kindergarten up until fifth grade to a peak of around 7 percentage points. It declines thereafter and reaches a small and insignificant effect by eighth grade.

We address attrition in several ways. First, we report tests for balance in the selected sample of students who later enroll in 6th grade. Appendix Table D1 shows that covariates remain balanced in these samples, which suggests that attrition may not be generating selection bias. Second, we find that estimates obtained when measuring our middle school outcomes in 8th rather than 6th grade are similar. The correspondence is reassuring as levels of attrition are insignificant by 8th grade. Finally, Appendix Table E1 reports similar results obtained from weighted 2SLS with weights equal to the inverse probability of having a non-missing outcome. Under an ignorability assumption on missingness, the weighted 2SLS estimand is the same as that of conventional 2SLS in the absence of attrition. Appendix C.6 provides details on the procedure.

3.4 First Stage

Table 2 reports various first stage effects of the RD eligibility and lottery offer instruments on G&T enrollment. At the cutoff, the probability of enrolling in kindergarten G&T the year after applying increases by 24 percentage points. The probability of ever enrolling in G&T increases by 16 percentage points. The difference between kindergarten enrollment and any K-5 enrollment reflects re-taking behavior in later grades. Eligible students at the cutoff enroll in G&T for 0.7 more years than those just missing the cutoff.

We depict eligibility effects in Figure 3. Ineligible applicants are neither assigned nor enroll in G&T. The probability of assignment increases by roughly 40 percentage points at the cutoff. The probability of enrollment increases by a lower 25 percentage points or so. The wedge between offer and enrollment effects is a manifestation of non-compliance among eligible applicants. Figure 3 also illustrates why we focus on the district G&T cutoff, as there is no first stage at the citywide qualifying cutoff.

Column 2 of Table 2 reports the effects of G&T offers in the lottery sample. The instrument increases the probability of kindergarten G&T enrollment by 41.2 percentage points, corresponding to a 41.4 percentage point increase in the probability of ever being enrolled in G&T. An offer increases time ever enrolled in G&T by 2.3 years.

3.5 Comparing RD and Lottery Designs

Throughout, we contrast estimates of G&T effects obtained from RD and lottery designs. Most of the prior work on G&T and tracking employs RD designs to estimate effects at a qualifying cutoff. A lottery design yields estimates pertaining to students both near and away from the cutoff, which could differ from RD estimates for several reasons.

First, the lottery design generates estimates that reflect a population with more variation in baseline ability than that when using RD. While the RD design yields a local effect for marginal students at the cutoff of 90, the lottery sample reflects a weighted average effect using students across the score distribution from 90 to 99. Differences across estimates might reflect heterogeneous effects for students who vary in baseline ability.

Second, the RD eligibility instrument only induces enrollment into programs that admit students who are marginally eligible for district programs. Since priorities are increasing in the entrance exam score, the programs that the marginal students end up in must not be the most popular ones (if they were they would have had a higher effective cutoff). By contrast, the offer instrument shifts students across programs with varying levels of demand.[16] Therefore, lottery effect estimates might also reflect heterogeneous effectiveness across programs.

Appendix Table D2 shows that lottery compliers are more likely to be White and less likely to be Black, Hispanic, Asian, or FRPL than RD compliers. Racial differences persist even among lower-scoring lottery compliers within the RD bandwidth, though FRPL shares are similar.

[16]Note that while the offer instrument is driven by oversubscription, rationing could also induce randomization at undersubscribed programs, permitting students to have non-degenerate assignment probability at programs with excess supply. To see this, suppose a student has $p_{is} = 0.5$ for a program s that they have ranked as their first choice. Their second choice s' is undersubscribed so the student is guaranteed a seat there if denied at s. Therefore, $p_{is'} = 0.5$.

4 Causal Effects of G&T

4.1 Kindergarten Classroom Environment

Enrolling in G&T significantly changes the classroom environment in ways that could plausibly generate long-run effects. Table 3 reports 2SLS estimates from the fuzzy RD design with the corresponding reduced forms depicted in Figure 4. For marginal students, G&T enrollment causes a sharp reduction in class size of about 4.1 students, relative to a mean of 22 students per class in non-G&T classrooms. The magnitude is somewhat smaller than the -6 to -8 student decrease generated by the Tennessee STAR experiment (Krueger, 1999). Class size reductions could boost achievement, though evidence on the matter is mixed (Angrist and Lavy, 1999; Angrist et al., 2019; Fredriksson et al., 2013; Hoxby, 2000).

For marginal students, G&T enrollment also generates large changes in peer exposure. Enrolling in G&T causes a 26 percentage point decline in the share of peers eligible for subsidized lunch (relative to a non-G&T mean of 45 percent) and a 14 percentage point increase in the share of peers who are White or Asian. Large changes in peers are consistent with prior work on tracking and G&T (e.g., Bui et al., 2014; Card and Giuliano, 2016a) and could impact secondary school choice by changing preferences for school demographics and the information families have about schools (Idoux and Corradini, 2023). In principle, peers could also impact achievement, though prior work emphasizes that variation in achievement levels across schools with different racial compositions is largely explained by selection bias (e.g., Angrist et al., 2024b).

Across all outcomes, G&T leads to smaller changes in the classroom environment for students in the lottery design. There is a small and statistically insignificant decline in class size of 0.1 students. The FRPL peer share decreases by 13 percentage points, and the White and Asian peer share increases by 9 percentage points. Evidently, untreated lottery compliers enroll in classes that are more similar to G&T classrooms than RD compliers.

4.2 Achievement

Changes in classroom characteristics notwithstanding, G&T does not appear to have positive impacts on achievement outcomes. Table 4 reports impacts on New York State elementary and middle school math and ELA assessments, as well as the Algebra I Regents exam (taken in either 8th or 9th grade). We stack the samples across grades for the math and ELA outcomes to increase precision.[17] Panels D through F of Figure 4 depict the RD reduced form.

The effect estimates are small and insignificant on all of the test score outcomes for both the RD and lottery design. The ELA estimate is 0.04σ in the RD design and -0.02σ in the lottery design, while the corresponding math estimates are -0.07σ and -0.04σ. The results are precisely estimated, with standard errors around 0.05σ. The stacked setup is over-identified, as there are grade-specific instruments for a single binary treatment. We report p-values obtained from testing the model's over-identifying restrictions and uniformly fail to reject the null, suggesting that there is little effect heterogeneity across grades. Appendix Table F2 also reports grade-specific estimates, where we fail to find significantly positive effects for any grade or outcome. We similarly find little impact on Algebra I performance, though the estimates are less precise.[18]

We generate even more precise estimates by stacking the RD and lottery samples. Column 3 of Table 4 reports the estimates from this setup, where we continue to stack across grades for math and ELA. We obtain point estimates of -0.05σ for math and virtually 0 for ELA. The magnitude of the standard errors is reduced by roughly 15% relative to the lottery estimates and around 40% relative to the RD estimates. Confidence intervals exclude the impacts of G&T on non-gifted students found in Card and Giuliano (2016a). The over-identification test in this setup fails to reject when we use grade- and research design-specific instruments, again suggesting limited heterogeneity. We conclude that G&T has little impact on test score outcomes, on average, for students across grades and research designs. Furthermore, our results suggest that large OLS test score effects, as reported in Appendix Table F1, are explained by selection bias.

Our findings are consistent with prior work. Cleveland (2023) estimates small short-run achievement effects on 3rd grade outcomes using a different research design and only for FRPL students. Bui et al. (2014) also find little impact of G&T enrollment on test score outcomes using both a RD and lottery design, though estimates for the latter are imprecise and complicated by large levels of differential attrition. While Card and Giuliano (2016a) find positive impacts on high-achieving non-gifted students who enroll in a G&T class, there were no achievement benefits for gifted students on the margin of qualification in their setting (Card and Giuliano, 2014). However, the program did have large impacts

[17] We stack across grades 3-8 for ELA and 3-6 for math. We omit grades 7-8 from the latter as many students are exempt if they instead take the Algebra I or Geometry Regents exams. See Appendix C.3 for details on the estimation procedure.

[18] The non-G&T mean is around 1σ above the mean, as potential G&T students are more advantaged than the NYC population as a whole. In estimated complier outcome distributions (unreported), we find that mass for untreated compliers is concentrated around 0.5-2σ. As the highest-achieving students score around 2.5σ above the mean, we conclude that "ceiling effects" do not explain the null G&T effects.

on college enrollment, course-taking, and grades for gifted boys at the cutoff (Card et al., 2024). It is worth emphasizing that our lottery estimates corroborate the RD-based null effects of prior work. The results weigh against rank effects (Urquiola et al., 2024; Barrow et al., 2020) as an important mechanism: such effects are less likely for inframarginal students who score above the cutoff.

On balance, G&T enrollment does not appear to positively impact elementary and middle school academic outcomes. This suggests that achievement is not a mechanism by which G&T impacts secondary school access. Nevertheless, G&T could generate pipeline effects by changing student aspirations, preferences, or information about schools. We turn next towards estimating effects on middle school applications and enrollment.

4.3 Access to Secondary Schools

As described above, we estimate impacts on the characteristics of middle school that students rank as their first choice and ultimately enroll. The first characteristic we focus on is the share of students at the middle school who later enroll in a specialized high school. We interpret this as a measure of school prestige, as it reflects the extent to which the school is a desirable feeder middle school. The second characteristic is the math value-added of the middle school. Unlike the specialized high school share, which may be compromised by selection bias, value-added is a measure of causal school effectiveness.

Column 1 of Table 5 reports RD estimates of G&T on the middle schools to which students apply and enroll. Figure 5 depicts the corresponding plots. We uniformly fail to reject the null across outcomes. For applicants at the margin, G&T does not cause them to change their most preferred school or where they enroll.

By contrast, students in the lottery design do change their behavior after enrolling in G&T. Column 2 reports that G&T increases the specialized high school share of the middle school they rank as their first choice by 7.1 percentage points. The increase on the enrollment margin is even larger (9.0 percentage points). The impacts are smaller for value-added: the first choice school of treated students has an estimated value-added that is only 0.03σ larger than those for the controls. There is a marginally significant 0.02σ increase in value-added on the enrollment margin.

We conclude that G&T does in fact facilitate access to middle schools that parents likely find desirable, though these schools improve student learning little, as measured by test score value-added.[19] Our results suggest that G&T admissions reforms have the po-

[19]In Appendix Table E3, we report qualitatively similar results when using mean test score levels as a proxy for middle school prestige instead of the specialized high school share. We also obtain similar results when measuring the outcome in 8th grade, where there are low and insignificant levels of differential

tential to change the composition of schools in which students enroll at later ages. We also show that G&T changes both application and enrollment behavior. The former suggests that G&T may be changing student aspirations or information about higher-performing schools, rather than simply increasing the probability of enrollment conditional on applications and offers. Changes in aspirations or information do not appear to be mediated by changes in test score performance, consistent with the findings of Carlana et al. (2022). Null impacts on marginal students suggest that aspirations may be a more important mechanism than information.

Finally, we ask whether the middle school effects persist into high school. We estimate the impact on taking the SHSAT, a prerequisite for applying to one of the specialized high schools, as well as whether students receive a specialized high school offer. The estimated impacts, reported in Appendix Table F3, are uniformly small and insignificant.

5 Heterogeneity

5.1 Baseline Ability

We first estimate how effects of kindergarten G&T vary across the distribution of entrance exam scores. Figure 6 plots the treatment effect estimates from the lottery design, estimated separately by four score bins in the lottery sample; for comparison, red markers show the RD estimates. We focus on elementary and middle school outcomes, as the high school estimates are less precise. Panels A and B depict treatment effects on math and ELA, where we continue to find effects close to zero or negative across the entrance exam score distribution.

The remaining panels show the gradient of treatment effects for our middle school access outcomes. We estimate positive effects that are largest among applicants with the highest entrance exam scores. Moreover, we find that most of the estimates increase in the exam score. For the specialized high school share outcomes, estimated effects range from a negative impact in the 90-92 score bin to large positive effects in the highest score bins. The pattern for the value-added outcomes is similar. In Appendix Figure F1, we plot effects by each unique value of the baseline score, as well as a linear fit. The score-specific estimates are noisier than the binned estimates, but the slopes remain positive. We conclude that G&T boosts access to desirable middle schools, but mostly for students who scored higher on the entrance exam at baseline.

attrition, instead of in 6th grade. Finally, we find similar results, reported in Appendix Table E1, when applying inverse probability weighted 2SLS to account for attrition. Appendix C.6 provides more details.

5.2 Family Income

Next, we examine heterogeneity in G&T impacts by family income, proxied by FRPL status. We focus on income heterogeneity for two reasons. First, as described in Section 2.3, the introduction of a recommendation-based admissions system markedly increased the share of FRPL students in G&T kindergarten programs. Second, prior work suggests that low-income students, especially boys, might especially benefit from tracking treatments (Card and Giuliano, 2016a; Cohodes, 2020; Card et al., 2024).

We find stronger patterns of heterogeneity when estimating impacts on school access.[20] The estimates in Table 6 show that in the lottery sample, G&T increases the specialized high school share of the preferred school by 11.9 percentage points for FRPL students and by 6.4 percentage points for non-FRPL students. The corresponding estimates for the specialized high school share of the enrolled school are 12 percentage points for FRPL and 8.2 percentage points for non-FRPL (though we fail to reject that the effects are equal due to smaller samples).[21] Results for value-added are similar, with significant differences between effects on the value-added of the enrolled middle school only for FRPL students. In the RD sample, enrollment effects generally go in the opposite direction. FRPL students experience a significant decline in the specialized high school share that is statistically different from the small effect for non-FRPL students. As reported in Appendix Table E3, heterogeneity by income follows a similar pattern when using alternative outcomes, such as 6th grade test score levels or the specialized high school share measured in 8th grade when the level of differential attrition is insignificant.

The patterns in Table 6 motivate an investigation of heterogeneity by FRPL and entrance exam score combined. Figure 7 reproduces Figure 6 but separately by FRPL status. We continue to find little to no effect on achievement across the score distribution for both groups. We also continue to find that treatment effects on school access generally increase in the entrance exam score, though the FRPL estimates are less precise.[22] Effects are also somewhat larger for FRPL students.

Finally, we report estimates for high school outcomes in Appendix Table F6. We find that G&T has little impact for either FRPL or non-FRPL students in the RD sample. For

[20] G&T generates larger changes in class size and peer exposure for FRPL students, as reported in Appendix Table F4. Appendix Table F5 reports small test score impacts for both groups.

[21] Although FRPL-specific estimates by gender are noisy, we find that effects are especially concentrated among low-income boys, consistent with prior work on G&T and tracking (Carlana et al., 2022; Card et al., 2024). We estimate that G&T increases the specialized high school share of the enrolled 6th grade school by $0.169 (SE = 0.045)$, the value-added of the enrolled 6th grade school by $0.124 (SE = 0.038)$, and the probability of receiving an SHSAT offer by $0.394 (SE = 0.164)$ for FRPL boys.

[22] See Appendix Figure F2 for a version that plots the effect estimates separately by each possible value of the entrance exam score.

FRPL students in the lottery sample, G&T increases the probability of taking the SHSAT by 18 percentage points and increases the offer probability by 12 percentage points. Therefore, while we find little impact on high school outcomes on average, we find that effects for high-achieving low-income students may be large and significant. However, the estimates are imprecise, so we interpret the point estimates as tentative.

5.3 Programs

Are G&T effects heterogeneous by program? A natural starting point compares district and citywide G&T. The contrast is salient for parents, and the more selective citywide programs appear to be especially desirable. Citywide programs exclusively enroll gifted students, which suggests that the citywide treatment could be more intensive than the district programs. Importantly, all five citywide programs continue through at least 8th grade. We also aim to test whether the effects on middle school access mechanically arise from students continuing at their citywide school.

We disentangle the effects by extending our lottery design to a multisector model with two treatments and two instruments. Specifically, we use 2SLS to instrument district and citywide enrollment with district and citywide offers. By the logic described in Section 3, district and citywide offers are conditionally ignorable given propensity scores for each (restricting the sample to applicants with non-degenerate probability of assignment to at least one of the sectors). In what follows, we also restrict the sample to students who are citywide eligible (entrance exam score of 97 or above) to facilitate comparability between district and citywide compliers.

The multisector strategy relies on stronger assumptions than conventional IV. In general, 2SLS estimation of models with multiple treatments and instruments yields weighted averages of treatment effects across treatments and complier types, not necessarily a LATE (e.g., Behaghel et al., 2013; Kirkeboen et al., 2016). Restrictions on instrument response types yield interpretable effects under arbitrary forms of treatment effect heterogeneity. However, modest violations of the behavioral restrictions likely matter little for bias unless effect heterogeneity is very large, as noted by Bhuller and Sigstad (2024) and Heinesen et al. (2022). In Appendix C.4, we support the validity of the multisector strategy by investigating the testable implications of the behavioral restrictions. We apply a bounding argument from Heinesen et al. (2022) to show that the share of applicants who violate the necessary conditions is small, suggesting little scope for bias.

We report multisector 2SLS results in Table 7. Neither district nor citywide appears to have significant impacts on test scores (except for a statistically significant citywide effect on math). There is more heterogeneity, however, in impacts on the specialized high school share of the enrolled middle school. Citywide and district generate gains of 16 and 4 percentage points. For FRPL students, the impacts are larger, with a 27 percentage point increase for citywide and 14 percentage point increase for district. As noted above, citywide programs continue through at least 8th grade, so it may not be surprising that citywide impacts are large. However, positive effects of district G&T suggest that the citywide continuation effect is not driving our overall results. That is, district programs increase the specialized high school share despite ending in 5th grade. Interestingly, the impact of district G&T is concentrated among FRPL students, perhaps because non-FRPL students would have aspirations or information to apply to high-performing middle schools even without G&T.

Impacts on value-added and high school outcomes are more modest but also reveal interesting patterns of heterogeneity. While there is little impact on value-added of the enrolled middle school in the full sample, both district and citywide programs increase this value-added by roughly 0.1σ within the FRPL sample. Citywide has larger impacts on SHSAT-taking and the probability of receiving a specialized high school offer, where again effects appear concentrated among FRPL students. Non-FRPL students appear to benefit little from any type of G&T program for these outcomes. Taken together, these results suggest that there is meaningful heterogeneity across district and citywide programs, but G&T effects on school access are also not an artifact of students remaining at their citywide school. District G&T has positive impacts on middle school access, and especially so for FRPL students.

We might also expect heterogeneity across the roughly 80 G&T programs as schools are granted large levels of autonomy in how they implement their programs. Credible value-added estimation is complicated by the lack of baseline test scores for elementary school.[23] We therefore investigate program-level heterogeneity via the instrumental variables value-added model (IV VAM) framework introduced by Angrist et al. (2024a). IV VAM models program-level causal effects as a function of a low-dimensional set of mediators, instrumenting them with program-specific offers. In our implementation, we use conventional VAM estimates as a scalar mediator. Following Kolesár et al. (2015), the identifying assumption is that exclusion violations where offers impact variation in program quality not captured by the mediator averages out across instruments. By Lemma 3 of Angrist et al. (2024a), this holds if residual variation in true value-added given con-

[23]This also hinders using value-added models to gauge fallback quality. Alternative strategies of instrumenting each program enrollment dummy with its offer or running 2SLS program-by-program are infeasible. The former has more treatments than instruments, and the latter yields unreliable estimates given the small number of students randomized at each program (see Walters, 2015).

ventional VAM estimates is unrelated to offer compliance rates. Appendix C.5 provides details.

This procedure yields estimates of the standard deviation of program-level causal effects, as reported in Appendix Table F8. We obtain estimates of 0.13 and 0.16 for 3rd grade math and ELA scores, suggesting there is meaningful heterogeneity in test score effects across programs. The corresponding estimates are 0.11 and 0.08 for the specialized high school share and value-added of the enrolled 6th grade school, again suggesting non-trivial variation (p-values are all below 0.01). Appendix Figure F3 depicts regression coefficients relating G&T value-added estimates to program-level characteristics. Programs that are citywide, in Brooklyn and Manhattan, and enroll fewer minority students predict larger causal effects on middle school access outcomes.

5.4 Race and Gender

We report impacts on achievement by race and gender in Appendix Table F9 and Appendix Table F10. There is little heterogeneity in RD estimates by race. Lottery estimates are close to zero for White and Asian students but large and negative for Black and Hispanic students. The estimates are imprecise but statistically significant and we reject homogeneous causal effects for each group. There is little heterogeneity in test score impacts by gender in either design. While our results appear to differ from those in Card and Giuliano (2016a), who find positive impacts concentrated among Black and Hispanic students, we note that their effects pertain to non-gifted students who enroll in a G&T class. By contrast, our effects are relevant for gifted students alone.

Likewise, we find little heterogeneity using the RD design for middle school access outcomes. Appendix Table F11 and Appendix Table F12 report estimates close to zero by both race and gender for marginal students. Effects on the specialized high school share of the enrolled middle school are larger for White, Asian, and male students in the lottery sample. Effects on enrolled value-added are concentrated among male students.

6 Simulated Impacts of Eliminating Exam Admissions

We apply our estimates of heterogeneous treatment effects to examine the potential impact of the 2021 NYC G&T admissions reform on middle school access. Our exercise asks how the policy impacted equity and efficiency. Our notion of equity is defined as how close the FRPL share in G&T matches that of NYC public school students overall. We define efficiency as the average treatment effect on the treated population (ATT) for our middle school application and enrollment outcomes, which reflects the extent to which G&T programs increase access to desirable middle schools.

6.1 Structural Model of G&T Demand

6.1.1 Model Setup

Let $s \in \{0, 1, ..., S\}$ index schooling alternatives with S G&T programs. The outside option, denoted $s = 0$, includes enrollment in non-G&T programs, private schools, and schools outside NYC. The model proceeds in three stages. First, applicants submit a ROL $A_i \in \mathcal{A}_i$ by solving an optimal portfolio problem to maximize the expected utility of either enrolling in G&T or an outside option. The set $\mathcal{A}_i$ contains all possible ROLs over G&T programs for which a student is eligible. Second, offers are generated by the DA mechanism and applicants either receive an offer to at most one G&T program or do not receive any offer. Third, applicants receive a preference shock then decide where to enroll by accepting or rejecting their G&T offer.[24] The additional shock rationalizes the presence of never-takers. We now describe each stage in more detail.

