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

Jimmy Chin† Geoffrey Kocks‡

Job Market Paper

November 14, 2025

Please click here for latest version

Abstract

We study the impacts of New York City kindergarten gifted and talented (G&T) programs on achievement and access to elite secondary schools. We combine two research designs: a regression discontinuity at a qualifying exam cutoff and a lottery design arising from centralized assignment. The latter permits the identification of effects for students scoring above the cutoff. While G&T markedly changes the classroom environment, there is no impact on achievement using both empirical strategies, with precise and insignificant effects smaller than 0.04σ when pooling the designs. In contrast, G&T boosts applications and enrollment in elite middle schools in the lottery sample. Effects on school access are largest among low-income students and those with high baseline ability. We use our estimates to predict how a recent admissions reform that replaced the G&T entrance exam with teacher recommendations affects elite school access. The reform increased the share of low-income students in G&T from 22 to 28 percent. A structural model of G&T demand shows that it also decreased the mean baseline ability. The decline in baseline ability outweighs the gains in low-income enrollment, lowering the average treatment effect on elite school access for G&T students. We trace out a policy possibilities frontier, comparing treatment effects and the low-income enrollment share under best-case policies. While neither exam nor recommendation admissions are on the frontier, exam-based admissions are closer.

*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, remains a cornerstone of the debate. Supporters argue that gifted and talented (G&T) programs, and tracking more broadly, 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).[2]

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.[3] 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 particularly 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

[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]Similar debates arise around tracking. For example, San Francisco and Cambridge restricted math tracking in the name of equity and diversity, only to partially reverse course following public backlash (Napolitano, 2024; Sifferlen and Farrell, 2023).

[3]To the extent that human capital interventions are more effective at younger ages, early identification could be a promising approach for improving long-term outcomes (Currie and Almond, 2011; Heckman, 2006).

city's high-performing secondary schools (Closson, 2024).[4]

To answer our first research question, we identify causal effects using two features of the G&T assignment system used during the exam-based admissions regime 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 uses 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 point estimate in absolute value for math and English Language Arts (ELA) is -0.04 standard deviations and confidence intervals exclude effects larger in magnitude than roughly ± 0.1 standard deviations. 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 $+0.5$ standard deviation 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

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

two characteristics of the middle schools to which students apply and enroll.[5] The first outcome is the share of a middle school's students that later enroll in the city's elite specialized high schools.[6] This outcome, which we refer to as the *specialized high school share*, 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 their enrolled middle school by 9.0 percentage points. The impacts on application 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, leverag-

[5] We focus on middle school effects to maximize precision, but also report results for high school.

[6] NYC has nine specialized high schools that require an exam for admission. The schools enroll extremely high-achieving students and admissions are intensely competitive.

ing 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 schools which students rank as their first choice decreases from $+5$ percentage points to -5 percentage points when moving from test to recommendation 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 recommendations-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; Strickland et al., 2024; Thompson, 2021), as well as to a closely related literature on the effects of tracking (Abdulkadiroğlu et al., 2014; Antonovics et al., 2022; Barrow et al., 2020; Card and Giuliano, 2016a; Cohodes, 2020; Carlana et al., 2022; Dobbie and Fryer, 2014; Duflo et al., 2011; Urquiola et al., 2024). Our study differs from prior work by employing a research design that identifies effects on inframarginal students. We provide novel evidence tracing out the gradient of the G&T treatment effects as a function of baseline ability.[7] We also study very early-age G&T which begins earlier than G&T programs featured in prior work. The studies most closely related to ours are Cleveland (2023) and Strickland et al. (2024), both of which study NYC G&T. The former estimates short-run impacts on test scores (only for FRPL students) and enrollment in public schools, while the latter estimates test score impacts using matching methods. We build on this work by using a research design that exploits all of the random variation embedded in the G&T lotteries, estimating effects on both FRPL and non-FRPL students and for longer-run outcomes, and by simulating the impacts of an admissions reform.

[7] Bui et al. (2014) estimate treatment effects on inframarginal students with a lottery design, though their results are derived from a single lottery and are less precise.

Our paper also relates to the literature on aspiration and preference formation in school choice settings (Ajayi et al., 2017; Andrabi et al., 2017; Corcoran et al., 2018; Corradini, 2024; Idoux and Corradini, 2023; Hahm and Park, 2024; Hastings and Weinstein, 2008). While past research focuses primarily on information interventions, we study the effects of a large change in the primary school environment on subsequent schooling decisions.[8] 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., Plucker et al., 2010; 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). For NYC, other work highlights policies like public pre-K and information interventions as ways to improve access to G&T programs (Lu and Weinberg, 2016; Lu et al., 2020). 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 effects.