6.1.2 Applications, Offers, and Enrollment

Families first choose an application portfolio to maximize expected utility. Let u_{is} denote the initial utility that applicant i has over program s. Applicants anticipate a preference shock after offers are realized but before the enrollment decision, such that they will have a new utility U_{is} prior to enrollment. The distribution of the enrollment stage preference shock is known. Therefore, expected utility at the application stage of an offer at program s relative to having only the outside option is given by $w_{is} = E[\max\{0, U_{is} - U_{i0}\} \mid u_{i0}, u_{is}]$. Offers provide option value, so applicants could apply to s even if $u_{is} < u_{i0}$.

Applicants choose A_i to solve

$$\max_{A \in \mathcal{A}_i} \sum_{s=1}^{S} w_{is} \times p_{is}(A) - C_i(A), \tag{3}$$

where w_{is} is defined as above, $p_{is}(A)$ denotes the perceived probability of receiving an offer (where we suppress dependence on priorities for notational simplicity), and $C_i(A)$ is a cost function.

[24] We abstract from the decision of whether to initially participate in the G&T match and assume a fixed set of applicants. While we cannot directly test for changes in the application decision, the absence of increases in the G&T application rates in Appendix Figure A2 or changes in the demographics of applicants in Appendix Figure A3 provides support for this assumption.

We parameterize utility u_{is} over G&T programs as a linear function of program fixed effects, distance, sibling and district priorities, and a Type 1 extreme value taste shock which is independent across students.[25] We normalize the value of the outside option so that $u_{i0} = \epsilon_{i0}$.

After offers are realized, but prior to enrolling, utilities are updated with an additional (independent) preference shock $\xi_{is} \sim \text{Gumbel}(0, \kappa)$ for each program. The shocks could arise from, for example, information learned on school visits, knowledge about where friends received offers, or changes in beliefs about school quality. Combining the second stage shock with first stage indirect utility yields enrollment-stage utilities that take the form: $U_{is} = u_{is} + \xi_{is}$. By standard arguments, expected utility can be written as $w_{is} = \kappa \log(\exp(\frac{u_{is} - u_{i0}}{\kappa}) + 1)$. That is, the expected benefit of an offer at s increases in the first stage utility u_{is} and the extent to which preferences can change, governed by κ.

Applicants have rational expectations over assignment probabilities as in Agarwal and Somaini (2018). That is, applicants know their true assignment probabilities across programs given their ROL and priorities, as well as the distribution of preferences and priorities in the district. Agarwal and Somaini (2018) show that a consistent estimate for p_{is} can be generated by bootstrapping the match and computing offer shares across simulations. Appendix G.1 provides further details on the procedure.

While there is no direct monetary cost for submitting the G&T application, applicants may incur time, psychological, or other monetary costs in the process of researching and applying to schools. Such costs rationalize ROLs that are shorter than the maximum length of 12. We assume that cost scales linearly in application length such that $C_i(A) = \zeta_i|A|$, $\zeta_i \sim \text{HN}(\sigma_\zeta)$, where HN denotes a half-normal distribution. The cost shocks are independent of taste shocks ϵ_{is} and ξ_{is}. We obtain similar results when parameterizing ζ_i to instead follow a truncated normal distribution, which yields estimates of location and lower bound parameters virtually equal to zero.

The DA algorithm generates either a single offer to a G&T program or no offer at all. Let $O_i \in \{0, 1, ..., S\}$ denote the offer outcome for each applicant. Applicants without any offer have $O_i = 0$. Preference shocks ξ_{is} are realized after assignments but before enrollment. Let $S_i \in \{0, 1, ..., S\}$ denote the school where families choose to enroll. Families with an offer at s enroll in s if and only if $U_{is} > U_{i0}$. Those who turn down their assignment enroll in the outside option.

[25]We define distance as the number of miles between the centroid of the applicant's tract of residence and the school.

6.1.3 Identification

Agarwal and Somaini (2018) show that indirect utilities are non-parametrically identified when a preference shifter has large support, enters the utility function additively, and is conditionally independent of unobserved tastes. In our application, distance plays the role of this shifter. As discussed in Idoux (2021), estimates of cost parameters are driven by variation in assignment probabilities conditional on observables. Holding utility fixed, such variation shifts the option value of adding a program to the list, so the margin at which applicants stop listing reveals the level of costs. Variation in ROL length among observationally-equivalent applicants pins down dispersion in application costs, σ_ς. Finally, dispersion in the distribution of post-offer preference shocks, κ, is driven by take-up behavior. Given indirect utility, κ increases in the rate at which applicants decline offers at programs they appear to prefer to the outside option.

6.1.4 Estimation

A student's likelihood contribution takes the form:

$$\begin{aligned} &L_i(A_i, O_i, S_i \mid W_i) \\ &= \int\int \Pr(A_i \mid W_i, \epsilon_i, c_i) \Pr(O_i \mid A_i, W_i) \Pr(S_i \mid Z_i, A_i, O_i, W_i, \epsilon_i) dF(\epsilon_i) dF(c_i). \end{aligned} \tag{4}$$

We estimate the model using data on the cohort of eligible applicants enrolling in G&T in 2019, as this is the last year prior to the admissions reform and COVID-19 pandemic. The pandemic may have influenced compliance behavior independent of the reform. We restrict to applicants who submitted a ROL and have non-missing distance data, which leaves a sample of 2,363 applicants (64% of the full sample of applicants who scored 90 or above).[26]

The number of potential ROLs renders estimation computationally infeasible, so we impose several restrictions to make the problem tractable. We restrict applicant choice sets to programs within a 3 mile radius and citywide programs (if eligible). Furthermore, we use only the top four choices from an applicant's ROL.[27]

[26]While we have location data for almost all families, many applicants are not in the structural estimation sample because families submit ROLs after receiving their G&T score. Therefore, it is possible for entrance exam test-takers to receive their score, learn they are ineligible, or learn they are eligible but decide to not submit a ROL.

[27]77% of ROLs include only programs within a 3 mile radius; 95% of top choices are programs within a 3 mile radius. 62% of applicants submit a ROL of four choices or fewer, and 97% enroll in a program that they ranked as one of their top four choices.

We estimate the parameters of our model via simulated maximum likelihood. The likelihood has no closed form, so we approximate it with a logit kernel smoother. Appendix G details the likelihood and optimization procedure.

6.1.5 Parameter Estimates and Model Fit

Our procedure recovers plausible parameter estimates that are consistent with prior work. As reported in Appendix Table G1, utility is strongly decreasing in distance and increasing in priority status. Applicants receive sizable shocks after offers are realized, which rationalizes never-taking behavior. Application costs are small but non-negligible, with an estimated 8% of applicants having costs smaller than 0.01.

The simulated distribution of students enrolled in G&T aligns closely with their observed counterparts, suggesting our model fits the data well. Panel A of Figure 8 shows that the simulated distribution aligns with observed FRPL and entrance exam score enrollment shares, despite the fact that neither was used to model utility. Appendix Table G2 shows that the model also reproduces other demographic shares.

6.2 Simulated Admissions Reform Impacts

To simulate the impacts of the reform, we eliminate the entrance exam priority groups, include citywide programs in the choice set of all applicants, and update subjective probabilities by re-estimating assignment probabilities.[28] We hold fixed the 2019 applicant pool for this exercise to approximate the pool that would have applied under the reform in the absence of other shocks.[29] All G&T applicants are eligible for G&T in our simulation, regardless of their score on the G&T exam. This is consistent with the extremely high share (91%) of applicants in the recommendation system successfully receiving a recommendation. We simulate enrollment under the reform with 200 draws for the parameters, latent random variables, and lottery tie-breakers.

We find that the reform slightly increased the FRPL share and greatly decreased the baseline ability of G&T students. Panel B of Figure 8 depicts estimates of the share of G&T students by baseline score and FRPL. We find a slight increase in the FRPL share to 27.9%. This closely matches the observed FRPL share of 28.2% in 2021, which we interpret

[28]The resulting perceived offer probabilities are not perfectly rational, as they do not take into account the changes in other applicants' ROLs. However, this may be realistic for the first year in which a new admissions regime is implemented, before applicants have a chance to observe how others respond.

[29]We support the validity of this approximation by noting that from 2019 to 2020, neither the overall application rate to G&T nor the demographic composition of applicants changed significantly. We interpret this as evidence of little extensive margin response. See Appendix Figure A2 and Appendix Figure A3.

as another out-of-sample validation of model fit. There is also a sharp decline in baseline ability for both FRPL and non-FRPL G&T students. Almost 70% of those who enroll after the reform would not have been eligible before it.

We simulate the ATT under the reform by re-weighting our reduced form estimates by entrance exam score and FRPL status to match the post-reform enrollment distribution. Two additional assumptions underlie this exercise: effects for ineligible students pre-reform equal those for the lowest-scoring eligible applicants (results are robust to alternative imputation methods), and effect heterogeneity reflects student-level characteristics rather than program-level sorting. We formalize that argument in Appendix C.8.

Table 8 reports the results for middle school application outcomes.[30] As a benchmark, column 1 reports estimates from re-weighting treatment effects on FRPL status alone. The ATTs change little under the reform. However, changes in the ATT differ markedly when we also re-weight on the entrance exam score, as reported in columns 2-4. Each column corresponds to an alternative method of estimating the conditional average treatment effects. Across specifications, the reform substantially decreases the ATT. For example, column 2 estimates group-specific treatment effects in the pre-reform period using FRPL-specific linear fits as depicted in Appendix Figure F2. The ATT on the specialized high school share of applicants' first choice school declined from $+5$ percentage points to -5 percentage points. The pattern is explained by the large decline in the baseline ability of new G&T students who also have smaller treatment effects. Declines in the value-added ATT are likewise sizable.

We conclude by asking how the existing policies compare to what is theoretically attainable. Figure 9 depicts a possibilities frontier with ATTs on the y-axis and FRPL shares on the x-axis. The solid line shows the maximum possible ATT at each FRPL share, representing the possibilities frontier for a policymaker who places positive weight on both equity and efficiency. The dashed line shows the minimum possible ATT at each FRPL share, representing a lower bound of G&T program performance. The red triangle shows where an exam-based admissions system falls between these two benchmarks and the blue circle corresponds to the recommendation system. Neither of the existing policies lies on the frontier. This implies that admissions reforms need not entail an equity-efficiency trade-off relative to either of the existing policies. The optimal policy, which admits high-ability FRPL students, could simultaneously increase the FRPL share and the ATT.

[30] Simulating impacts on middle school enrollment is complicated by congestion effects in the middle school match. Nevertheless, we find similar results on the enrollment margin.

7 Conclusion

This paper studies the extent to which kindergarten G&T programs impact achievement and access to elite secondary schools. We do so using unique features of the NYC setting that permit identification of effects for both marginal students at a qualifying cutoff and inframarginal students who score well above it. We find that G&T has little effect on achievement through middle school, with little evidence of positive effects either by demographic subgroups or by entrance exam score. An interesting question for future work is whether NYC G&T generates gains in long-run outcomes despite small impacts on achievement, as in Card et al. (2024).

Moving beyond achievement, we find that kindergarten G&T shapes middle school applications and enrollment, especially for low-income students and those scoring well above the qualifying cutoff. An important topic for future work is disentangling mechanisms by which G&T and similar tracking programs impact school choice decisions.

Finally, we find that a reform that replaced an entrance exam with an admissions system based on teacher recommendations increased the representation of low-income students but significantly reduced the positive effects of G&T on middle school access. Our possibilities frontier, however, suggests that there need not be such an equity-efficiency trade-off. Hypothetical alternatives could improve on both policies by simultaneously increasing diversity and average treatment effects. Designing realistic policies that do not incur this trade-off, by moving closer to the frontier or by expanding the frontier itself, remains a key task for future work. Universal screening, which NYC implemented in 2022, could prove promising.

References

Abdulkadiroğlu, A., J. D. Angrist, and P. A. Pathak (2014): "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools," *Econometrica*, 82, 137–196.

Abdulkadiroğlu, A., J. D. Angrist, Y. Narita, and P. A. Pathak (2017): "Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation," *Econometrica*, 85, 1373–1432.

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

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

Andrabi, T., J. Das, and A. I. Khwaja (2017): "Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets," *American Economic Review*, 107, 1535–1563.

Angrist, J., P. Hull, P. A. Pathak, and C. Walters (2024a): "Credible School Value-Added with Undersubscribed School Lotteries," *Review of Economics and Statistics*, 106, 1–19.

Angrist, J. D., P. D. Hull, P. A. Pathak, and C. R. Walters (2017): "Leveraging Lotteries for School Value-Added: Testing and Estimation," *Quarterly Journal of Economics*, 132, 871–919.

——— (2024b): "Race and the Mismeasure of School Quality," *American Economic Review: Insights*, 1, 20–37.

Angrist, J. D. and V. Lavy (1999): "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, 114, 533–575.

Angrist, J. D., V. Lavy, J. Leder-Luis, and A. Shany (2019): "Maimonides Rule Redux," *American Economic Review: Insights*, 1, 309–324.

Angrist, J. D. and M. Rokkanen (2015): "Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away From the Cutoff," *Journal of the American Statistical Association*, 110, 1331–1344.

Barrow, L., L. Sartain, and M. de la Torre (2020): "Increasing Access to Selective High Schools Through Place-Based Affirmative Action: Unintended Consequences," *American Economic Journal: Applied Economics*, 12, 135–63.

Behaghel, L., B. Crépon, and M. Gurgand (2013): "Robustness of the Encouragement Design in a Two-Treatment Randomized Control Trial," Working Paper.

Bhuller, M. and H. Sigstad (2024): "2SLS with Multiple Treatments," *Journal of Econometrics*, 242, 105785.

Bleemer, Z. (2020): "Top Percent Policies and the Return to Postsecondary Selectivity," Working Paper.

Bryan, C. (2024): "Why Seattle Public Schools is Closing its Highly Capable Cohort Program," *The Seattle Times*, https://www.seattletimes.com/education-lab/why-seattle-public-schools-is-closing-its-highly-capable-cohort-program/.

BUI, S. A., S. G. CRAIG, AND S. A. IMBERMAN (2014): "Is Gifted Education a Bright Idea? Assessing the Impact of Gifted and Talented Programs on Students," *American Economic Journal: Economic Policy*, 6, 30–62.

CALONICO, S., M. D. CATTANEO, M. H. FARRELL, AND R. TITIUNIK (2019): "Regression Discontinuity Designs Using Covariates," *Review of Economics and Statistics*, 101, 442–451.

CALONICO, S., M. D. CATTANEO, AND R. TITIUNIK (2014): "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82, 2295–2326.

CARD, D., E. CHYN, AND L. GIULIANO (2024): "Can Gifted Education Help Higher-Ability Boys from Disadvantaged Backgrounds?" NBER Working Paper No. 33282.

CARD, D. AND L. GIULIANO (2014): "Does Gifted Education Work? For Which Students?" NBER Working Paper No. 20453.

——— (2016a): "Can Tracking Raise the Test Scores of High-Ability Minority Students?" *American Economic Review*, 106, 2783–2816.

——— (2016b): "Universal Screening Increases the Representation of Low-income and Minority Students in Gifted Education," *Proceedings of the National Academy of Sciences*, 113, 13678–13683.

CARLANA, M., E. LA FERRARA, AND P. PINOTTI (2022): "Goals and Gaps: Educational Careers of Immigrant Children," *Econometrica*, 90, 1–29.

CHETTY, R., M. O. JACKSON, T. KUCHLER, J. STROEBEL, N. HENDREN, R. B. FLUEGGE, S. GONG, F. GONZALEZ, A. GRONDIN, M. JACOB, ET AL. (2022): "Social Capital II: Determinants of Economic Connectedness," *Nature*, 608, 122–134.

CLEVELAND, C. (2023): "Gifted or Gone: The Effects of Gifted Education on Public School Enrollment and Achievement," Working Paper.

CLOSSON, T. (2024): "Court Allows Case Challenging Segregation in N.Y.C. Schools to Advance," *The New York Times*, https://www.nytimes.com/2024/05/02/nyregion/nyc-schools-segregation.html.

COHODES, S. R. (2020): "The Long-Run Impacts of Specialized Programming for High-Achieving Students," *American Economic Journal: Economic Policy*, 12, 127–66.

CORCORAN, S. P., J. L. JENNINGS, S. R. COHODES, AND C. SATTIN-BAJAJ (2018): "Leveling the Playing Field for High School Choice: Results from a Field Experiment of Informational Interventions," NBER Working Paper No. 24471.

CORRADINI, V. (2024): "Information and Access in School Choice Systems: Evidence from New York City," Working Paper.

CUNHA, F. AND J. HECKMAN (2007): "The Technology of Skill Formation," *American Economic Review*, 97, 31–47.

DOBBIE, W. AND R. G. FRYER, JR (2014): "The Impact of Attending a School with High-

Achieving Peers: Evidence from the New York City Exam Schools," *American Economic Journal: Applied Economics*, 6, 58–75.

DONG, Y. (2018): "Alternative Assumptions to Identify LATE in Fuzzy Regression Discontinuity Designs," *Oxford Bulletin of Economics and Statistics*, 80, 1020–1027.

DUFLO, E., P. DUPAS, AND M. KREMER (2011): "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya," *American Economic Review*, 101, 1739–1774.

ELLISON, G. AND P. A. PATHAK (2021): "The Efficiency of Race-Neutral Alternatives to Race-Based Affirmative Action: Evidence from Chicago's Exam Schools," *American Economic Review*, 111, 943–975.

FADULU, L. (2022): "New York City to Expand Gifted and Talented Program but Scrap Test," *The New York Times*, https://www.nytimes.com/2022/04/14/nyregion/nyc-gifted-talented.html.

FREDRIKSSON, P., B. ÖCKERT, AND H. OOSTERBEEK (2013): "Long-Term Effects of Class Size," *Quarterly Journal of Economics*, 128, 249–285.

GOLDSTEIN, D. (2018): "Rethinking What Gifted Education Means, and Whom It Should Serve," *The New York Times*, https://www.nytimes.com/2018/09/13/us/education-gifted-students.html.

GOLDSTEIN, Y., V. LAVY, AND G. RACHKOVSKI (2026): "Gifted Children Programs' Short and Long-Term Impact: Higher Education, Earnings, and the Knowledge Economy," NBER Working Paper No. 29779.

HAHM, D. W. AND M. PARK (2024): "A Dynamic Framework of School Choice: Effects of Middle Schools on High School Choice," Working Paper.

HAHN, J., P. TODD, AND W. VAN DER KLAAUW (2001): "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69, 201–209.

HASTINGS, J. S. AND J. M. WEINSTEIN (2008): "Information, School Choice, and Academic Achievement: Evidence from Two Experiments," *Quarterly Journal of Economics*, 123, 1373–1414.

HEINESEN, E., C. HVID, L. KIRKEBØEN, E. LEUVEN, AND M. MOGSTAD (2022): "Instrumental Variables with Unordered Treatments: Theory and Evidence from Returns to Fields of Study," NBER Working Paper No. 30574.

HOXBY, C. AND C. AVERY (2013): "The Missing "One-Offs": The Hidden Supply of High-Achieving, Low-Income Students," *Brookings Papers on Economic Activity*, 2013, 1–65.

HOXBY, C. M. (2000): "The Effects of Class Size on Student Achievement: New Evidence from Population Variation," *Quarterly Journal of Economics*, 115, 1239–1285.

IDOUX, C. (2021): "Integrating New York City Schools: The Role of Admission Criteria and Family Preferences," MIT Department of Economics, Working Paper, June.

IDOUX, C. AND V. CORRADINI (2023): "Overcoming Racial Gaps in School Preferences: the Role of Peer Diversity on School Choice," NBER Working Paper No. 33179.

IMBENS, G. W. AND K. KALYANARAMAN (2012): "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 79, 933–959.

INTEGRATENYC V. NEW YORK (2024): "2024 NY Slip Op 02369," Appellate Division, First Department, decided May 2, 2024.

KAPOR, A. (2024): "Transparency and Percent Plans," NBER Working Paper No. 32372.

KIRKEBOEN, L. J., E. LEUVEN, AND M. MOGSTAD (2016): "Field of Study, Earnings, and Self-Selection," *Quarterly Journal of Economics*, 131, 1057–1111.

KOLESÁR, M., R. CHETTY, J. FRIEDMAN, E. GLAESER, AND G. W. IMBENS (2015): "Identification and Inference With Many Invalid Instruments," *Journal of Business & Economic Statistics*, 33, 474–484.

KRUEGER, A. B. (1999): "Experimental Estimates of Education Production Functions," *Quarterly Journal of Economics*, 114, 497–532.

LEE, J. (2018): "See Where New York City's Elite High Schools Get Their Students," *The New York Times*, https://www.nytimes.com/interactive/2018/06/29/nyregion/nyc-high-schools-middle-schools-shsat-students.html.

LOVELESS, T. (2013): "The Resurgence of Ability Grouping and Persistence of Tracking," *Part II of the 2013 Brown Center Report on American Education, the Brookings Institution*, 1.

REARDON, S. F. (2018): "The Widening Academic Achievement Gap Between the Rich and the Poor," in *Social Stratification*, Routledge, 536–550.

RODA, A. (2015): *Inequality in Gifted and Talented Programs: Parental Choices About Status, School Opportunity, and Second-Generation Segregation*, Springer.

ROSENBAUM, P. R. AND D. B. RUBIN (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55.

SHAPIRO, E. (2021): "De Blasio to Phase Out N.Y.C. Gifted and Talented Program," *The New York Times*, https://www.nytimes.com/2021/10/08/nyregion/gifted-talented-nyc-schools.html.

STRICKLAND, K. J., W. CHAN, M. GOTTFRIED, J. HUANG, AND D. HILDRETH (2024): "Exploring the Impact of New York Citys Gifted and Talented Program: A Matched Comparison Study," *AERA Open*, 10, 1–15.

THOMPSON, O. (2021): "Gifted & Talented Programs and Racial Segregation," NBER Working Paper No. 29546.

URQUIOLA, M. S., O. MALAMUD, A. MITRUT, AND C. POP-ELECHES (2024): "Self-, Peer-, and Teacher Perceptions Under School Tracking," NBER Working Paper No. 32892.

WALTERS, C. R. (2015): "Inputs in The Production of Early Childhood Human Capital: Evidence from Head Start," *American Economic Journal: Applied Economics*, 7, 76–102.

Tables

Table 1. Summary Statistics

	All kindergarten students	G&T kindergarten applicants	RD sample	Lottery sample
	(1)	(2)	(3)	(4)
A: Race and Ethnicity				
Hispanic	0.41	0.17	0.12	0.07
Black, non-Hispanic	0.22	0.15	0.09	0.05
White, non-Hispanic	0.17	0.27	0.34	0.41
Asian, non-Hispanic	0.17	0.23	0.27	0.27
Missing or other race	0.03	0.18	0.19	0.20
B: Other Demographics				
Female	0.49	0.52	0.52	0.53
FRPL	0.72	0.35	0.24	0.16
Student with disabilities	0.16	0.06	0.05	0.04
English language learner	0.21	0.01	0.01	0.02
C: Qualification and Enrollment				
District eligible	0.06	0.26	0.56	1.00
Citywide eligible	0.02	0.11	0.00	0.59
G&T offer	0.03	0.17	0.27	0.53
G&T kindergarten enrollment	0.03	0.11	0.16	0.41
G&T ever enrollment	0.03	0.16	0.26	0.48
Entrance exam score		70.3	90.0	96.4
N	769,881	171,971	27,885	5,001

Notes: This table reports descriptive statistics for NYC kindergarten students from 2011-2012 to 2020-2021. The mean value of each variable is shown, based on the sample indicated in each column. Column 1 reports means for all NYC kindergarten students. Column 2 reports means for applicants to kindergarten G&T. Column 3 reports means for kindergarten G&T applicants in the RD sample. Column 4 reports means for kindergarten G&T applicants in the lottery IV sample. Means in Panel B are limited to students for whom demographics are non-missing.

Table 2. First Stage Effects of Eligibility and Lottery Offer Instruments

	RD (1)	Lottery (2)
Enroll next year	0.239***	0.412***
	(0.007)	(0.014)
N	27,885	5,001
Control mean	0.001	0.188
Ever enroll	0.155***	0.414***
	(0.011)	(0.016)
N	20,418	3,790
Control mean	0.129	0.256
Years enrolled	0.748***	2.292***
	(0.050)	(0.083)
N	20,418	3,790
Control mean	0.465	1.071

Notes: This table reports estimates of the effect of each instrument on first-stage outcomes. The enroll next year outcome is a dummy variable that equals one if the applicant enrolled in G&T in the year following application. The ever enroll outcome indicates whether an applicant ever enrolled in G&T in grades K-5. Years enrolled tracks the number of years in G&T through 5th grade. The ever enroll and years enrolled outcomes use cohorts through the 2018-19 school year, the last cohort for which we observe 5th grade enrollment. Column 1 reports effects of the RD eligibility instrument. Column 2 reports effects of the G&T offer instrument. Control means are the means of each outcome for applicants in the sample who do not enroll in G&T. Robust standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3. Effects on Kindergarten Classroom Characteristics

	RD (1)	Lottery (2)
Class Size	–4.104***	–0.112
	(0.352)	(0.278)
N	20,854	5,001
First stage F	1517	874
Non-G&T Mean	22.332	24.267
FRPL Share	–0.261***	–0.126***
	(0.024)	(0.016)
N	20,854	4,047
First stage F	1517	1017
Non-G&T Mean	0.452	0.355
White/Asian Share	0.138***	0.092***
	(0.021)	(0.013)
N	20,854	4,047
First stage F	1517	1017
Non-G&T Mean	0.627	0.669

Notes: This table reports 2SLS estimates of the effect of G&T enrollment on kindergarten classroom characteristics. Column 1 reports fuzzy RD estimates and column 2 reports estimates using the lottery design. Class size is the number of students in a student's kindergarten class. The FRPL share is the share of FRPL-eligible classmates in the kindergarten class (excluding the G&T applicant). The White/Asian share is the same share of classmates who are White or Asian. The table reports Kleibergen-Paap F-statistics and outcome means for students who do not enroll in G&T. Robust standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4. Effects on Achievement

	RD (1)	Lottery (2)	Pooled (3)
Math	–0.071	–0.039	–0.051
	(0.063)	(0.045)	(0.037)
N (estimation)	41,126	8,097	49,223
N (students)	17,119	3,266	19,442
First stage F	282	236	253
Non-G&T Mean	1.07	1.30	1.10
Over-id p-value	[0.159]	[0.340]	[0.283]
ELA	0.040	–0.023	0.002
	(0.059)	(0.041)	(0.034)
N (estimation)	50,104	9,712	59,816
N (students)	17,254	3,281	19,587
First stage F	191	172	179
Non-G&T Mean	1.02	1.21	1.04
Over-id p-value	[0.702]	[0.300]	[0.540]
Algebra Regents	0.081	–0.003	0.018
	(0.100)	(0.049)	(0.045)
N (estimation)	4,306	918	5,224
N (students)	4,306	918	5,008
First stage F	370	693	530
Non-G&T Mean	1.02	1.15	1.03
Over-id p-value			[0.449]

Notes: This table reports 2SLS estimates of the G&T causal effects on test scores. Column 1 reports fuzzy RD estimates and column 2 reports estimates using the lottery design. ELA and math achievement scores are New York State assessments, standardized to be mean zero and unit variance by test, grade, and year. Samples for ELA are stacked across grades 3-8, and samples for math are stacked across grades 3-6. The estimation procedure is described in Appendix C.3. The Algebra I outcome is based on a student's maximum score on the Regents exam, standardized to be mean zero and unit variance by year. Column 3 reports estimates that pool across the RD and lottery samples using the procedure described in Appendix C.3. The table reports Kleibergen-Paap F-statistics and outcome means for students who do not enroll in G&T. Standard errors are in parentheses. For ELA, math, and pooled outcomes, standard errors are clustered by student. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 5. Effects on 6th Grade Applications and Enrollment

	RD (1)	Lottery (2)
A: First Choice 6th Grade School		
Specialized HS Share	–0.002	0.071***
	(0.019)	(0.013)
N	13,101	2,582
First stage F	971.5	795.1
Non-G&T Mean	0.198	0.248
Value-Added	–0.001	0.033***
	(0.019)	(0.013)
N	13,101	2,582
First stage F	971.5	795.1
Non-G&T Mean	0.065	0.087
B: Enrolled 6th Grade School		
Specialized HS Share	–0.026	0.090***
	(0.018)	(0.013)
N	10,400	2,099
First stage F	798.7	779.7
Non-G&T Mean	0.149	0.202
Value-Added	–0.003	0.021*
	(0.020)	(0.012)
N	10,400	2,099
First stage F	798.7	779.7
Non-G&T Mean	0.051	0.080

Notes: This table reports 2SLS estimates of the effect of G&T enrollment on the characteristics of middle schools to which students apply and enroll. Column 1 reports fuzzy RD estimates and column 2 reports estimates using the lottery design. The specialized high school share is the proportion of 6th grade students at each middle school who later enroll in a specialized high school. Value-added is the estimated value-added of the middle school on 6th grade math scores. Panel A reports estimates for outcomes corresponding to the school that students ranked first on their middle school application. Panel B reports estimates corresponding to the school that students enrolled in for 6th grade. The table reports Kleibergen-Paap F-statistics and outcome means for students who do not enroll in G&T. Robust standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 6. Effects on 6th Grade Applications and Enrollment by FRPL Status

	RD		Lottery	
	FRPL	Non-FRPL	FRPL	Non-FRPL
	(1)	(2)	(3)	(4)
A: First Choice 6th Grade School				
Specialized HS Share	0.002	–0.004	0.119***	0.064***
	(0.032)	(0.024)	(0.040)	(0.013)
N	3,318	9,756	459	2,121
First stage F	324.7	636.6	111.0	684.2
Non-G&T Mean	0.184	0.203	0.233	0.251
p-value	[0.851]		[0.202]	
Value-Added	0.028	–0.014	0.082**	0.026*
	(0.030)	(0.023)	(0.033)	(0.013)
N	3,318	9,756	459	2,121
First stage F	324.7	636.6	111.0	684.2
Non-G&T Mean	0.071	0.063	0.082	0.088
p-value	[0.249]		[0.129]	
B: Enrolled 6th Grade School				
Specialized HS Share	–0.073**	–0.008	0.120***	0.082***
	(0.029)	(0.023)	(0.031)	(0.014)
N	2,750	7,642	389	1,710
First stage F	264.1	517.3	119.5	664.1
Non-G&T Mean	0.125	0.158	0.164	0.211
p-value	[0.071]		[0.263]	
Value-Added	–0.029	0.007	0.074**	0.009
	(0.034)	(0.025)	(0.030)	(0.013)
N	2,750	7,642	389	1,710
First stage F	264.1	517.3	119.5	664.1
Non-G&T Mean	0.044	0.054	0.056	0.085
p-value	[0.402]		[0.055]	

Notes: This table reports 2SLS estimates of the effect of G&T enrollment on the characteristics of middle schools to which students apply and enroll, separately by FRPL status. See Table 5 for a description of the outcomes. The table reports Kleibergen-Paap F-statistics and outcome means for students who do not enroll in G&T. Robust standard errors are in parentheses. p-values for tests of the difference in G&T effects between FRPL and non-FRPL students are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 7. Multisector Model Estimates

	Achievement		Middle school enrollment	
	Math	ELA	Specialized HS Share	Value-Added
	(1)	(2)	(3)	(4)
A: All Applicants				
District	–0.022	0.020	0.039**	–0.005
	(0.055)	(0.052)	(0.016)	(0.016)
Citywide	–0.108**	–0.043	0.159***	0.015
	(0.044)	(0.042)	(0.011)	(0.012)
N (estimation)	12,554	14,967	3,292	3,292
N (students)	4,801	4,828	3,292	3,292
First stage F	92.0	63.0	301	301
B: FRPL				
District	–0.060	–0.079	0.138***	0.101**
	(0.150)	(0.160)	(0.040)	(0.043)
Citywide	–0.171	–0.180	0.268***	0.110***
	(0.134)	(0.141)	(0.033)	(0.037)
N (estimation)	2,107	2,533	568	568
N (students)	753	758	568	568
First stage F	14.0	10.0	52.0	52.0
C: Non-FRPL				
District	–0.018	0.030	0.021	–0.021
	(0.059)	(0.055)	(0.017)	(0.017)
Citywide	–0.100**	–0.032	0.143***	0.003
	(0.046)	(0.043)	(0.012)	(0.013)
N (estimation)	10,443	12,430	2,724	2,724
N (students)	4,046	4,068	2,724	2,724
First stage F	78.0	53.0	249	249

Notes: This table reports 2SLS estimates of the effect of district and citywide G&T enrollment. Estimates are obtained from a multisector model with two treatments and two offer instruments as described in Section C.4. The sample is restricted to those who are eligible for both district and citywide (scored above 97) and with non-degenerate assignment risk to either the district or citywide G&T sector. See Tables 4 and 5 for a description of the outcomes. The table reports Kleibergen-Paap F-statistics. Standard errors are in parentheses. For achievement outcomes, standard errors are clustered by student. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 8. ATT Estimates under Alternative Admissions Regimes

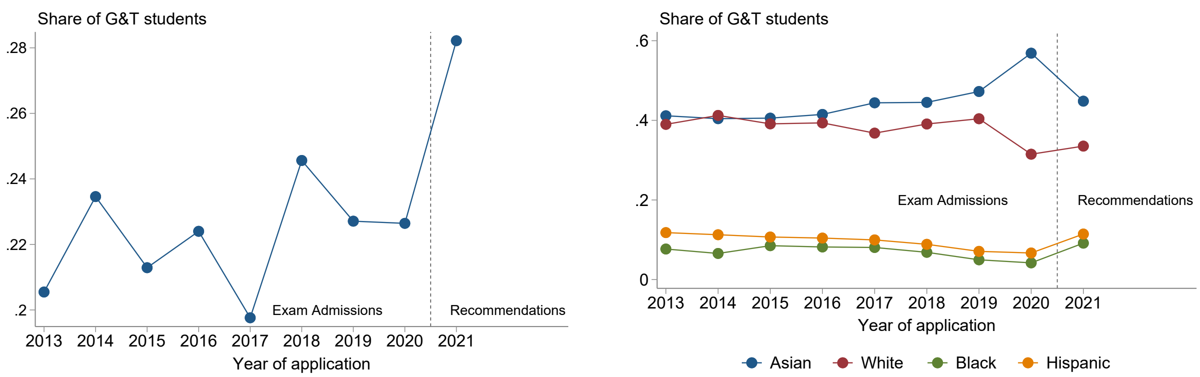
Weighting scheme:		FRPL and Exam Score		
	FRPL (1)	Linear (2)	Two Bins (3)	Four Bins (4)
A: Specialized HS Share of First Choice 6th Grade School				
Exam	0.077***	0.046**	0.062***	0.036*
	(0.014)	(0.020)	(0.016)	(0.020)
Recommendation	0.080***	–0.052	–0.010	–0.121
	(0.015)	(0.052)	(0.043)	(0.077)
B: Value-Added of First Choice 6th Grade School				
Exam	0.039***	0.016	0.029*	0.004
	(0.013)	(0.018)	(0.015)	(0.017)
Recommendation	0.042***	–0.053	–0.012	–0.075
	(0.013)	(0.047)	(0.039)	(0.063)

Notes: This table reports ATT estimates for G&T enrollment on characteristics of schools ranked first on students' middle school applications. See Table 5 for outcome descriptions. Each panel reports ATTs under exam- and recommendation-based admissions, obtained by reweighting conditional LATEs as described in Section 6. Column 1 reweights FRPL-specific effects. Columns 2-4 reweight effects estimated separately by FRPL and the entrance exam score. Column 2 uses 2SLS estimates that treat enrollment and its interaction with the exam score as endogenous variables, instrumented by lottery offers and their interaction with exam score. Columns 3-4 use 2SLS estimates by exam score bins (two bins: $\leq$94 and $\geq$95; four bins: $\leq$92, 93-95, 96-98, 99). Students in the recommendations period scoring below 90 are assigned the treatment effect of the lowest scoring group. * significant at 10%; ** significant at 5%; *** significant at 1%.

Figures

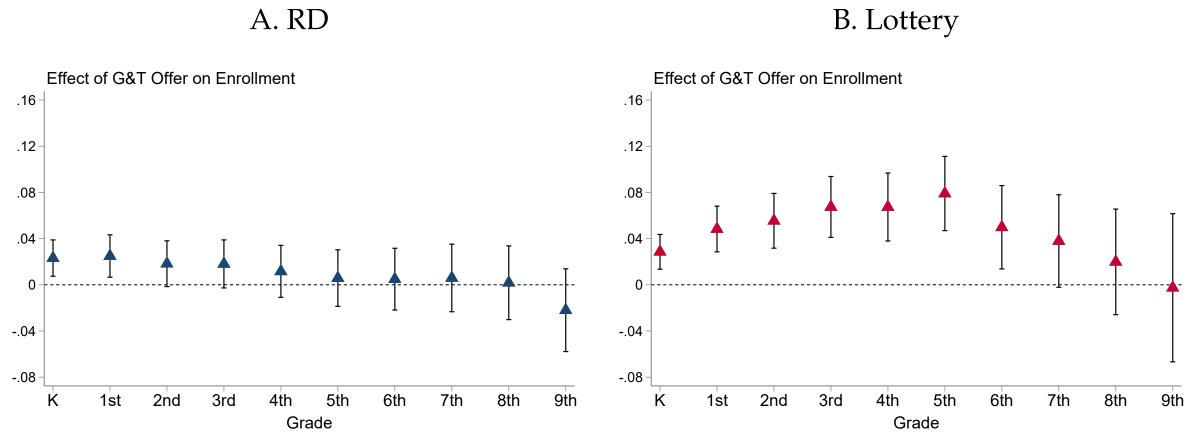
Figure 1. Trends in Enrollment Shares

A. FRPL B. Race and Ethnicity



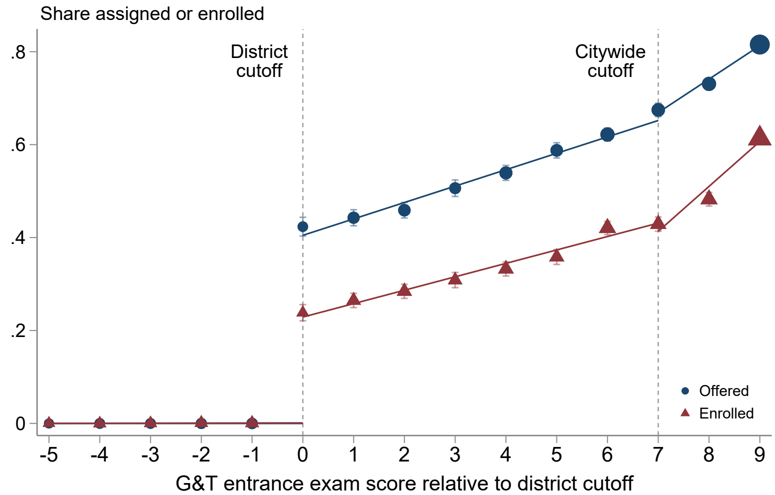
Notes: This figure depicts demographic shares of kindergarten G&T students over time. Panel A plots the share of students who are FRPL-eligible. Panel B plots the shares of students by race and ethnicity. Shares are computed in the sample of students with non-missing demographic data.

Figure 2. G&T Offer Effects on Retention in NYC Public Schools



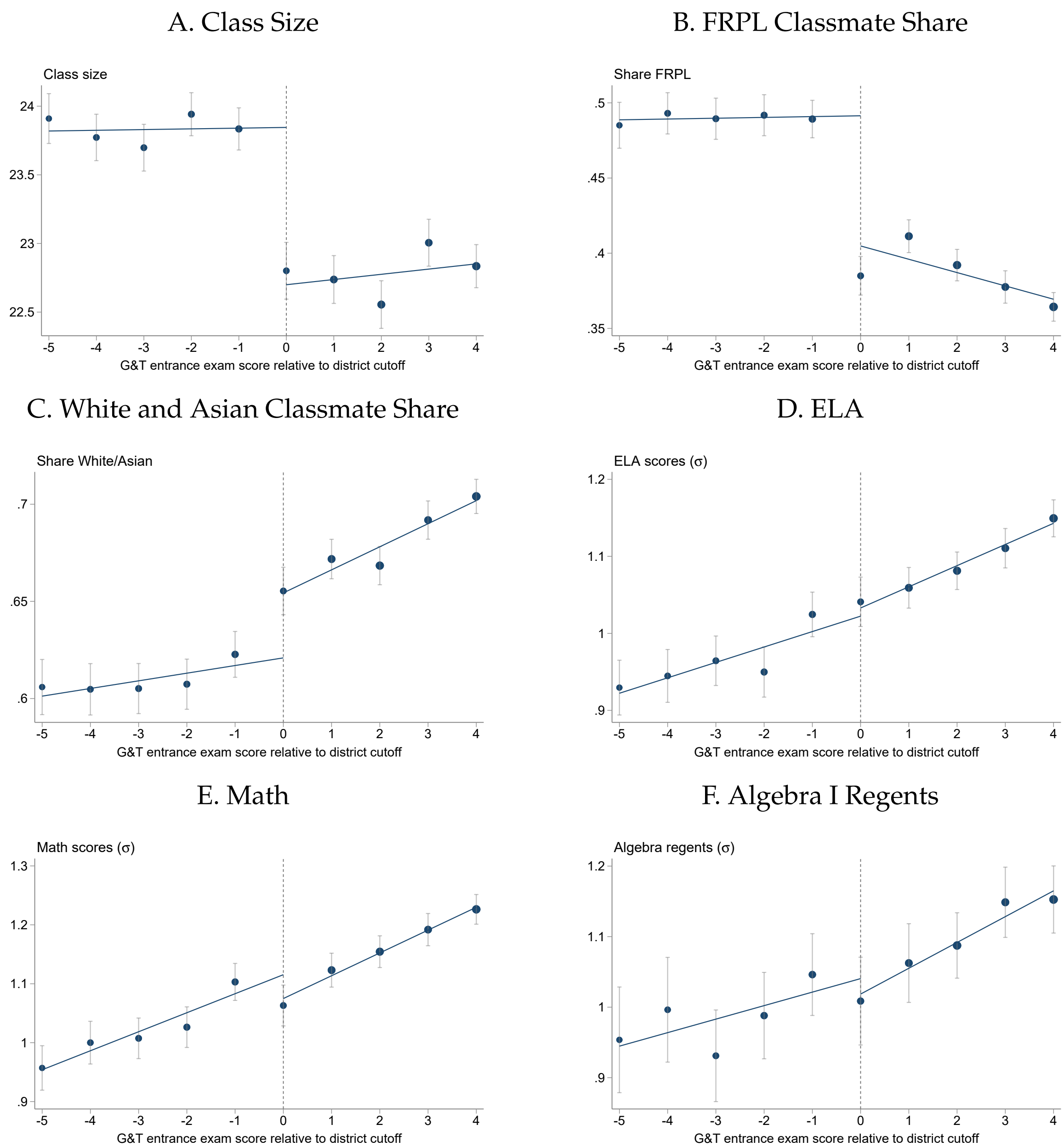
Notes: This figure depicts effect estimates of the impact of G&T offers on enrollment in a non-charter NYC public school. Panel A plots estimates obtained from a fuzzy RD procedure that instruments offer with eligibility. Panel B plots estimates from the reduced form effect of offers in the lottery IV design. Marker sizes are proportional to the number of students attaining each score value. Whiskers depict 95-percent confidence intervals.

Figure 3. G&T Qualification Increases the Probability of Assignment and Enrollment



Notes: This figure depicts the rate at which students receive offers and enroll in G&T as a function of the G&T entrance exam score. The blue circles correspond to offer shares, and the red triangles correspond to enrollment shares. The running variable is normalized so that 0 corresponds to the district G&T cutoff of 90. Dashed vertical lines indicate the district and citywide cutoffs. Marker sizes are proportional to the number of students attaining each score value. Whiskers depict 95% confidence intervals for the conditional means.

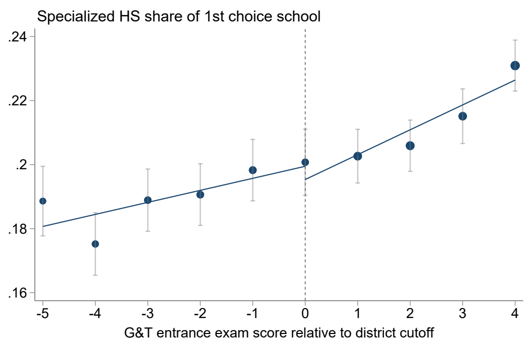
Figure 4. RD Effects on Kindergarten Classrooms and Achievement



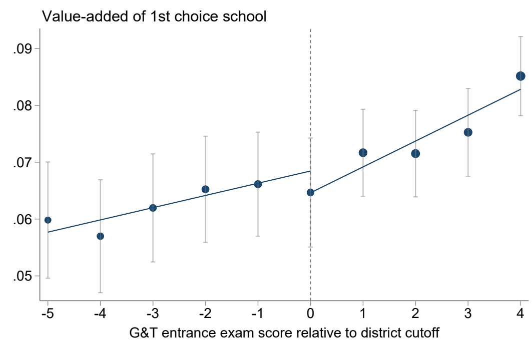
Notes: This figure depicts the reduced form impacts of G&T eligibility on kindergarten classroom characteristics and test scores in grades 3-8. Panel A depicts the impact on kindergarten class size. Panels B and C depict the impact on the share of classmates who are FRPL-eligible and who are White and Asian. The outcomes in Panels D, E, F are New York State assessments and Algebra I Regents. ELA outcomes stack across grades 3-8, and math outcomes stack samples across grades 3-6. See the notes to Tables 3 and 4 for further description of the outcomes. Marker sizes are proportional to the number of students attaining each score value. Whiskers depict 95-percent confidence intervals for the conditional means.

Figure 5. RD Effects on Middle School Outcomes

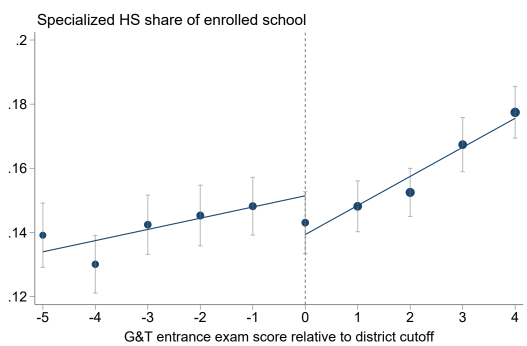
A. Specialized HS Share of First Choice Middle School



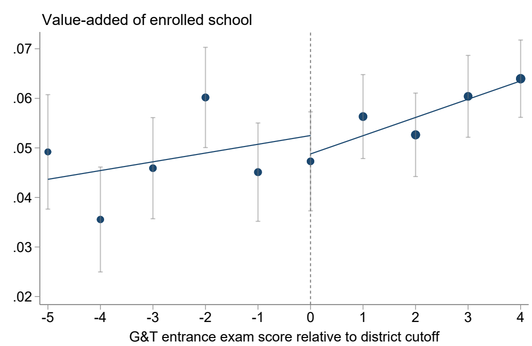
B. Value-Added of First Choice Middle School



C. Specialized HS Share of Enrolled Middle School

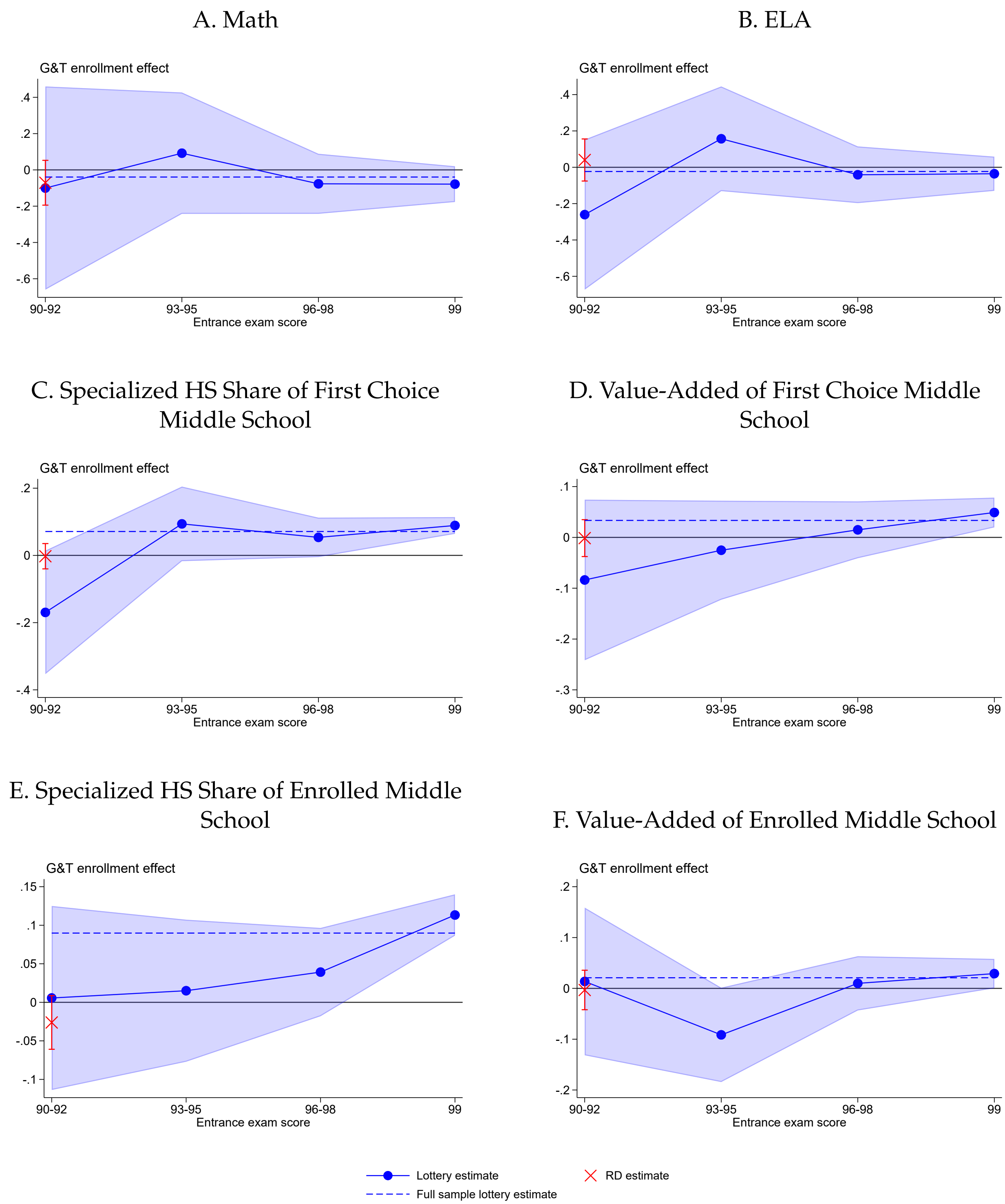


D. Value-Added of Enrolled Middle School



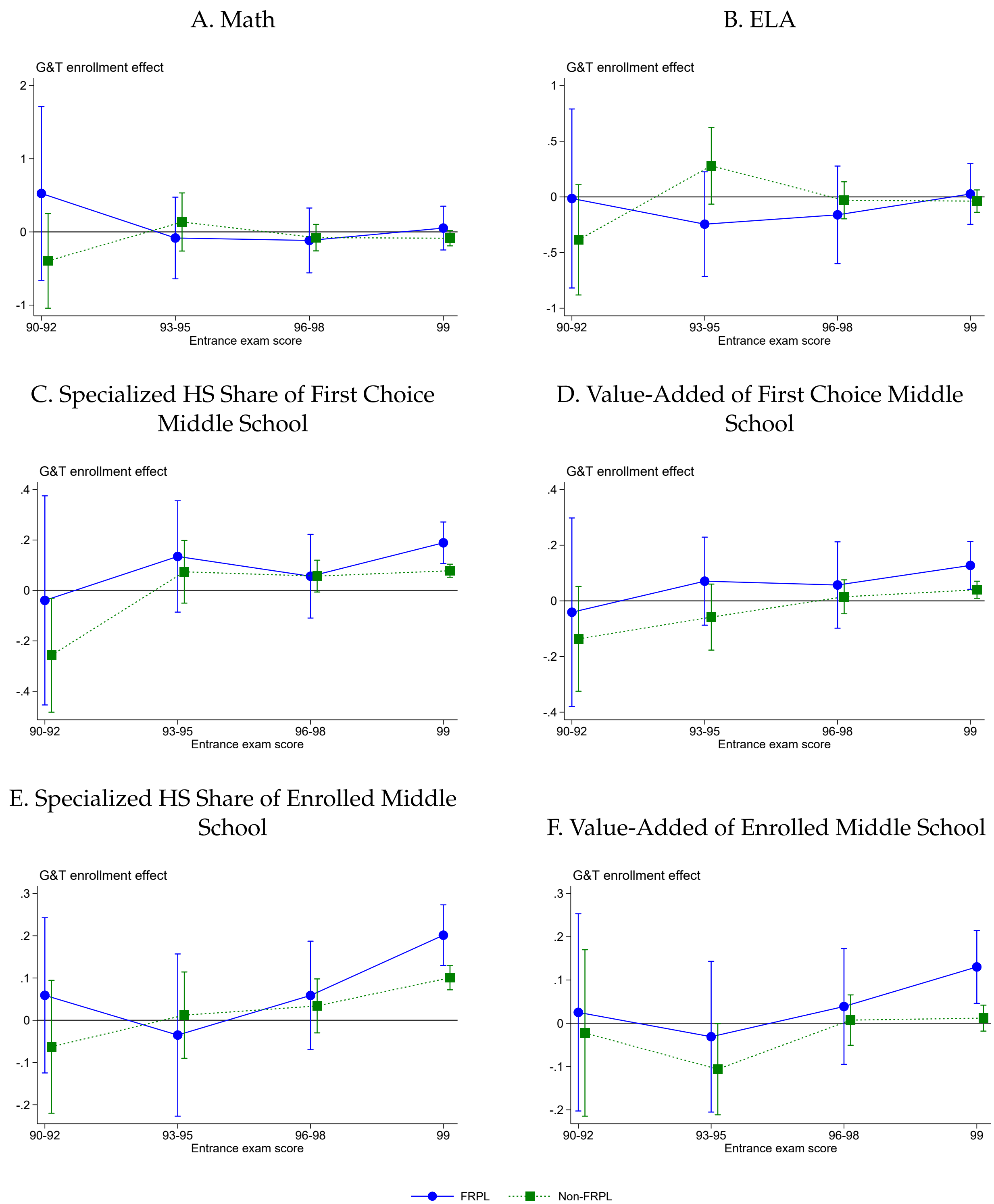
Notes: This figure depicts the reduced form impact of kindergarten G&T eligibility on secondary school application and enrollment outcomes. Panels A and B depict impacts on the specialized high school share and math value-added of a student's first choice on their middle school application. Panels C and D depict the corresponding effects on the schools at which students enroll for 6th grade. Marker sizes are proportional to the number of students attaining each score value. Whiskers depict 95-percent confidence intervals for the conditional means.

Figure 6. Effects by Entrance Exam Score



Notes: These figures depict lottery 2SLS estimates of the effects of G&T enrollment by bins of the entrance exam score. The blue shading corresponds to 95-percent confidence intervals. The blue dashed line depicts the 2SLS estimate obtained using the full sample. The red marker and whiskers show the point estimate and 95 percent confidence interval for the RD estimate.

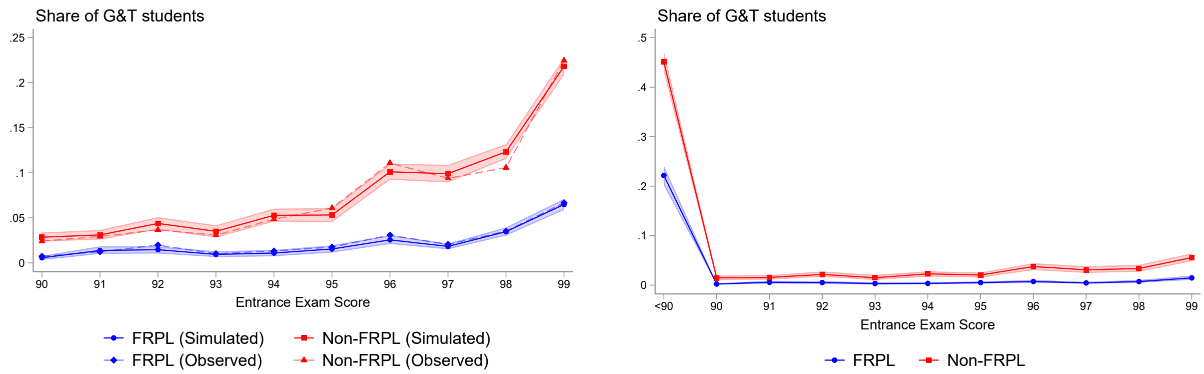
Figure 7. Effects by Entrance Exam Score and FRPL Status



Notes: These figures depict lottery 2SLS estimates of the effects of G&T enrollment by FRPL status and bins of the entrance exam score. The blue line corresponds to FRPL students, and the green line corresponds to non-FRPL students. The whiskers show the 95-percent confidence intervals.

Figure 8. Simulated G&T Enrollment

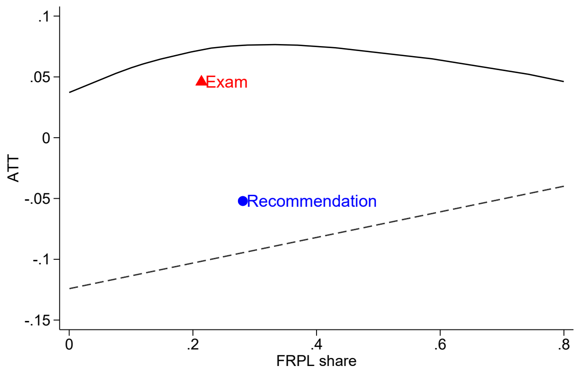
A. Exam Admissions B. Teacher Recommendation Admissions



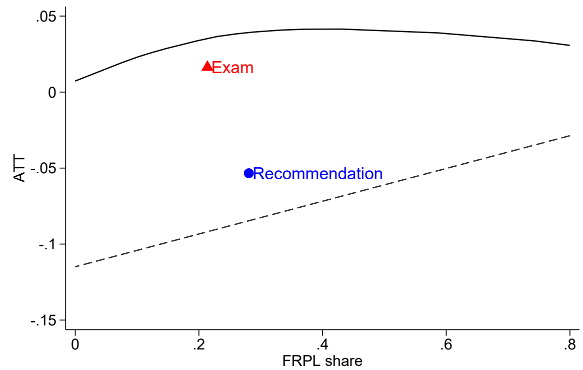
Notes: Panel A depicts the observed and simulated share of G&T students by entrance exam score and FRPL status in 2019. Enrollment is simulated by drawing parameters from a multivariate normal centered at the point estimates with empirical covariance matrix, solving for ROLs, running the DA algorithm, and simulating whether applicants accept or reject their offer. Program capacities are equal to the number of observed offers given by each program in our 2019 estimation sample. This procedure is repeated 200 times. 95 percent prediction intervals for the simulated shares are depicted by the shaded regions. Panel B depicts simulated enrollment shares under a recommendation system. See Section 6 for details.

Figure 9. Policy Possibility Frontiers

A. Specialized HS Share of First Choice School



B. Value-Added of First Choice School



Notes: This figure depicts possibility frontiers of hypothetical policies on the FRPL share and ATT of G&T enrollment. The solid and dashed lines correspond to the largest and smallest possible ATT at each value of the FRPL share. The lines are computed using estimates obtained from a 2SLS procedure corresponding to column 2 of Table 8. The red triangle and blue circle correspond to simulated outcomes under exam- and teacher recommendation-based admissions.

Sorting or Supporting? The Effect of Gifted Education on Achievement and Access: Supplemental Appendix

Jimmy Chin, UC Berkeley

Geoffrey Kocks, MIT

July 2026

A Institutional Details

A.1 NYC Kindergarten Options

NYC public school students begin kindergarten during the calendar year in which they turn five. Kindergarten students may enroll in either traditional non-G&T or G&T programs. The G&T and non-G&T have separate assignment processes. Families who apply to both G&T and non-G&T submit separate rank-ordered lists over programs, and lists are processed independently via separate matches (both using deferred acceptance).

The non-G&T programs do not use any tests or recommendations for admissions. Non-G&T zoned programs give priority to students living in the surrounding geographic area while unzoned schools may only give geographic priority to students living within the same district. NYC has 32 school districts, but each zoned kindergarten has its own specific non-overlapping zone. Within the non-G&T kindergarten programs, some may still be specialized, such as dual language or STEM-focused. A map of kindergarten G&T locations is shown in Appendix Figure A1.

A.2 G&T Program Details

NYC G&T consists of district and citywide programs. District G&T programs consist of "gifted" classrooms within otherwise integrated schools, and students may share some class time with students not identified as "gifted" in non-core classes like gym, music, and lunch. For district programs, admissions priority is given to students who live in the same school district. Citywide programs are in schools that only include "gifted" students, and there is no geographic priority based on district for these programs. There are five citywide G&T programs: NEST+M, the Anderson School, and TAG (Talented and Gifted School for Young Scholars) in Manhattan; Brooklyn School of Inquiry in Brooklyn; and PS/IS 300 (the 30th Avenue School) in Queens. Most district programs end in 5th grade, while all of the citywide programs except for NEST+M end in 8th grade. NEST+M continues through 12th grade.

Interviews with NYC G&T teachers suggest that while the G&T curriculum is largely the same as the standard curriculum, more students are near grade-level and teachers may use more project-based learning than a standard kindergarten classroom. There is no mandated curriculum for G&T. Teachers may have a Gifted Education Extension Certificate, but this is not a requirement for teaching in the program. Kindergarten is the main entry point for most programs, but some do not admit students until 3rd grade.

A.3 Admissions Exams

The admissions exam was introduced in 2008, was free of charge, offered in several languages, and administered by NYC teachers. From 2013 onwards, the G&T examination consisted of two subtests: the Naglieri Nonverbal Ability Test (NNAT) and Otis-Lennon School Ability Test (OLSAT). Prior to 2013, NYC used the Bracken School Readiness Assessment (BSRA) instead of the NNAT. Students were given an age-adjusted percentile rank from 1 to 99 on each exam. The ranks were converted to normal curve equivalents (NCE), which are obtained by applying the following transformation to the percentile ranks q: $N_t = 21.06 \times \Phi^{-1}(q/100) + 50$ for test t, where Φ^{-1} is the standard normal inverse CDF. The two NCEs are averaged to obtain $\bar{N} = 0.5(N_{NNAT} + N_{OLSAT})$, then converted back into an integer-valued score R_i by rounding the result from $100 \times \Phi((\bar{N} - 50)/21.06)$.

The NNAT is a test of non-verbal reasoning and problem-solving skills. It includes items such as pattern completion and spatial visualization. The OLSAT is a test of verbal and non-verbal reasoning. For example, a sample OLSAT question from 2020 presents students with four different pictures of store windows and the following instructions are read aloud: “Mark under the picture that shows this: In a store window, there are two things to wear and one thing to play with." Versions of these tests are commonly used as components of gifted identification nationally. For example, the NNAT is used for gifted identification in the programs studied by Card et al. (2024) and Bui et al. (2014), and the OLSAT is used for gifted identification by the Los Angeles Unified School District.

A.4 Recommendation Admissions

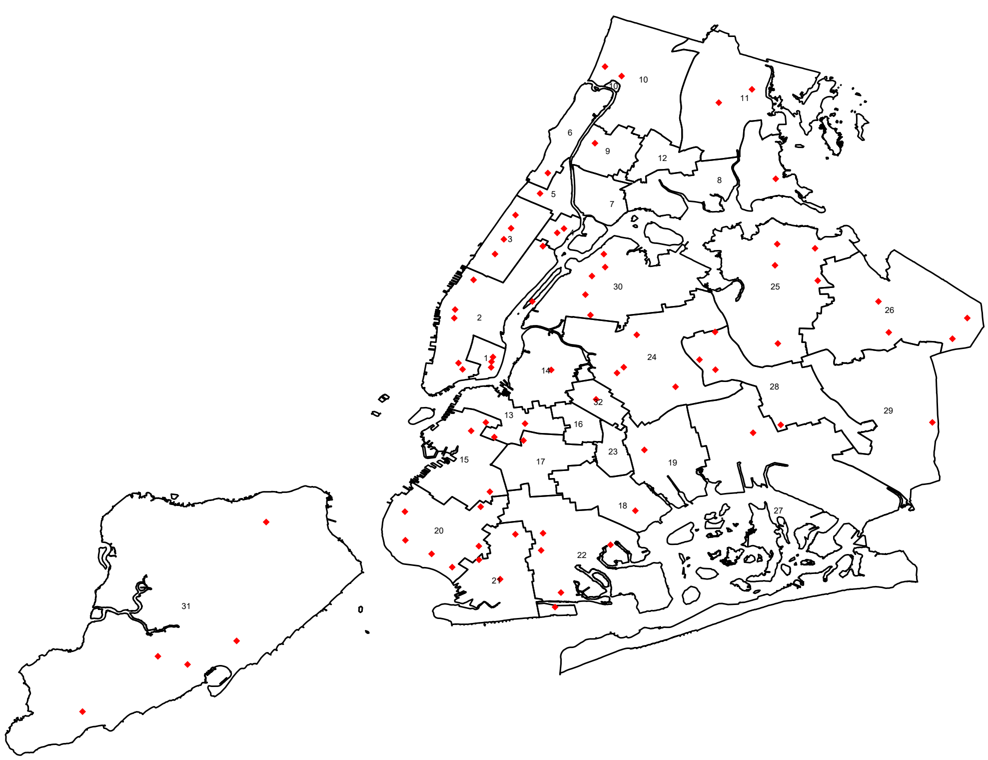
Under the recommendation system, kindergarten students qualify for G&T either by obtaining a recommendation from a NYC pre-K teacher or, for those not enrolled in NYC pre-K, obtaining a recommendation via interview with a staff member from the Division of Early Childhood Education. For admissions in later grades, eligibility is based on elementary course grades. Preschool teachers are only required to recommend students if they had any students apply to G&T programs, but there is no limit on the number of students that each teacher can recommend. When students are recommended, they are automatically eligible for both district and citywide programs.

Teachers or interviewers are asked to look for the following types of behaviors in assessing giftedness: curiosity and initiative (e.g., “Asks questions and communicates about the environment, people, events, and/or everyday experiences in and out of the classroom” or “Creatively expresses ideas verbally and non-verbally”), approaches to learning (e.g., “Willing to take risks and experiment” or “Becomes absorbed in topics, tasks, and activities"), and perceptiveness and self-direction (e.g., "Demonstrates pro-social problem solving skills" and "Follows through with plans and decisions.")

A.5 Admissions Priorities

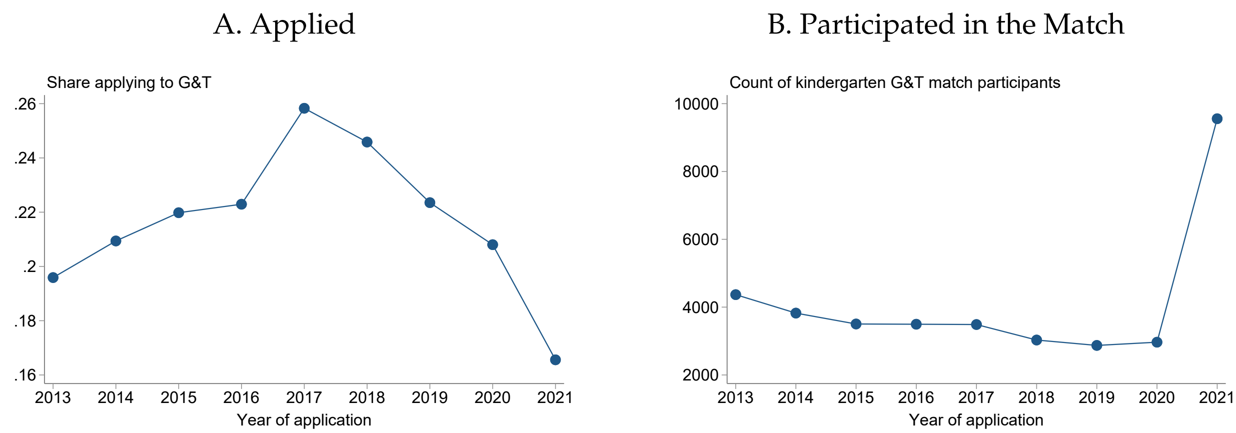
The following example illustrates how priorities are defined. Consider priorities at district programs that do not participate in DIA. Define $P_{is} = 100 \times \text{Sibling}_{is} + 10 \times \text{District}_{is} + R_i - 90$ if $R_i \geq 90$ and $-\infty$ otherwise, where Sibling_{is} indicates a sibling at s and District_{is} indicates being in-district for s. The priority ϱ_{is} is the rank of P_{is}, such that $\varrho_{is} = 1$ corresponds to applicants with the largest value of P_{is} (those with sibling and district priority and $R_i = 99$). Then, $\varrho_{is} = 2$ corresponds to applicants with sibling and district priority and $R_i = 98$, and so on. In this example, there are 40 values in the support of P_{is} (excluding $P_{is} = -\infty$) so $K = 40$. Those with $R_i < 90$ are ineligible, so we set $\varrho_{is} = \infty$. An analogous formula applies for district programs with DIA priorities and citywide programs.

Appendix Figure A1. Kindergarten G&T Locations



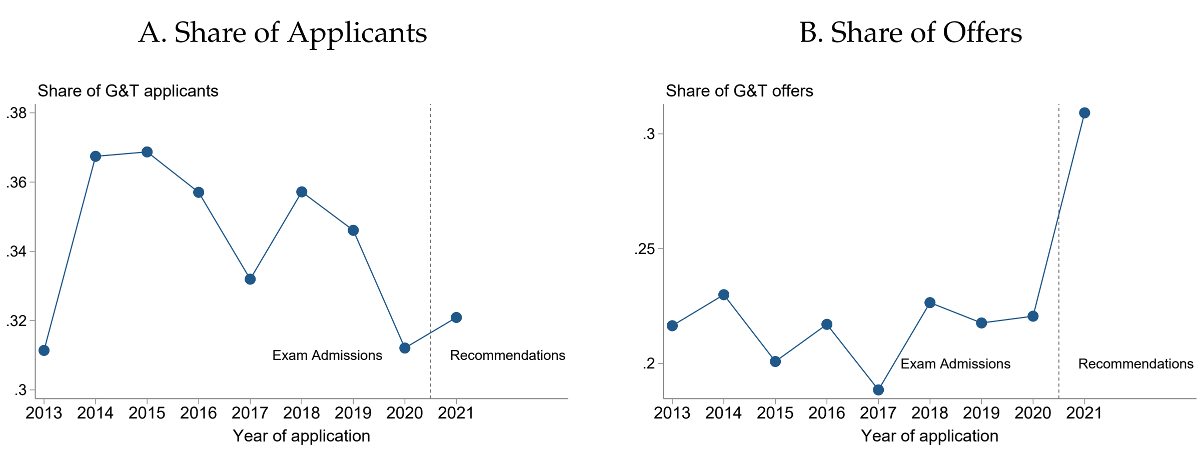
Notes: This figure depicts the locations of kindergarten G&T programs, including both district and citywide programs, during the 2020-2021 school year. The numbers indicate each of the 32 school districts comprising the NYC public school system.

Appendix Figure A2. Participation Trends in the G&T Application Process



Notes: Panel A shows the share of students who submitted an application for kindergarten G&T by year. The denominator is all students who either applied to kindergarten G&T or were enrolled in a traditional NYC public kindergarten. NYC students who did not apply to G&T and attended other schools are not included because they do not appear consistently in the data during the full time period. Panel B shows the number of gifted-eligible students who participated in the match. In 2020 and earlier, applicants who scored 90 or higher on an entrance exam were eligible for G&T. In 2021 and later, eligibility was determined by teacher recommendations and staff interviews.

Appendix Figure A3. Trends in FRPL G&T Applicants and Offers



Notes: Panel A shows the share of kindergarten G&T applicants who are FRPL-eligible, by year. Panel B shows the share of G&T offers given to FRPL-eligible applicants.

B Data

The data used in this paper are constructed using files provided to us by New York City Public Schools. Our data include information on applications, school and course enrollment, demographics, achievement, and school-to-student distances. Students are linked across different files and years. All files are de-identified.

Gifted and Talented Applications and Assignments

In our main analysis, we use G&T application data for cohorts applying for kindergarten G&T from school years 2011-2012 to 2020-2021. When plotting the trend in G&T application and enrollment shares, as in Figure 1, we extend the data through the 2021-22 school year. The application files record students' rank-order preference lists, priorities, and entrance exam scores (during the exam period). The application data omit lottery numbers. Therefore, we compute propensity scores by simulation following the procedure described in Section 3.2.

For cohorts starting kindergarten in 2015-16 onward, we have complete sibling priority data on the schools where an applicant has priority. For cohorts from 2011-12 through 2014-15, sibling priorities are missing, so we assume that no one has sibling priority.

Our files also omit data on priorities arising from the Diversity in Admissions (DIA) initiative which began in 2016. DIA allows G&T programs that opt-in to offer assignment priority to certain groups of applicants. Most programs do not offer DIA priorities but those that do can offer priorities on the basis of free- and reduced-price (FRPL) lunch status, English language learner status, temporary or public housing status, and residential neighborhood. For example, the Brooklyn School of Inquiry, a citywide G&T program, offers priority to applicants eligible for FRPL for 20% of its kindergarten seats and to applicants who live in Districts 18 or 19 for 20% of its kindergarten seats.

We impute DIA priorities wherever possible. When we observe an applicant's residential neighborhood or FRPL or English language learner status, we give them an additional priority at the relevant programs. We do not observe temporary or public housing status so do not impute those priorities.

Since sibling and DIA priorities are imputed (and sometimes missing completely), our simulated propensity scores deviate from the true probability $\Pr(Z_i = 1|\theta_i)$. However, the empirical strategy remains valid if the simulated propensity score eliminates omitted variables bias. Statistical tests for balance, reported in Appendix Table D1, support the interpretation that our propensity scores eliminate OVB.

During the post-reform teacher recommendations period, G&T application files include priorities for all students who applied for G&T programs, including an indicator for whether applicants received a recommendation.

Enrollment and Demographics

NYCPS collects student enrollment and demographic information in the June of each academic year. We use enrollment files through the 2023-24 school year. The files record the school and classroom in which students are enrolled. They also record indicators for student race, gender, FRPL status, English language learner status, and disability status.

In the exam admissions period, we define a class as G&T if at least 90% of its students were G&T-eligible. We have data on the classrooms officially identified as G&T beginning in the 2015-2016 school year and have validated that classes we classify as G&T align closely with the official coding. We use our proxy measure to maintain a consistent definition of G&T classrooms throughout our sample period.

In the recommendations admissions period beginning in 2021, we use the official classroom code since there are no G&T exams available, and defining classrooms based on 90% of the students receiving a G&T recommendation does not line up closely with the official classroom codes. For comparability, the time series analyses that compare G&T enrollment across years then all use the official G&T classroom codes, rather than the previous 90% threshold measure.

Student-to-School Distances

We measure student-to-school proximity by computing the great circle distance (in miles) between the school and the centroid of the student's residential census tract. This calculation is based on the latitude and longitude coordinates of both locations.

Achievement

We use exam scores from the New York State (NYS) Assessments for school years 2012-13 through 2023-2024 as our primary achievement outcome. The standardized exams are administered annually to students in grades 3-8. We use a student's first attempt as their exam score for each grade. The exam scores are standardized to be mean zero and unit variance by subject-grade-year in the population of all NYC students. Testing was suspended in the school years 2019-20 and 2020-21 because of the COVID-19 pandemic. Testing resumed in school year 2021-22.

The New York Regents Exams are standardized tests administered to students enrolled in high school-level courses. Students are required to pass a certain number of Regents exams to graduate high school. Students enrolled in Algebra I and Geometry courses during 7th or 8th grade are required to take the Regents exams and are exempt from the NYS Assessments for math. Since many middle school students opt out for this reason, we exclude 7th and 8th grade NYS math assessments as outcomes in our achievement results. We similarly standardize the Algebra Regents scores to be mean zero and unit variance at the grade-year level in the population of NYC students.

Middle School Applications and Assignments

We obtain data on middle school applications and assignments for school years 2016-2017 through 2024-2025. These data are generated by centralized assignment for middle school admissions and record students' rank-ordered preferences over programs, their priorities at those programs, and the program to which they are assigned. We link middle school assignments to enrollment files to track the schools to which students apply and enroll.

We focus primarily on two middle school outcomes. The first, which we call the specialized high school share, is the share of 6th grade students who later enroll in a specialized high school in 9th grade, averaged across students enrolled in 6th grade between 2016 and 2019. We begin in 2016 because this is the first year that students in our sample begin 6th grade. The second middle school outcome is the school's estimated value-added using data from the same years (see Appendix C.2 for details). We use data up to 2019 because state exams were canceled during the COVID-19 pandemic in 2020 and 2021.

Specialized High School Admissions Test

The Specialized High School Admissions Test (SHSAT) is the exam required for admission to NYC specialized high schools. We obtain these data from the high school match files, for cohorts entering high school in the 2019-20 through 2024-25 school years. We construct dummy variables for whether students took the SHSAT exam and whether they received an offer to any specialized high school. Students who do not take the SHSAT exam, but were enrolled in NYC schools in 8th grade, are coded as not receiving an offer to a specialized high school.

C Methodological Details

C.1 Attrition Specifications

Within the RD sample, we estimate the following model via 2SLS, separately by grade:

$$\begin{aligned} Z_i &= \pi 1(R_i \geq 90) + \lambda_1 R_i + \gamma_1 R_i 1(R_i \geq 90) + \nu_i \\ Y_i &= \beta Z_i + \lambda_2 R_i + \gamma_2 R_i 1(R_i \geq 90) + \varepsilon_i, \end{aligned}$$

where outcome Y_i is an indicator for enrollment in a non-charter NYC public school.

Within the lottery sample, we estimate the following reduced form separately by grade, where Z_i is G&T assignment:

$$Y_i = \rho Z_i + \delta p_i + X_i' \psi + \eta_i.$$

C.2 School Value-Added Estimation

One primary outcome is the value-added of the middle schools to which students apply and enroll. To obtain value-added estimates we estimate the following model via OLS:

$$Y_i = D_i' \alpha + X_i' \Lambda + \upsilon_i,$$

where Y_i is sixth grade math achievement, D_i are sixth grade school enrollment dummies, and X_i are student demographics and fifth grade achievement, interacted with year. Achievement scores are normalized to be mean zero and unit variance by grade-subject-year. The control vector includes indicators for race, gender, FRPL, English language learner, and disability status, as well as cubic functions of fifth grade math and ELA achievement. Angrist et al. (2024) show that value-added estimates $\hat{\alpha}$ obtained from models of this form are nearly unbiased for sixth grade math and modestly biased for sixth grade ELA in NYC middle schools. Accordingly, we opt for sixth grade math value-added as the primary outcome in our analysis (though results are similar for ELA).

In Appendix Table F7, we report results on high school value-added, estimated by a regression of total SAT scores on ninth grade enrollment dummies and eighth grade demographics and achievement. Angrist et al. (2024) show that estimates obtained in this manner predict causal effects but are modestly biased in NYC.

C.3 Stacked 2SLS

For achievement outcomes, we stack the data across grades to improve precision. That is, we index the estimating equations by grade and estimate them jointly via system 2SLS, imposing a common coefficient for the causal effect across grades. Wooldridge (2010) overviews system 2SLS, and Angrist et al. (2023) describes several applications.

Formally, we have the following grade-specific second stage equations:

$$\begin{aligned}
Y_{i1} &= D_{i1}\tau + X_{i1}'\Gamma_1 + \varepsilon_{i1}, \\
&\vdots \\
Y_{iG} &= D_{iG}\tau + X_{iG}'\Gamma_G + \varepsilon_{iG},
\end{aligned}$$

where $g \in \{1, ..., G\}$ indexes grade and X_{ig} is a $M \times 1$ vector. Note that D_{ig} are equal across g for the same i, as the treatment is the same across samples. Suppose each student is observed in each grade so that

$$\mathbf{Y}_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{iG} \end{bmatrix}, \quad \mathbf{D}_i = \begin{bmatrix} D_{i1} \\ D_{i2} \\ \vdots \\ D_{iG} \end{bmatrix}, \quad \mathbf{X}_i = \begin{bmatrix} X_{i1}' & 0 & \cdots & 0 \\ 0 & X_{i2}' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_{iG}' \end{bmatrix}, \quad \boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iG} \end{bmatrix}.$$

We also have a matrix of instruments given by

$$\mathbf{Z}_i = \begin{bmatrix} Z_{i1} & 0 & \cdots & 0 \\ 0 & Z_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & Z_{iG} \end{bmatrix}.$$

Again, note that Z_{ig} are equal across grades for the same student since the instrument is the same across samples. Stacking across students yields the following system

$$\begin{aligned}
\mathbf{Y} &= \mathbf{D}\tau + \mathbf{X}\Gamma + \boldsymbol{\varepsilon} \\
\mathbf{D} &= \mathbf{Z}\Pi + \mathbf{X}\Psi + \boldsymbol{\nu},
\end{aligned}$$

where $\mathbf{Y}$, $\mathbf{D}$, $\boldsymbol{\varepsilon}$, and $\boldsymbol{\nu}$ are $NG \times 1$ vectors, Π is a $G \times 1$ vector, $\mathbf{X}$ is a $NG \times MG$ matrix, and $\mathbf{Z}$ is a $NG \times G$ matrix. The control coefficients Γ and Ψ form $MG \times 1$ vectors, but we restrict the causal effect τ to be common across grades.

We estimate this system by 2SLS, clustering standard errors by student. In practice, note that G will vary by student as not all students are observed in every grade. The procedure when stacking across grades in the RD design, or when stacking across research designs, is analogous.

C.4 Multisector 2SLS

Instrumenting either citywide or district G&T enrollment alone yields an effect relative to a composite counterfactual. District lottery losers enroll in a non-G&T or citywide program and citywide losers enroll in a non-G&T or district program. The counterfactual is harder to interpret and motivates a multisector model that jointly estimates district and citywide effects relative to non-G&T enrollment. Let s denote one of three school sectors:

non-G&T, district G&T, and citywide G&T, numbering them 0, 1, and 2. The model is

$$D_{is} = \varphi_s Z_{is} + \omega_{1s} p_{is} + X_i' \kappa_{1s} + \zeta_{is}, \quad s \in \{1, 2\}$$
$$Y_i = \tau_1 D_{i1} + \tau_2 D_{i2} + \omega_{21} p_{i1} + \omega_{22} p_{i2} + X_i' \kappa_2 + \eta_i,$$

where non-G&T students, $D_{i0} = 1$, form the omitted group. We estimate this model in the sample of students who have non-degenerate assignment risk at either G&T sector ($p_{i1} \in (0,1)$ or $p_{i2} \in (0,1)$). Prior work discusses the conditions under which the multisector model yields interpretable treatment effects (Behaghel et al., 2013; Bhuller and Sigstad, 2024; Heinesen et al., 2022; Kirkeboen et al., 2016). Under restrictions on effect heterogeneity and/or behavioral responses to the instrument, τ_1 and τ_2 are average treatment effects on different types of compliers shifted into district and citywide G&T.

First stage regressions are informative about the extent to which the 2SLS might deviate from a positively-weighted average of treatment effects on compliers. Bhuller and Sigstad (2024) and Heinesen et al. (2022) discuss how this sort of bias arises when there are defiers, and the difference in treatment effects for defiers and compliers are large. With three treatment values, Heinesen et al. (2022) describe "irrelevance" defiers as units for which $Z_{is} = 1 \Rightarrow D_{is'} = 1$ and $Z_{is} = 0 \Rightarrow D_{is} = D_{is'} = 0$, and "next-best" defiers as those where $Z_{is} = 1 \Rightarrow D_{is} = 1$ and $Z_{is} = 0 \Rightarrow D_{is'} = 1$. Furthermore, their Proposition 4 shows that defier shares are partially identified with bounds given by functions of first stage parameters. By applying their result with the first stage estimates from Appendix Table D3, we obtain an upper bound of 0.00 for the share of irrelevance and next-best defiers for district G&T. We obtain an upper bound of 0.10 for next-best defier share and 0.02 for the irrelevance defier share for citywide G&T. (The first stage estimates are precisely estimated, so accounting for sampling error would only slightly increase the upper bounds.) Since defier shares are small, substantial bias only arises if effect heterogeneity is extremely large. We conclude that the magnitude of potential bias is likely small.

C.5 IV VAM

Let β_j denote the causal effect of program j. The IV VAM framework assumes that program-level causal effects can be written as $\beta_j = M_j' \varphi + \nu_j$, where M_j is a low-dimensional set of program-level mediators. In our application, we let M_j be scalar and equal to an estimate of the causal effect for j obtained from a conventional value-added model (see Appendix Section C.2).

The IV VAM procedure instruments M_j with program-specific offer instruments, controlling for program-specific assignment propensity scores. The primary concern is that the instruments violate the exclusion restriction via direct effects on ν_j. The key assumption, following Kolesár et al. (2015), requires that exclusion violations across instruments average out in a many-instruments asymptotic sequence (Assumption MIV in Angrist et al. (2024)). The procedure involves a bias-correction to account for many instruments bias, implemented with a suitable k-class estimator.

Our objective is to estimate the forecast coefficient φ, the standard deviation of true value-added σ_β, and the standard deviation of the forecast residual σ_ν. We compute the standard deviation of value-added as $\sigma^2_\beta = \varphi^2 Var(M_j) + \sigma^2_\nu$. The estimator for σ_ν can be negative, in which case we report a value of zero. Appendix Table F8 reports estimates of these objects for achievement and middle school enrollment outcomes. Note that the forecast coefficients for the test score outcomes are well below 1, suggesting that conventional VAMs are biased. However, the forecast coefficients are close to 1 for the middle school access outcomes. The over-identification passes for the value-added outcome and rejects at the 5% level for the specialized high school share outcome. On balance, VAMs for the middle school outcomes seem to provide better guides to program-level causal effects.

C.6 Inverse Probability Weighted 2SLS

We use a weighted 2SLS estimator to address the potential for selection bias arising from differential attrition. The weights reflect the inverse probability of having a non-missing outcome. Let H_i be a dummy variable that equals 1 if the outcome Y_i is observed and 0 otherwise. In what follows, Y_i, D_i, and Z_i are scalars. Denote potential outcomes as $Y_i(1)$ when treated and $Y_i(0)$ when untreated. Observed outcomes are given by $Y_i = D_iY_i(1) + (1 - D_i)Y_i(0)$ if $H_i = 1$ and missing otherwise. In addition to the standard IV assumptions, we require H_i to be conditionally ignorable.

Assumption C.1 (Outcome is missing at random)**.**

$$Y_i(d) \perp\!\!\!\perp H_i \mid D_i, X_i, Z_i, \qquad \textit{for } d \in \{0, 1\}.$$

The key assumption is that missingness is as-good-as-random given treatment, the instrument, and exogenous controls. Under this assumption, weighted 2SLS applied to the selected sample with $H_i = 1$, with weights equal to the inverse probability that $H_i = 1$, yields the same estimand as that of the unweighted 2SLS estimator applied to the unselected sample. To see this, note the moment condition that defines weighted 2SLS is

$$\mathbb{E}[w_i\tilde{Z}_iH_i(Y_i - \tilde{D}_i'\beta)] = 0,$$

where $w_i = 1/\Pr(H_i = 1|D_i, X_i, Z_i)$ and $\tilde{D}_i = D_i - X_i\Gamma_X$ and $\tilde{Z}_i = Z_i - X_i\Gamma_Z$ are residuals from auxiliary regressions of D_i and Z_i on the controls X_i. By iterating expectations, the above equals

$$\begin{aligned}
&\mathbb{E}[w_i\tilde{Z}_i(\mathbb{E}[Y_iH_i|D_i, X_i, Z_i] - \tilde{D}_i'\beta\mathbb{E}[H_i|D_i, X_i, Z_i])] \\
&= \mathbb{E}[w_i\tilde{Z}_i\mathbb{E}[Y_iH_i|D_i, X_i, Z_i] - \tilde{Z}_i\tilde{D}_i'\beta] \\
&= \mathbb{E}[w_i\tilde{Z}_i\mathbb{E}[Y_i(D_i)H_i|D_i, X_i, Z_i] - \tilde{Z}_i\tilde{D}_i'\beta].
\end{aligned}$$

Since $Y(D) = 1\{D = 1\}Y(1) + 1\{D = 0\}Y(0)$, we obtain

$$\begin{aligned}
&\mathbb{E}[w_i\tilde{Z}_i\{1\{D_i = 1\}\mathbb{E}[Y_i(1)H_i|D_i, X_i, Z_i] + 1\{D_i = 0\}\mathbb{E}[Y_i(0)H_i|D_i, X_i, Z_i]\} - \tilde{Z}_i\tilde{D}_i'\beta] \\
&= \mathbb{E}[w_i\tilde{Z}_i\{1\{D_i = 1\}\mathbb{E}[Y_i(1)|D_i, X_i, Z_i]w_i^{-1} + 1\{D_i = 0\}\mathbb{E}[Y_i(0)|D_i, X_i, Z_i]w_i^{-1}\} - \tilde{Z}_i\tilde{D}_i'\beta] \\
&= \mathbb{E}[\tilde{Z}_i\{1\{D_i = 1\}\mathbb{E}[Y_i(1)|D_i, X_i, Z_i] + 1\{D_i = 0\}\mathbb{E}[Y_i(0)|D_i, X_i, Z_i]\} - \tilde{Z}_i\tilde{D}_i'\beta],
\end{aligned}$$

where the second equality follows by Assumption C.1. Finally, simplify to obtain

$$
\begin{aligned}
&\mathbb{E}[\tilde{Z}_i\{\mathbb{E}[1\{D_i=1\}Y_i(1)+1\{D_i=0\}Y_i(0)|D_i,X_i,Z_i]\}-\tilde{Z}_i\tilde{D}_i'\beta]\\
&=\mathbb{E}[\tilde{Z}_i\{\mathbb{E}[Y_i(D_i)|D_i,X_i,Z_i]\}-\tilde{Z}_i\tilde{D}_i'\beta]\\
&=\mathbb{E}[\mathbb{E}[\tilde{Z}_i(Y_i-\tilde{D}_i'\beta)|D_i,X_i,Z_i]]\\
&=\mathbb{E}[\tilde{Z}_i(Y_i-\tilde{D}_i'\beta)].
\end{aligned}
$$

The final line shows that we obtain the moment condition defining 2SLS in the unselected population. We include demographics and assignment propensity scores as exogenous controls and estimate the propensity score for H_i by logit. The estimated propensity scores are bounded away from zero, so we do not encounter issues caused by extreme weights. Standard errors are computed via bootstrap.

C.7 ATT Identification

Our reweighting exercise aims to estimate what the ATT would be under different treated populations. We require an additional identification assumption to impute treatment effect values for the always-takers in the treated population. First, note that the 2SLS estimand with assignment instruments and propensity score controls is a weighted average of treatment effects for applicants sharing the same value of the propensity score. The 2SLS estimand while conditioning on $X_i = x$ is:

$$
\tau_x=\sum_p \frac{\Pr(p_i=p \mid X_i=x)\pi_{px}p(1-p)}{\sum_{p'}\Pr(p_i=p' \mid X_i=x)\pi_{p'x}p'(1-p')}\tau_{px},
$$

where $\pi_{px}=\Pr[D_i(1)>D_i(0)|p_i,X_i]$ and $\tau_{px}=\mathbb{E}[Y_i(1)-Y_i(0)|D_i(1)>D_i(0),p_i,X_i]$ are the first stage and LATE for strata defined by $p_i=p$ and $X_i=x$. The following assumption enables extrapolation of LATEs to the full treated population:

Assumption C.2 (Conditional effect ignorability)**.**

$$
\mathbb{E}[Y_i(1)-Y_i(0) \mid p_i, X_i, D_i(1), D_i(0)]=\mathbb{E}[Y_i(1)-Y_i(0) \mid X_i].
$$

This condition, a variant of Assumption 3 in Angrist and Fernández-Val (2013), is satisfied when X_i is the only source of heterogeneity in average treatment effects. It rules out heterogeneity by compliance status or by propensity score, such that $\tau_{px}=\tau_x$. Under Assumption C.2 and the usual independence, relevance, and monotonicity assumptions:

Proposition C.1 (LATE-Reweight)**.**

$$
\mathbb{E}[Y_i(1)-Y_i(0) \mid D_i=1]=\int \tau_x \omega dF_X(x),
$$

where $\omega=\Pr[D_i=1 \mid X_i=x]/\Pr[D_i=1]$.

Proof. Iterate expectations and write D_i as a function of potential values and the instrument Z_i to obtain

$$\begin{aligned}\mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 1] &= \mathbb{E}[\mathbb{E}\{Y_i(1) - Y_i(0) \mid D_i = 1, p_i, X_i\} \mid D_i = 1] \\ &= \mathbb{E}[\mathbb{E}\{Y_i(1) - Y_i(0) \mid (1 - Z_i)D_i(0) + Z_i D_i(1) = 1, p_i, X_i\} \mid D_i = 1].\end{aligned}$$

By conditional effect ignorability and conditional ignorability of the instrument, the above equals

$$\mathbb{E}[\mathbb{E}\{Y_i(1) - Y_i(0) \mid X_i\} | D_i = 1] = \mathbb{E}[\tau_x \mid D_i = 1].$$

Then by Bayes' theorem,

$$\begin{aligned}\mathbb{E}[\tau_x \mid D_i = 1] &= \int \tau_x dF_X(x \mid D_i = 1) \\ &= \int \tau_x \frac{\Pr(D_i = 1 \mid X_i = x)}{\Pr(D_i = 1)} dF_X(x).\end{aligned}$$

□

The proposition and proof, again following Angrist and Fernández-Val (2013), demonstrate that a weighted average of LATEs identifies the ATT under an appropriate restriction on effect heterogeneity.

C.8 Student vs Program-Level Heterogeneity

By the arguments in Appendix C.7, we identify conditional average treatment effects (CATE):

$$\tau_x = \mathbb{E}[Y(1) - Y(0) \mid X = x].$$

Our objective is to reweight $\widehat{\tau}_x$ to obtain estimates of $\mathbb{E}[Y(1) - Y(0) \mid D = 1]$. One concern is that differences in CATEs might reflect program-level heterogeneity rather than student-level characteristics per se. For example, effect heterogeneity such that $\tau_x - \tau_{x'} \neq 0$ could reflect the fact that students with $X = x$ and $X = x'$ enroll in different types of programs. Forcing such students to enroll in the same set of programs could generate $\tau_x = \tau_{x'}$. That would mean all of the heterogeneity is being explained by program-level heterogeneity that is correlated with student characteristics.

We can interpret differences in the CATEs as arising from student-level characteristics if the degree of selection-on-gains into different programs is limited. Let S_i denote the program that student i enrolls in. Suppose we contrast τ_x and $\tau_{x'}$. An Oaxaca-Blinder decomposition yields

$$
\begin{aligned}
&\tau_x - \tau_{x'} \\
&= \sum_s (E[Y_1 - Y_0 \mid X = x, S = s] - E[Y_1 - Y_0 \mid X = x', S = s]) \Pr(S = s \mid X = x') \\
&\quad + \sum_s E[Y_1 - Y_0 \mid X = x, S = s](\Pr(S = s \mid X = x) - \Pr(S = s \mid X = x')).
\end{aligned}
$$

The first term is a convex-weighted average of within-program CATE differences due to differences in X. The second term reflects differences in the enrollment mix for students with different values of X. A small value of the second term supports the interpretation that CATE differences are driven by student-level characteristics rather than program-level heterogeneity.

Note the second term can be written as a covariance. Multiply and divide it by some probability distribution over S, which we will call $\pi(S)$.

$$
\begin{aligned}
&\sum_s E\{Y_1 - Y_0 \mid X = x, S = s\}\Big(\Pr(S = s \mid X = x) - \Pr(S = s \mid X = x')\Big) \qquad (1) \\
&= \sum_s E\{Y_1 - Y_0 \mid X = x, S = s\} \left(\frac{\Pr(S = s \mid X = x) - \Pr(S = s \mid X = x')}{\pi(S)} \right) \pi(S) \\
&= E_{S\sim\pi}\left[E\{Y_1 - Y_0 \mid X = x, S\} \underbrace{\frac{\Pr(S \mid X = x) - \Pr(S \mid X = x')}{\pi(S)}}_{\equiv W(S)} \right].
\end{aligned}
$$

Note that $W(S)$ is mean zero:

$$
\begin{aligned}
E_{S\sim\pi}[W(S)] &= \sum_s \pi(S) \frac{\Pr(S = s \mid X = x) - \Pr(S = s \mid X = x')}{\pi(S)} \\
&= \sum_s \Big(\Pr(S = s \mid X = x) - \Pr(S = s \mid X = x')\Big) \\
&= 0.
\end{aligned}
$$

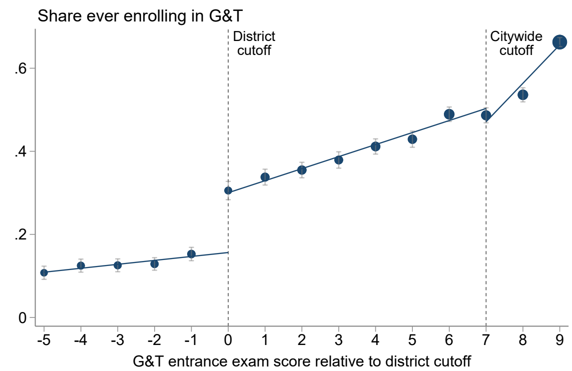
Therefore, (1) is a covariance:

$$
\begin{aligned}
&\mathrm{Cov}\Big(E\{Y_1 - Y_0 \mid X = x, S\}, W(S)\Big) \\
&= \mathrm{Cov}\left(E\{Y_1 - Y_0 \mid X = x, S\}, \frac{\Pr(S \mid X = x) - \Pr(S \mid X = x')}{\pi(S)} \right).
\end{aligned}
$$

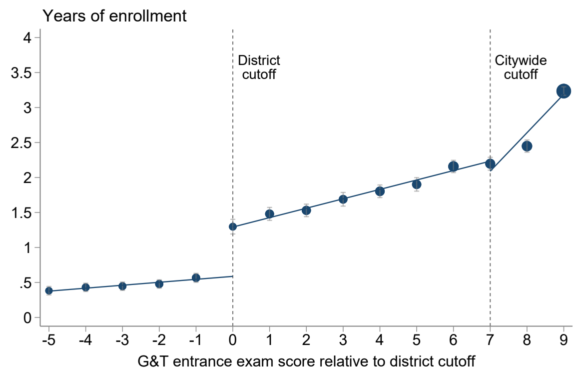
This term is small as long as treatment effects are uncorrelated with the difference in propensities given by $W(S)$. Selection-on-gains implies that this covariance is positive, but the term is small if there isn't much Roy sorting. Intuitively, if programs matter but students randomly sort into them, then heterogeneity should not distort the contrast between average treatment effects.

Appendix Figure C1. Alternative First Stage Effects of G&T Eligibility

A. Ever Enrolled



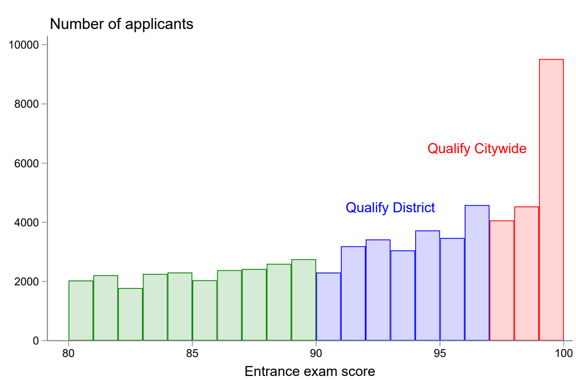
B. Years Enrolled



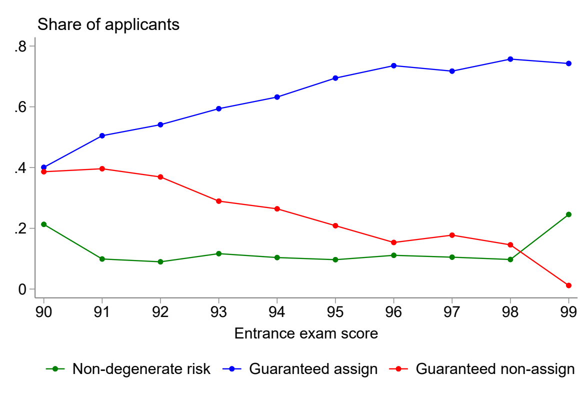
Notes: This figure depicts the effect of kindergarten G&T eligibility on alternative first-stage outcomes. Panel A depicts the effect on ever enrolling in a G&T program between grades K-5. Panel B depicts the effect on the number of years enrolling in a G&T program between grades K-5. The sample only includes cohorts through 2018-19 for whom we can observe G&T classroom enrollment for five full years after kindergarten. Marker sizes are proportional to the number of students attaining each score value. Whiskers depict 95-percent confidence intervals for the conditional means.

Appendix Figure C2. Distribution of Entrance Exam Scores and Assignment Probabilities

A. Entrance Exam Scores



B. Assignment Probabilities



Notes: Panel A depicts the frequency of G&T entrance exam scores for kindergarten applicants from 2011-2020. Scores in green (less than 90) do not qualify the student for any G&T program, scores in blue (90-96) qualify the student for district G&T programs only, and scores in red (97 and above) qualify the student for citywide G&T programs. Panel B depicts the distribution of assignment risk status by entrance exam score. The blue line shows the share of applicants who are guaranteed assignment at some G&T program ($p_i = 1$). The red line shows the share of applicants who are guaranteed non-assignment at all G&T programs ($p_i = 0$). The resulting green line is the share of applicants with assignment risk to the G&T sector $p_i \in (0, 1)$.

D Research Design Checks

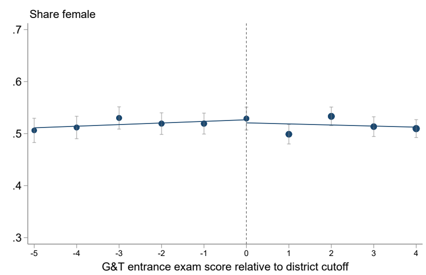
Appendix Table D1. Covariate Balance

	RD sample		Lottery sample		
	Full sample (1)	6th grade (2)	Uncontrolled (3)	Controlled (4)	Controlled, 6th grade (5)
Female	–0.007	0.007	–0.005	–0.014	–0.032
	(0.014)	(0.021)	(0.004)	(0.018)	(0.024)
Black	0.008	0.012	–0.106***	0.008	0.006
	(0.008)	(0.012)	(0.002)	(0.007)	(0.010)
White	0.010	0.001	0.085***	0.007	0.002
	(0.013)	(0.020)	(0.003)	(0.018)	(0.024)
Hispanic	0.007	0.015	–0.112***	0.006	0.015
	(0.010)	(0.015)	(0.002)	(0.009)	(0.013)
Asian	–0.034***	–0.036*	0.119***	–0.036**	–0.032
	(0.013)	(0.020)	(0.003)	(0.016)	(0.022)
FRPL	–0.017	–0.007	–0.148***	–0.002	0.009
	(0.012)	(0.018)	(0.003)	(0.013)	(0.018)
Disability	–0.003	0.003	–0.024***	–0.003	0.002
	(0.006)	(0.009)	(0.001)	(0.007)	(0.009)
ELL	–0.003	–0.007	0.006***	0.008	0.004
	(0.003)	(0.006)	(0.001)	(0.005)	(0.007)
N	24,196	10,567	142,025	4,322	2,127
Joint test					
$\chi^2(8)$	14.5	10.3	7392	11.4	6.09
p-value	[0.070]	[0.247]	[0.000]	[0.178]	[0.637]

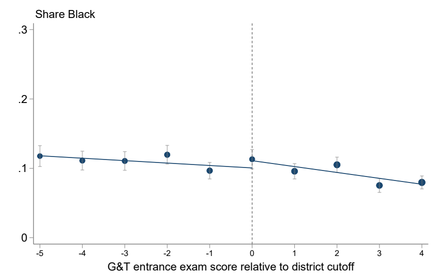
Notes: This table reports statistical tests for balance. Column 1 reports estimates from regressions of baseline demographic variables on the RD district eligibility instrument with RD controls. The RD model imposes linear fits for the conditional means and a bandwidth of length five. Column 2 reports estimates from the same model as column 1 in the sample of students who enroll in a NYC public school in sixth grade. Column 3 reports estimates from regressions of demographics on an indicator for G&T assignment and application year dummies in the sample of all G&T applicants. Column 4 reports estimates that additionally control for assignment risk in the sample of students with non-degenerate assignment risk. Column 5 reports estimates from the same model as column 4 but in the sample of students who have non-degenerate risk and enroll in a NYC public school in sixth grade. FRPL indicates free- or reduced-price lunch status and ELL indicates English language learner status. Robust standard errors in parentheses. The last two rows report chi-squared statistics and p-values for Wald tests of the null hypothesis that the coefficients in each column are all equal to zero. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Figure D1. Baseline Covariates are Smooth Through the G&T Cutoff

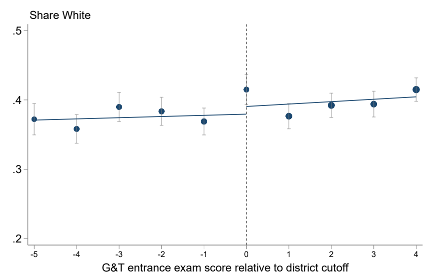
A. Female



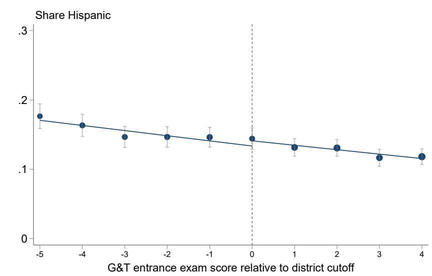
B. Black



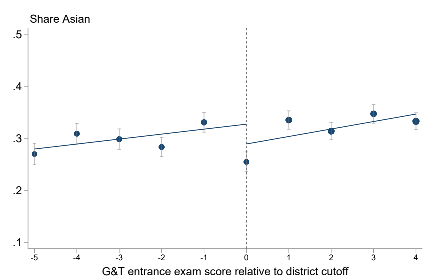
C. White



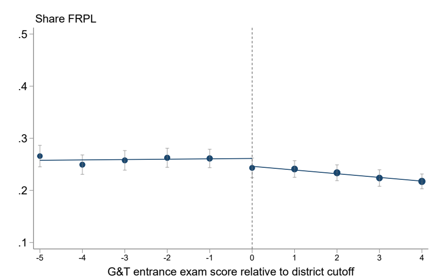
D. Hispanic



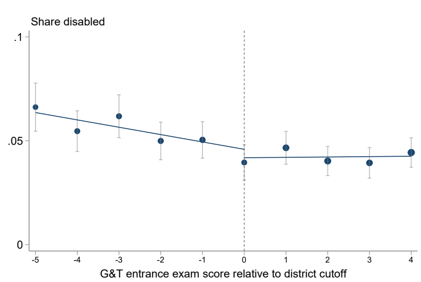
E. Asian



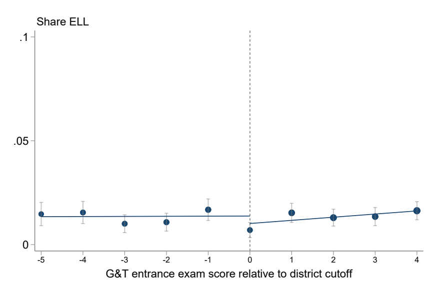
F. FRPL



G. Disability



H. English Language Learner



Notes: This figure depicts tests for balance using the RD eligibility instrument. Each panel plots means of the outcome at each score relative to the district eligibility cutoff. Outcomes are indicators for each characteristic. Marker sizes are proportional to the number of students attaining each score value. Whiskers depict 95-percent confidence intervals for the conditional means.

Appendix Table D2. Complier Characteristics

	RD Sample	Lottery Sample	Lottery, Entrance Exam Score 90-94	Lottery, Entrance Exam Score 95-99
	(1)	(2)	(3)	(4)
Female	0.542	0.500	0.562	0.497
	(0.026)	(0.019)	(0.101)	(0.018)
Black	0.139	0.029	–0.014	0.034
	(0.015)	(0.007)	(0.054)	(0.006)
White	0.271	0.503	0.508	0.503
	(0.024)	(0.019)	(0.101)	(0.018)
Hispanic	0.163	0.081	0.213	0.068
	(0.017)	(0.009)	(0.064)	(0.008)
Asian	0.402	0.258	0.429	0.244
	(0.023)	(0.017)	(0.096)	(0.015)
Other/Missing Race	0.024	0.128	–0.136	0.152
	(0.020)	(0.015)	(0.079)	(0.014)
FRPL	0.354	0.131	0.412	0.106
	(0.023)	(0.013)	(0.091)	(0.011)
N	37292	5004	1395	3609

Notes: This table reports estimates of complier covariate means. Sample sizes show the total number of observations used for estimation. Columns 1-2 report estimates for compliers in the RD and lottery samples. Column 3 reports estimates for compliers in the intersection of the lottery and RD samples. Column 4 reports estimates for compliers in the lottery sample but not in the RD sample. Estimates are obtained by stacking the data to jointly estimate treated and untreated complier means, imposing a common coefficient. We instrument indicators for treatment status with indicators for instrument status, with interactions of treatment status and the baseline covariate as the outcome. Standard errors clustered by student are in parentheses.

Appendix Table D3. Multisector First Stages

	District enrollment (1)	Citywide enrollment (2)
District offer	0.460***	–0.022**
	(0.016)	(0.011)
Citywide offer	–0.091***	0.717***
	(0.012)	(0.013)
Constant	0.104***	0.000
	(0.010)	(0.008)
N	6,927	6,927

Notes: This table reports first stage regressions on district and citywide offer instruments. The dependent variable in column 1 is a dummy variable that equals one if enrolling in district G&T. The dependent variable in column 2 is a dummy variable that equals one if enrolling in citywide G&T. The regressors are the two district and citywide offers, propensity score controls for each instrument, and a constant. The sample is restricted to applicants with a non-degenerate probability of assignment at either the district or citywide sector. Robust standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

E Robustness

Appendix Table E1. Lottery Estimates with Inverse Probability Weighting

	Full sample (1)	FRPL (2)	Non-FRPL (3)
Math	–0.047	0.020	–0.055
	(0.047)	(0.129)	(0.049)
First stage F	198	28.4	201
Over-id p-value	0.623	0.233	0.102
Non-G&T Mean	1.30	1.17	1.33
N (estimation)	8,093	1,449	6,644
N (students)	3,264	552	2,712
ELA	–0.045	–0.100	–0.029
	(0.044)	(0.126)	(0.048)
First stage F	160	20.0	148
Over-id p-value	0.419	0.281	0.103
Non-G&T Mean	1.21	1.09	1.23
N (estimation)	9,702	1,762	7,940
N (students)	3,278	555	2,723
Specialized HS Share, Enrolled Middle School	0.090***	0.124***	0.082***
	(0.015)	(0.036)	(0.015)
First stage F	669	113	596
Non-G&T Mean	0.202	0.163	0.211
N (estimation)	2,100	390	1,710
Value-Added, Enrolled Middle School	0.020	0.076**	0.013
	(0.014)	(0.035)	(0.014)
First stage F	669	113	596
Non-G&T Mean	0.080	0.055	0.085
N (estimation)	2,100	390	1,710

Notes: This table reports estimates of the effects of G&T using weighted 2SLS, with weights equal to the inverse probability of having a non-missing outcome. Achievement results are obtained using samples stacked across grades 3-8 for ELA and 3-6 for math. See Tables 4 and 5 for further description of the outcomes. See Appendix C.6 for details of the estimation procedure. Standard errors computed via bootstrap are reported in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table E2. RD Robustness

			Bandwidth Length							
	Baseline	Quadratic fit	7	6	4	3	No controls	Honest RD	K-3 Pooled	K-3 Over-id
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Math	–0.071	–0.178**	0.023	–0.029	–0.065	–0.130	–0.143**	–0.379***	–0.109	–0.094*
	(0.063)	(0.085)	(0.053)	(0.057)	(0.071)	(0.085)	(0.069)	(0.088)	(0.071)	(0.052)
N	41,126	59,261	59,261	49,572	32,609	24,593	41,126	20,259	83,730	83,730
ELA	0.040	0.029	0.081	0.086	0.053	0.020	0.032	–0.181**	–0.041	–0.034
	(0.059)	(0.079)	(0.049)	(0.053)	(0.067)	(0.079)	(0.062)	(0.074)	(0.069)	(0.048)
N	50,104	72,399	72,399	60,390	39,768	30,041	50,104	25,649	106,174	106,174
Specialized HS Share, 1st Choice	–0.002	–0.014	0.011	–0.002	–0.009	0.009	–0.014	–0.017	–0.005	0.005
	(0.019)	(0.026)	(0.016)	(0.017)	(0.022)	(0.026)	(0.020)	(0.037)	(0.020)	(0.018)
N	13,101	18,982	18,982	15,803	10,377	7,844	13,101	6,682	25,064	25,064
Value-Added, 1st Choice	–0.001	–0.008	0.010	–0.000	–0.001	0.008	–0.012	–0.004	0.002	0.005
	(0.019)	(0.025)	(0.016)	(0.017)	(0.021)	(0.025)	(0.019)	(0.035)	(0.019)	(0.017)
N	13,101	18,982	18,982	15,803	10,377	7,844	13,101	6,682	25,064	25,064
Specialized HS Share, Enrolled	–0.026	–0.036	–0.007	–0.020	–0.031	–0.011	–0.036*	–0.025	–0.033*	–0.022
	(0.018)	(0.024)	(0.015)	(0.016)	(0.020)	(0.024)	(0.019)	(0.036)	(0.019)	(0.016)
N	10,400	15,004	15,004	12,505	8,262	6,250	10,400	5,160	20,966	20,966
Value-Added, Enrolled	–0.003	–0.001	–0.001	–0.004	–0.014	0.014	–0.009	0.054	0.006	0.000
	(0.020)	(0.027)	(0.017)	(0.018)	(0.023)	(0.027)	(0.020)	(0.039)	(0.021)	(0.019)
N	10,400	15,004	15,004	12,505	8,262	6,250	10,400	5,160	20,966	20,966

Notes: This table reports RD estimates obtained from several alternative specifications. Column 1 reproduces the baseline results. Column 2 adds quadratic running variable controls and increases the bandwidth to be length seven. Columns 3-6 vary the length of bandwidth with a linear fit. Column 7 removes demographic control variables. Column 8 reports results using the RDHonest procedure (Kolesár and Rothe, 2018). We implement the procedure with a triangular kernel, a bandwidth optimized for the fixed length confidence interval, the Armstrong and Kolesár (2020) plug-in smoothness parameter for the reduced form, and by enforcing linearity in the first-stage. Column 9 reports a pooled specification using all applicants to grades K-3. Column 10 allows the first stage to vary by application grade, generating four eligibility instruments in an over-identified fuzzy RD setup. Robust standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table E3. Effects on Additional Middle School Enrollment Outcomes

	RD			Lottery		
	Overall	FRPL	Non-FRPL	Overall	FRPL	Non-FRPL
	(1)	(2)	(3)	(4)	(5)	(6)
Mean Test Score Levels, 6th Grade	–0.080	–0.188*	–0.042	0.351***	0.401***	0.329***
	(0.061)	(0.109)	(0.075)	(0.042)	(0.108)	(0.044)
N	10,400	2,750	7,642	2,099	389	1,710
First stage F	798.7	264.1	517.3	779.7	119.5	664.1
Non-G&T Mean	0.561	0.418	0.611	0.750	0.585	0.788
p-value		[0.256]			[0.543]	
Specialized HS Share, 8th Grade	–0.037*	–0.072**	–0.020	0.082***	0.144***	0.072***
	(0.021)	(0.033)	(0.027)	(0.013)	(0.035)	(0.015)
N	6,842	1,904	4,937	1,367	264	1,103
First stage F	582.1	214.8	361.8	824.2	100.1	712.8
Non-G&T Mean	0.147	0.121	0.156	0.199	0.161	0.208
p-value		[0.224]			[0.061]	

Notes: This table reports 2SLS estimates of the effect of G&T enrollment on additional characteristics of the middle schools where students enroll. Columns 1-3 report RD estimates, and columns 4-6 report lottery estimates. The first panel reports estimates on the average math and ELA score of the school in which a student enrolled in sixth grade. The scores are normalized to be mean zero and unit variance by test-grade-year. The second panel reports estimates on the specialized high school share outcome, computed as in Table 5, but measured based on where a student is enrolled in 8th grade instead of 6th grade. The table reports Kleibergen-Paap F-statistics and outcome means for students who do not enroll in G&T. Robust standard errors are in parentheses. p-values for tests of the difference in G&T effects between FRPL and non-FRPL students are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

F Additional Results

Appendix Table F1. OLS Effects

	Uncontrolled (1)	Controlled (2)
Math	0.634***	0.161***
	(0.006)	(0.007)
N (estimation)	265,214	265,214
N (students)	97,169	97,169
ELA	0.562***	0.043***
	(0.006)	(0.007)
N (estimation)	278,710	278,710
N (students)	97,426	97,426
Specialized HS Share, Enrolled Middle School	0.110***	0.023***
	(0.002)	(0.002)
N	57,457	57,457
Value-Added, Enrolled Middle School	0.055***	0.009***
	(0.002)	(0.002)
N	57,457	57,457

Notes: This table reports OLS estimates of the effect of G&T enrollment. The first column reports estimates from a regression of the outcome on a dummy for G&T enrollment and year dummies. The second column additionally controls for demographics, demographic-year interactions, indicators for values of the G&T entrance exam, and a dummy for having a missing value for the G&T exam. See Tables 4, 5, and Appendix Table F3 for a description of the outcomes. Estimates for math and ELA are stacked across grades 3-6 for math and grades 3-8 for ELA, with standard errors in parentheses clustered by student. Robust standard errors are in parentheses for the remaining outcomes. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table F2. Effects on Achievement by Grade

	RD		Lottery	
	Math (1)	ELA (2)	Math (3)	ELA (4)
Grade 3	–0.014	0.039	–0.036	0.040
	(0.075)	(0.076)	(0.052)	(0.052)
	[1.034]	[1.008]	[1.258]	[1.218]
Grade 4	–0.161**	–0.027	–0.080	–0.026
	(0.078)	(0.081)	(0.053)	(0.058)
	[1.082]	[1.034]	[1.322]	[1.234]
Grade 5	–0.113	0.030	0.029	0.010
	(0.092)	(0.095)	(0.068)	(0.070)
	[1.055]	[1.012]	[1.305]	[1.214]
Grade 6	0.031	0.114	–0.073	–0.018
	(0.113)	(0.106)	(0.097)	(0.099)
	[1.138]	[1.024]	[1.331]	[1.187]
Grade 7		0.174		–0.117
		(0.117)		(0.092)
		[1.040]		[1.200]
Grade 8		0.031		–0.139*
		(0.123)		(0.073)
		[0.957]		[1.102]

Notes: This table reports 2SLS estimates of the effects of G&T enrollment on test scores by grade. See Table 4 for a description of the outcome variables. Robust standard errors are in parentheses. The mean of each outcome in each grade, among those who do not enroll in G&T, is shown in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table F3. Effects on Specialized High School Outcomes

	RD (1)	Lottery (2)
SHSAT Tested	–0.042	0.016
	(0.062)	(0.033)
N	7,825	1,490
First stage F	614.8	839.1
Non-G&T Mean	0.731	0.802
Specialized HS Offer	0.010	–0.015
	(0.064)	(0.044)
N	7,825	1,490
First stage F	614.8	839.1
Non-G&T Mean	0.302	0.453

Notes: This table reports 2SLS estimates of the effect of G&T enrollment on high school application and offer outcomes. Column 1 reports fuzzy RD estimates and column 2 reports estimates using the lottery design. The first panel reports estimates for effects on an indicator for taking the Specialized High School Admissions Test. The second panel reports estimates of effects on an indicator for receiving an offer to a specialized high school. The table reports Kleibergen-Paap F-statistics and outcome means for students who do not enroll in G&T. Robust standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table F4. Effects on Kindergarten Classroom Characteristics by FRPL Status

	RD		Lottery	
	FRPL	Non-FRPL	FRPL	Non-FRPL
	(1)	(2)	(3)	(4)
Class Size	–5.972***	–4.002***	–1.553*	0.229
	(0.672)	(0.464)	(0.816)	(0.287)
N	5,474	17,533	701	3,619
First stage F	463	976	124	788
Non-G&T Mean	22.952	23.052	23.097	23.171
p-value	[0.016]		[0.050]	
FRPL Share	–0.291***	–0.252***	–0.277***	–0.102***
	(0.036)	(0.030)	(0.046)	(0.016)
N	4,946	15,909	650	3,397
First stage F	481	1012	130	879
Non-G&T Mean	0.663	0.389	0.617	0.306
p-value	[0.399]		[0.000]	
White/Asian Share	0.102***	0.150***	0.160***	0.084***
	(0.038)	(0.026)	(0.043)	(0.013)
N	4,946	15,909	650	3,397
First stage F	481	1012	130	879
Non-G&T Mean	0.544	0.652	0.576	0.687
p-value	[0.286]		[0.067]	

Notes: This table reports 2SLS estimates of the effect of G&T enrollment on kindergarten classroom characteristics, separately by FRPL status. Columns 1-2 report fuzzy RD estimates, and columns 3-4 report estimates using the lottery design. See Table 3 for a description of the outcomes. The table reports Kleibergen-Paap F-statistics and outcome means for students who do not enroll in G&T. Robust standard errors are in parentheses. p-values for tests of the difference in G&T effects between FRPL and non-FRPL students are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table F5. Effects on Achievement by FRPL Status

	RD		Lottery	
	FRPL (1)	Non-FRPL (2)	FRPL (3)	Non-FRPL (4)
Math	–0.192*	–0.011	0.025	–0.048
	(0.109)	(0.078)	(0.119)	(0.048)
N (estimation)	10,768	30,338	1,449	6,648
N (students)	4,380	12,723	552	2,714
First stage F	90.2	188.6	28.9	206.1
Non-G&T Mean	0.964	1.107	1.173	1.326
p-value	[0.178]		[0.585]	
ELA	–0.087	0.103	–0.079	–0.011
	(0.100)	(0.073)	(0.115)	(0.044)
N (estimation)	13,282	36,802	1,764	7,948
N (students)	4,408	12,830	556	2,725
First stage F	63.2	127.1	22.3	149.4
Non-G&T Mean	0.891	1.061	1.084	1.234
p-value	[0.204]		[0.371]	
Algebra Regents	–0.081	0.137	0.222	–0.043
	(0.184)	(0.122)	(0.145)	(0.053)
N (estimation)	1,080	3,226	139	779
First stage F	104.0	257.8	105.4	573.9
Non-G&T Mean	0.880	1.060	0.899	1.197
p-value	[0.096]		[0.483]	

Notes: This table reports 2SLS estimates of the effect of G&T enrollment on test scores, separately by FRPL status. Columns 1-2 report fuzzy RD estimates, and columns 3-4 report estimates using the lottery design. See Table 4 for a description of the outcomes and estimation procedure. The table reports Kleibergen-Paap F-statistics and outcome means for students who do not enroll in G&T. Robust standard errors are in parentheses. p-values for tests of the difference in G&T effects between FRPL and non-FRPL students are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table F6. Effects on SHSAT Outcomes by FRPL Status

	RD		Lottery	
	FRPL (1)	Non-FRPL (2)	FRPL (3)	Non-FRPL (4)
SHSAT Tested	–0.142	0.017	0.183**	–0.017
	(0.098)	(0.080)	(0.088)	(0.035)
N	2,205	5,619	289	1,201
First stage F	232.7	377.4	100.8	734.0
Non-G&T Mean	0.691	0.746	0.741	0.816
p-value	[0.206]		[0.038]	
Specialized HS Offer	–0.041	0.042	0.119	–0.040
	(0.087)	(0.086)	(0.112)	(0.048)
N	2,205	5,619	289	1,201
First stage F	232.7	377.4	100.8	734.0
Non-G&T Mean	0.213	0.335	0.316	0.485
p-value	[0.497]		[0.201]	

Notes: This table reports 2SLS estimates of the effect of G&T enrollment on high school outcomes, separately by FRPL status. See Appendix Table F3 for a description of the outcomes. The table reports Kleibergen-Paap F-statistics and outcome means for students who do not enroll in G&T. Robust standard errors are in parentheses. p-values for tests of the difference in G&T effects between FRPL and non-FRPL students are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table F7. Effects on Grade 9 Enrollment

	RD			Lottery		
	Overall	FRPL	Non-FRPL	Overall	FRPL	Non-FRPL
	(1)	(2)	(3)	(4)	(5)	(6)
Specialized High School, 9th Grade	0.073	0.094	0.068	–0.037	0.076	–0.060
	(0.078)	(0.112)	(0.105)	(0.050)	(0.126)	(0.055)
N	4,916	1,489	3,427	1,037	205	832
First stage F	412.7	154.1	253.1	650.4	69.0	571.6
Non-G&T Mean	0.303	0.214	0.340	0.444	0.284	0.484
p-value		[0.865]			[0.330]	
Value-Added, 9th Grade	0.034	0.044	0.035	0.001	0.036	–0.010
	(0.051)	(0.076)	(0.068)	(0.031)	(0.083)	(0.033)
N	4,948	1,501	3,446	1,041	204	837
First stage F	418.8	158.1	256.9	664.2	69.5	586.4
Non-G&T Mean	0.329	0.254	0.361	0.451	0.346	0.477
p-value		[0.927]			[0.614]	

Notes: This table reports 2SLS estimates of the effect of G&T enrollment on the characteristics of the school where a student enrolls in 9th grade. Columns 1-3 report RD estimates, and columns 4-6 report lottery estimates. The outcome in the first panel is a dummy variable indicating enrollment in a specialized high school. The outcome in the second panel is the estimated value-added on SAT scores at the school in which students enrolled in 9th grade. Value-added estimates are computed by following the procedure described in Appendix Section C.2. The table reports Kleibergen-Paap F-statistics and outcome means for students who do not enroll in G&T. Robust standard errors are in parentheses. p-values for tests of the difference in G&T effects between FRPL and non-FRPL students are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table F8. IV VAM Estimates

	3rd Grade Math (1)	3rd Grade ELA (2)	6th Grade Specialized HS Share (3)	6th Grade Math Value-Added (4)
Forecast coefficient	0.558	0.697	1.07	0.946
	(0.097)	(0.077)	(0.050)	(0.044)
Over-id (χ^2)	102	103	130	91.9
df	96	97	94	94
p-value	0.31	0.33	0.01	0.54
σ_β	0.13	0.16	0.11	0.08
σ_ν	0.05	0	0	0
First-stage F	190	156	152	116
N	5,937	5,943	4,751	4,751

Notes: This table reports estimates from the IV VAM procedure as described in Appendix Section C.5. The first row reports forecast coefficients obtained from instrumenting conventional value-added model estimates of G&T program effectiveness with program-specific offers. Value-added models include controls for baseline demographics interacted with year. The outcomes in columns 1 and 2 are 3rd grade math and ELA achievement scores. Columns 3 and 4 report impacts on the specialized high school share and 6th grade math value-added of the enrolled school in 6th grade. The middle rows report over-identification test statistics, degrees of freedom, and p-values. The last rows report estimates of the standard deviation of true value-added across G&T programs, the standard deviation of the forecast residual as defined in Appendix Section C.5, and the Kleibergen-Paap F-statistic. Estimates of σ_ν can be negative, in which case we report a value of 0. Robust standard errors are in parentheses.

Appendix Table F9. Effects on Achievement by Race/Ethnicity

	RD		Lottery	
	Black/Hispanic (1)	White/Asian (2)	Black/Hispanic (3)	White/Asian (4)
Math	–0.065	–0.122*	–0.462***	0.003
	(0.122)	(0.074)	(0.161)	(0.048)
N (estimation)	9,187	29,422	1,051	6,466
N (students)	3,851	12,218	427	2,596
First stage F	87.8	187	24.2	198
Non-G&T Mean	0.776	1.164	0.993	1.347
p-value	[0.682]		[0.007]	
ELA	0.006	–0.008	–0.383**	0.013
	(0.109)	(0.071)	(0.159)	(0.045)
N (estimation)	11,195	35,921	1,285	7,747
N (students)	3,901	12,292	431	2,607
First stage F	61.0	126	16.0	145
Non-G&T Mean	0.805	1.079	1.019	1.234
p-value	[0.838]		[0.010]	
Algebra Regents	–0.044	0.102	–0.052	0.004
	(0.178)	(0.123)	(0.182)	(0.052)
N (estimation)	1,046	3,074	128	738
First stage F	139	231	69.4	578
Non-G&T Mean	0.680	1.132	0.858	1.213
p-value	[0.596]		[0.254]	

Notes: This table reports 2SLS estimates of the effect of G&T enrollment on test scores, by student race/ethnicity. Columns 1-2 report fuzzy RD estimates and columns 3-4 report estimates using the lottery design. See Table 4 for a description of the outcomes. The table reports Kleibergen-Paap F-statistics and outcome means for students who do not enroll in G&T. Standard errors are in parentheses. Standard errors for math and ELA outcomes are clustered by student. p-values for tests of the difference in G&T effects between Black/Hispanic and White/Asian students are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table F10. Effects on Achievement by Gender

	RD		Lottery	
	Female (1)	Male (2)	Female (3)	Male (4)
Math	–0.092	–0.035	–0.103	0.025
	(0.092)	(0.086)	(0.064)	(0.065)
N (estimation)	21,534	19,572	4,440	3,657
N (students)	8,899	8,204	1,744	1,522
First stage F	147	135	117	111
Non-G&T Mean	1.030	1.114	1.278	1.324
p-value	[0.676]		[0.302]	
ELA	0.049	0.045	–0.055	–0.007
	(0.082)	(0.084)	(0.056)	(0.061)
N (estimation)	26,238	23,846	5,245	4,467
N (students)	8,991	8,247	1,752	1,529
First stage F	101	90.7	88.9	79.3
Non-G&T Mean	1.107	0.918	1.297	1.098
p-value	[0.647]		[0.171]	
Algebra Regents	0.078	0.106	0.062	–0.051
	(0.139)	(0.142)	(0.071)	(0.067)
N (estimation)	2,248	2,058	483	435
First stage F	197	175	350	300
Non-G&T Mean	1.007	1.026	1.138	1.171
p-value	[0.885]		[0.266]	

Notes: This table reports 2SLS estimates of the effect of G&T enrollment on test scores, by student gender. Columns 1-2 report fuzzy RD estimates and columns 3-4 reports estimates using the lottery design. See Table 4 for a description of the outcomes. The table reports Kleibergen-Paap F-statistics and outcome means for students who do not enroll in G&T. Standard errors are in parentheses. Standard errors for math and ELA outcomes are clustered by student. p-values for tests of the difference in G&T effects between female and male students are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table F11. Effects on 6th Grade Applications and Enrollment by Race/Ethnicity

	RD		Lottery	
	Black/Hispanic	White/Asian	Black/Hispanic	White/Asian
	(1)	(2)	(3)	(4)
A: First Choice 6th Grade School				
Specialized HS Share	0.025	–0.014	–0.001	0.076***
	(0.028)	(0.025)	(0.042)	(0.014)
N	2,952	9,364	343	2,056
First stage F	315.5	627.0	90.4	653.9
Non-G&T Mean	0.127	0.221	0.181	0.257
p-value	[0.293]		[0.089]	
Value-Added	–0.021	0.007	0.035	0.025*
	(0.031)	(0.023)	(0.033)	(0.014)
N	2,952	9,364	343	2,056
First stage F	315.5	627.0	90.4	653.9
Non-G&T Mean	0.029	0.077	0.061	0.094
p-value	[0.471]		[0.802]	
B: Enrolled 6th Grade School				
Specialized HS Share	0.017	–0.051**	0.049	0.091***
	(0.021)	(0.024)	(0.039)	(0.014)
N	2,343	7,490	284	1,678
First stage F	276.1	498.0	85.5	631.9
Non-G&T Mean	0.081	0.170	0.124	0.212
p-value	[0.035]		[0.321]	
Value-Added	–0.030	–0.003	0.052	0.016
	(0.033)	(0.025)	(0.037)	(0.014)
N	2,343	7,490	284	1,678
First stage F	276.1	498.0	85.5	631.9
Non-G&T Mean	0.011	0.064	0.038	0.088
p-value	[0.506]		[0.375]	

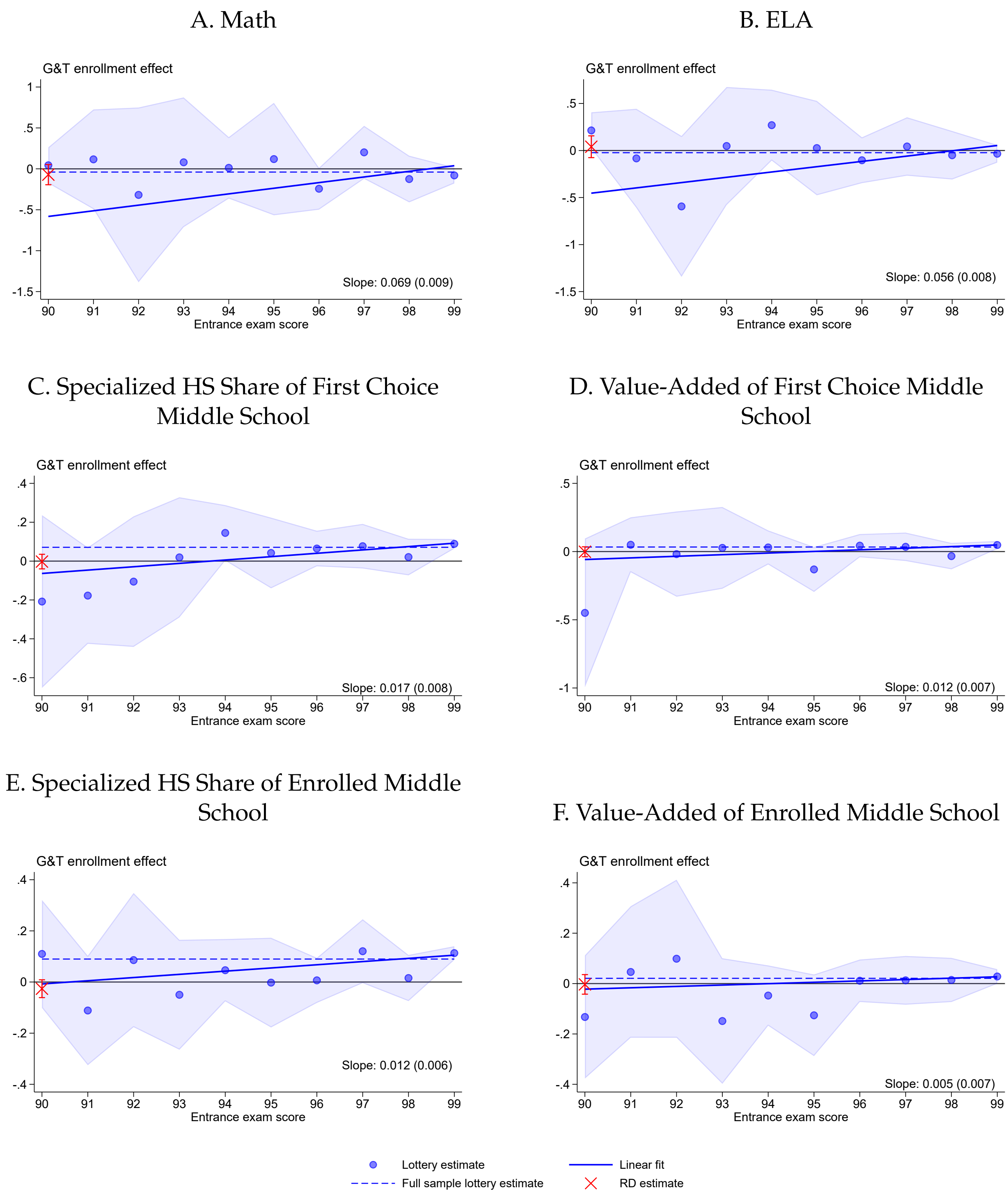
Notes: This table reports 2SLS estimates of the effect of G&T enrollment on characteristics of the 6th grade schools where students apply as their first choice and enroll, by race/ethnicity. Columns 1-2 report fuzzy RD estimates and columns 3-4 report estimates using the lottery design. See Table 5 for a description of the outcomes. The table reports Kleibergen-Paap F-statistics and outcome means for students who do not enroll in G&T. Robust standard errors are in parentheses. p-values for tests of the difference in G&T effects between Black/Hispanic and White/Asian students are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table F12. Effects on 6th Grade Applications and Enrollment by Gender

	RD		Lottery	
	Female (1)	Male (2)	Female (3)	Male (4)
A: First Choice 6th Grade School				
Specialized HS Share	–0.029	0.029	0.056***	0.085***
	(0.027)	(0.028)	(0.018)	(0.019)
N	6,865	6,209	1,399	1,181
First stage F	522.8	441.3	397.1	357.5
Non-G&T Mean	0.198	0.199	0.245	0.252
p-value	[0.150]		[0.288]	
Value-Added	–0.013	0.012	0.032*	0.038**
	(0.026)	(0.026)	(0.018)	(0.017)
N	6,865	6,209	1,399	1,181
First stage F	522.8	441.3	397.1	357.5
Non-G&T Mean	0.064	0.066	0.080	0.096
p-value	[0.523]		[0.795]	
B: Enrolled 6th Grade School				
Specialized HS Share	–0.056**	0.004	0.063***	0.118***
	(0.026)	(0.025)	(0.017)	(0.020)
N	5,496	4,896	1,135	964
First stage F	413.0	381.5	420.9	333.0
Non-G&T Mean	0.150	0.149	0.205	0.198
p-value	[0.088]		[0.040]	
Value-Added	–0.016	0.011	0.008	0.041**
	(0.029)	(0.027)	(0.017)	(0.019)
N	5,496	4,896	1,135	964
First stage F	413.0	381.5	420.9	333.0
Non-G&T Mean	0.051	0.051	0.079	0.081
p-value	[0.522]		[0.206]	

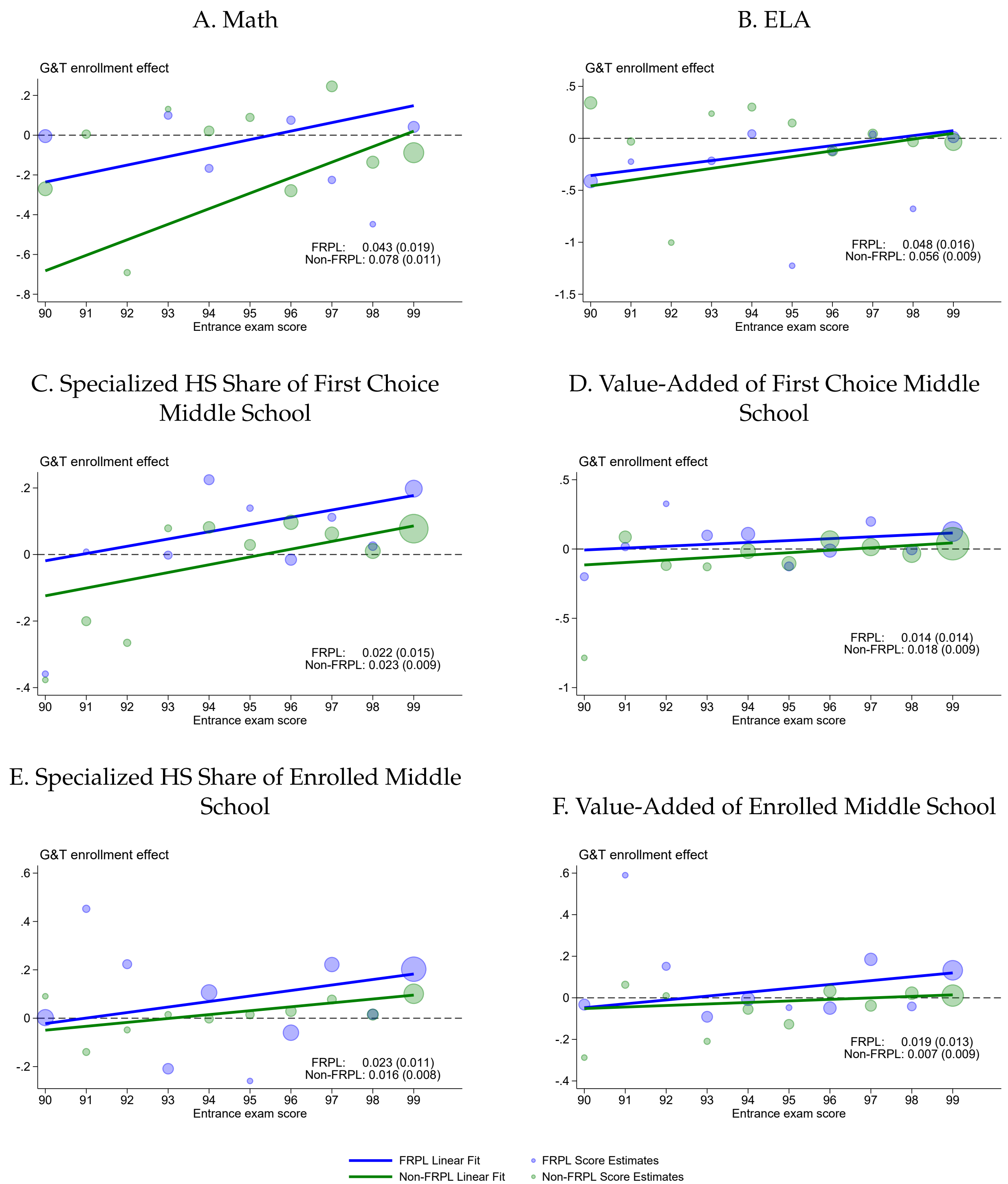
Notes: This table reports 2SLS estimates of the effect of G&T enrollment on characteristics of the 6th grade schools where students apply as their first choice and enroll, by gender. Columns 1-2 report fuzzy RD estimates and columns 3-4 report estimates using the lottery design. See Table 5 for a description of the outcomes. The table reports Kleibergen-Paap F-statistics and outcome means for students who do not enroll in G&T. Robust standard errors are in parentheses. p-values for tests of the difference in G&T effects between male and female students are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Figure F1. G&T Causal Effects by Admissions Exam Score



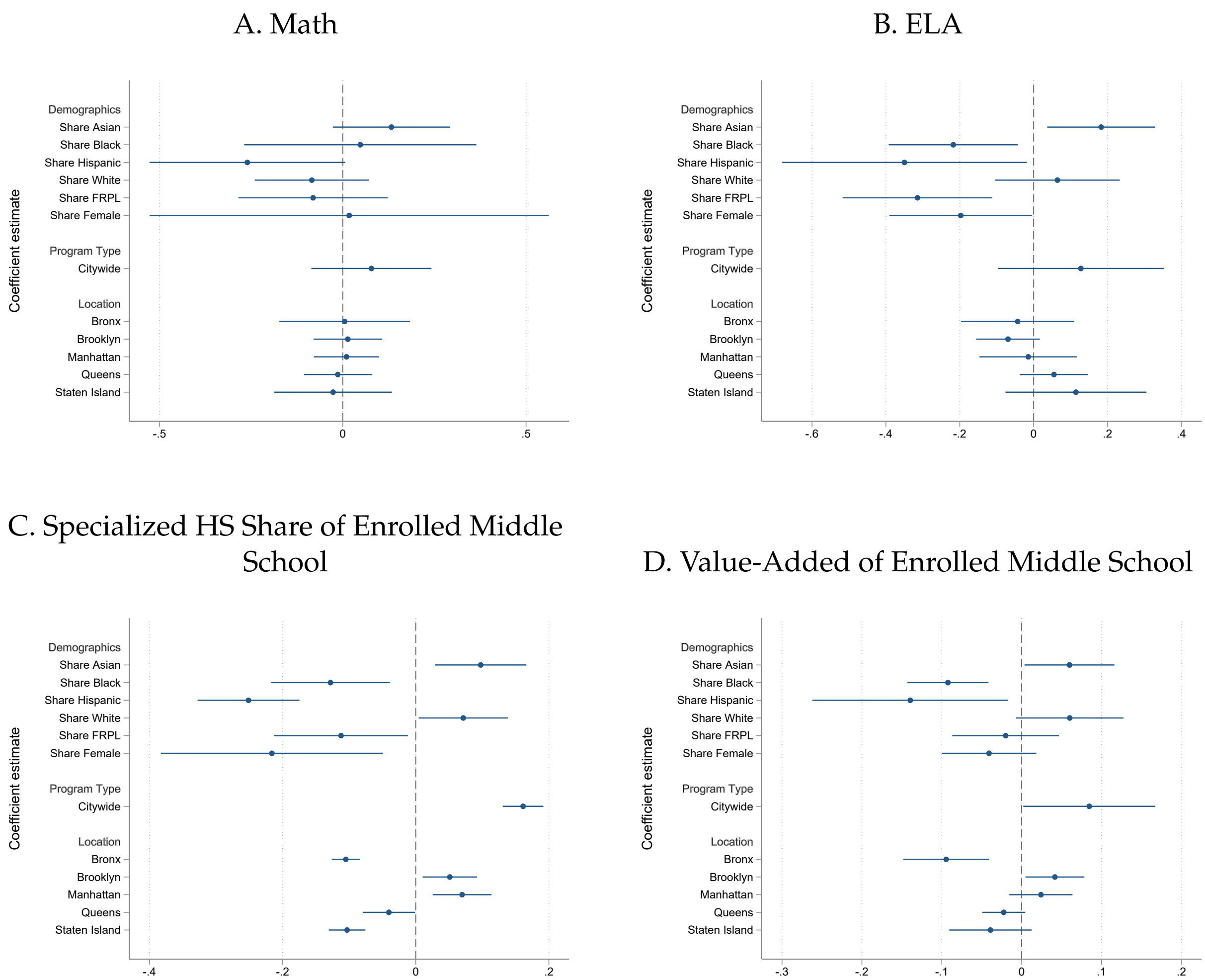
Notes: The plots in this figure depict how G&T effects vary by entrance exam score. The blue markers correspond to 2SLS estimates obtained via estimation in separate samples for each value of the score. Marker size is proportional to the inverse of the standard error on the estimate. The blue line is obtained from a model with a dummy for G&T enrollment and the dummy interacted with the exam score as two endogenous variables. The red cross corresponds to the fuzzy RD estimate.

Appendix Figure F2. G&T Effects Across the Score Distribution by FRPL Status



Notes: The plots in this figure depict how G&T effects vary by entrance exam score, separately by FRPL status. The blue markers correspond to 2SLS estimates obtained via estimation in separate samples for each value of the score. Marker size is proportional to the inverse of the standard error on the estimate. Markers are omitted if the standard error on the point estimate is greater than 0.8. The blue line is obtained from a model with a dummy for G&T enrollment and the dummy interacted with the exam score as two endogenous variables.

Appendix Figure F3. Correlates of G&T Value-Added



Notes: This figure depicts the relationship between program-level value-added estimates and program-level characteristics. Value-added estimates are obtained from a regression of the outcome on G&T program dummies and controls for baseline demographics. Outcomes in Panels A and B are 3rd grade math and ELA scores. Outcomes in Panels C and D are the specialized high school share and math value-added of the enrolled school in 6th grade. The coefficient estimates in this figure are obtained from bivariate regressions of the value-added estimate on program characteristics. Whiskers depict 95% confidence intervals.

G School Choice Model Details

G.1 Estimating Admissions Probabilities

We obtain a consistent estimate of admissions probabilities under exam admissions by following the bootstrap procedure from Agarwal and Somaini (2018). We use $B = 10,000$ bootstrap iterations. We generate each bootstrap sample $b \in \{1, ..., B\}$ by sampling with replacement from all applicants, holding fixed their preferences and priorities. Denote ϱ_i as the priority bucket of applicant i (a cell defined by their sibling, district, exam score, and DIA priorities). Then we estimate admissions probabilities at each program as follows:

1. Draw a new lottery tie-breaker $t_i^b \sim U[0, 1]$ for each applicant. (The tie-breaker is used across all programs.)

2. Run the student-proposing DA algorithm with t_i^b. Capacities are set equal to the number of offers observed for each program in the actual 2019 match, only among the set of applicants in the structural estimation sample.

3. Denote ϱ_s^b as the marginal priority group for program s and denote τ_{sp}^b as the empirical fraction of applicants admitted to program s within priority group g. (The marginal priority group is the lowest priority group for which the assignment is not guaranteed. The priority group is defined by an applicant's sibling, district, exam score, and DIA-awarded priorities.)

4. Denote S_{is} as the fraction of bootstrap samples where an applicant i is guaranteed admission by being in a strictly higher priority group than the marginal priority group. Then, the simulated assignment probability at s for i (if s had been ranked first) is:

$$q_{is} = S_{is} + \frac{1}{B} \sum_{b=1}^{B} 1(\varrho_{is} = \varrho_s^b) \times \tau_{sp}^b$$

With an estimate of the assignment probabilities q_{is}, we can estimate the rational expectations probability of receiving an offer for program s for each program: $p_{is}(A_i, q_i)$. Recall that applicants receive at most one G&T offer. We assume that applicants do not internalize that a single random number is used in the lotteries across all programs, which implies that p_{is} takes the following form:

$$p_{is} = q_{is} \times \prod_{j \in A_{k-1}} (1 - q_{ij})$$

where A_{k-1} denotes the truncated ROL of all programs ranked above program s.

Theorem 2 of Agarwal and Somaini (2018) demonstrates that the estimate $\hat{p}_{is}$, obtained from plug-in estimation of the formulas above, is consistent for p_{is}. The probability p_{is} captures ex-ante uncertainty in admissions that comes both from uncertainty in the composition of the applicant pool and uncertainty in lottery numbers.

We estimate admissions probabilities under recommendation-based admissions using the same procedure, but removing priorities based on entrance exam scores.

G.2 Parameter Estimation Details

We estimate the parameters of the model via simulated maximum likelihood. The likelihood in Equation 4 has no closed form, so we approximate the likelihood as follows. We apply a logit kernel smoother to approximate application probabilities as

$$\Pr(A_i = a \mid X_i, \epsilon_i, c_i) \approx \frac{\exp((\sum_{s \in a} w_{is} \times p_{is}(a) - C_i(a))/\lambda)}{\sum_{a'} \exp((\sum_{s \in a'} w_{is} \times p_{is}(a') - C_i(a'))/\lambda)},$$

where the logit smoother is set to $\lambda = 0.05$.

Offer probabilities, given by the second part of the integrand in Equation 4, do not depend on Ω so can be pulled out of the likelihood.

The last part of the integrand gives the probability of enrollment after preference shocks ξ_{is} are realized. At this stage, the likelihood of enrolling in G&T is given by

$$\Pr(S_i = s \mid Z_i, A_i, O_i, X_i, \epsilon_i) = \frac{\exp(u_{is}/\kappa)}{\exp(u_{is}/\kappa) + \exp(u_{i0}/\kappa)}$$

The expression for an individual's likelihood contribution across all stages is given by Equation 4. We approximate the likelihood via simulation using 128 realizations of ϵ_i and c_i drawn using a Sobol sequence.

For always-takers, who are observed enrolling in G&T programs despite not having a recorded offer (for example, if they received a waitlist offer that we do not observe), we only use the observed application decision and continue to record no acceptances in the simulation.

G.3 Parameter Estimates and Model Fit

Table G1. Model Parameter Estimates

	Estimate
	(1)
Distance coefficient	–0.658
	(0.008)
District priority coefficient	1.558
	(0.038)
Sibling priority coefficient	3.767
	(0.057)
Marginal cost scale parameter (σ_ζ)	0.096
	(0.002)
Scale parameter for preference shock (κ)	0.716
	(0.030)
Mean citywide fixed effect	4.139
Mean district fixed effect	–2.389
Share with cost below 0.01	0.083
N	2,363

Notes: This table reports parameter estimates from maximizing equation 4 via simulated maximum likelihood. The procedure normalizes the initial utility of the outside option u_{i0} to 0 and the scale parameter on the application-stage preference shock ϵ_{is} to 1. Distance is the great circle distance in miles between the centroid of a student's residential tract and the school. Standard errors, computed using the estimated Hessian, are in parentheses. See Appendix G.2 for more details on the procedure.

Table G2. Observed and Simulated G&T Demographic Shares

	Observed (1)	Simulated (2)	95% Prediction Interval (3)
FRPL	0.212	0.214	[0.201, 0.225]
Black	0.033	0.035	[0.028, 0.041]
Asian	0.485	0.475	[0.462, 0.489]
Hispanic	0.072	0.071	[0.063, 0.079]
White	0.408	0.415	[0.401, 0.428]

Notes: This table reports observed and simulated demographic shares of G&T students. Column 1 reports observed shares of kindergarten G&T students in 2019, among those for whom demographic information is non-missing. Column 2 reports mean demographic shares computed using 200 simulations of the model. Column 3 reports the 95% prediction intervals across 200 simulations. Prediction intervals are computed by drawing parameters Ω from a multivariate normal with mean $\hat{\Omega}$ and covariance matrix obtained from the estimated Hessian. We then use the parameters to draw applicant-specific shocks $(c_i, \epsilon_{is}, \xi_{is})$ from their relevant distributions, along with uniformly distributed lottery tie-breakers. We solve for optimal ROLs, run the DA algorithm, and simulate whether the applicant would accept or reject their G&T offer.

H Appendix References

AGARWAL, N. AND P. SOMAINI (2018): "Demand Analysis Using Strategic Reports: An Application to a School Choice Mechanism," *Econometrica*, 86, 391–444.

ANGRIST, J., P. HULL, P. A. PATHAK, AND C. WALTERS (2024): "Credible School Value-Added with Undersubscribed School Lotteries," *Review of Economics and Statistics*, 106, 1–19.

ANGRIST, J., P. HULL, AND C. WALTERS (2023): "Methods for Measuring School Effectiveness," in *Handbook of the Economics of Education*, ed. by E. A. Hanushek, S. Machin, and L. Woessmann, Amsterdam, Elsevier, vol. 7, 1–60.

ANGRIST, J. D. AND I. FERNÁNDEZ-VAL (2013): "ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework," in *Advances in Economics and Econometrics*, ed. by D. Acemoglu, M. Arellano, and E. Dekel, Cambridge University Press, vol. III of *Econometric Society Monographs, Econometrics*, chap. 11, 401–434.

ARMSTRONG, T. B. AND M. KOLESÁR (2020): "Simple and Honest Confidence Intervals in Nonparametric Regression," *Quantitative Economics*, 11, 1–39.

BEHAGHEL, L., B. CRÉPON, AND M. GURGAND (2013): "Robustness of the Encouragement Design in a Two-Treatment Randomized Control Trial," Working Paper.

BHULLER, M. AND H. SIGSTAD (2024): "2SLS with Multiple Treatments," *Journal of Econometrics*, 242, 105785.

BUI, S. A., S. G. CRAIG, AND S. A. IMBERMAN (2014): "Is Gifted Education a Bright Idea? Assessing the Impact of Gifted and Talented Programs on Students," *American Economic Journal: Economic Policy*, 6, 30–62.

CARD, D., E. CHYN, AND L. GIULIANO (2024): "Can Gifted Education Help Higher-Ability Boys from Disadvantaged Backgrounds?" NBER Working Paper No. 33282.

HEINESEN, E., C. HVID, L. KIRKEBØEN, E. LEUVEN, AND M. MOGSTAD (2022): "Instrumental Variables with Unordered Treatments: Theory and Evidence from Returns to Fields of Study," NBER Working Paper No. 30574.

KIRKEBOEN, L. J., E. LEUVEN, AND M. MOGSTAD (2016): "Field of Study, Earnings, and Self-Selection," *Quarterly Journal of Economics*, 131, 1057–1111.

KOLESÁR, M., R. CHETTY, J. FRIEDMAN, E. GLAESER, AND G. W. IMBENS (2015): "Identification and Inference With Many Invalid Instruments," *Journal of Business & Economic Statistics*, 33, 474–484.

KOLESÁR, M. AND C. ROTHE (2018): "Inference in Regression Discontinuity Designs with a Discrete Running Variable," *American Economic Review*, 108, 2277–2304.

WOOLDRIDGE, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data*, MIT Press.