2 Setting and Data

2.1 Gifted and Talented Programs in NYC

G&T programs are extensive within NYC. In the 2022-23 school year, approximately three percent of kindergarten students were enrolled across 85 G&T programs.[9] Gifted and non-gifted students are sorted into separate classrooms and schools, which distinguishes it from gifted programs in other school 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 a subject of 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

[8] Two exceptions are Idoux and Corradini (2023) and Hahm and Park (2024), which study middle school peer effects.

[9] In NYC, a school may contain distinct programs, each of which could vary in admissions criteria.

high schools.[10]

The district offers G&T programs for students enrolled in kindergarten through fifth grade, with kindergarten 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.[11] There are two types of G&T programs, referred to as *district* and *citywide* programs. 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.[12] Citywide programs are likewise open to all students but do not offer district priority. The programs are hosted at five schools that do not offer other district or non-G&T programs. Four of the five citywide schools enroll students from grades K-8. The fifth, New Explorations into Science, Technology and Math, enrolls students from K-12. By contrast, 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 analysis. We then apply our estimates to understand the potential effects of the admissions reform.

2.2 Exam Admissions

During the period with exam-based admissions, families initiate the application process by requesting the G&T entrance exam.[13] Two underlying 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.

[10] For example, those in favor of eliminating G&T argued that it disproportionately advantaged White and Asian students and "gives them an advantage to later be admitted into selective middle and high-school programs" (Dangor, 2021) Meanwhile, Eric Adams argued: "The levels of segregation at the NYC specialized high schools are exacerbated by the inability of the DOE to effectively make G&T education widely accessible and inclusive because strong G&T programs historically have had high placement rates."

[11] See more information here: `https://www.schools.nyc.gov/enrollment/enroll-grade-by-grade/gifted-talented`.

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

[13] 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, with families knowing the score at the time of application, and families were notified of assignments in June 2020. G&T assignment was decentralized prior to 2008.

Let $i = 1, ..., N$ index students and $s = 1, ..., S$ index G&T programs. For each student i, percentile ranks achieved on the two tests were combined to form an integer-valued entrance exam score $R_i \in \{0, 1, ..., 99\}$, which served as the basis for G&T qualification. Students needed a composite score of at least 90 to qualify for a district program and at least 97 to qualify for citywide programs. See Appendix A for more details about the exams and composite score formula.

Families ranked 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" (Gale and Shapley, 1962). Each student had 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. When an applicant is ineligible at s, their priority is $\varrho_{is} = \infty$. Priority groups ϱ_{is} are determined by unique combinations of sibling status, in-district status, and the entrance exam score R_i.[14]

All district programs used the same priority ordering. The highest priorities (smaller values of ϱ_{is}) 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. Therefore, 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.[15]

At oversubscribed programs, ties among equally qualified applicants in same priority group were broken via a random lottery number. Specifically, each applicant was as-

[14] Applicants had priority at programs located in their school district. Some programs offered priority to applicants in neighboring districts when there were no G&T programs offered in the neighboring district. Appendix Figure A1 depicts the geographic distribution of kindergarten programs across districts.

[15] The following example illustrates how priorities are defined. Consider priorities at district programs who 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 $\varrho_{is} = \infty$. An analogous formula applies for district programs with DIA priorities and citywide programs.

signed a tie-breaker $t_i \overset{iid}{\sim} U[0,1]$. Note that the tie-breaker is fixed for each applicant across all programs. Define *applicant rank* as $\chi_{is} = \varrho_{is} + t_i$. Within 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: $\varrho_s = \max_i\{\varrho_{is} | i \text{ assigned to } s\}$. 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.

The assignment process for G&T was completely independent from the non-G&T assignment process (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.[16] 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 score cutoff 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 participated in the match. Panel A of Appendix Figure A2 shows that, in fact, there were

[16] 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.

fewer G&T applicants in 2021 than 2020 or 2019, continuing the downward trend in the number of applications 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 (see Panel B).

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). Moreover, the increase in the FRPL share was much larger than for other demographic groups, such as those for race and ethnicity (see Appendix Table A1).

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.[17] We use information generated by the G&T match, which include student rank-ordered preferences over programs, priorities at those programs, and the program to which they were assigned.[18] 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.[19] See Appendix B for more details on data sources.

We construct three outcomes to measure the elementary school classroom environment. The first outcome is class size, which we compute using enrollment class codes. The remaining two outcomes are the share of peers in the classroom that are FRPL-eligible and the share who are White or Asian.

Our achievement outcomes are New York State standardized assessments in Math and ELA, which are administered in grades 3-8, and performance on the Algebra I Regents exam, which is typically administered in grades 8-9. All scores are standardized to be mean zero and unit variance for each test-grade-year. Tests are taken in the spring and

[17]When examining time series patterns in the G&T program during the admissions reform, we also use data on the 2021-22 school year.

[18]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 also are missing sibling priority information for some early years. We discuss further below.

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

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

Our school access outcomes are the characteristics of middle schools to which students rank as their first choice and enroll.[20] 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 is a measure of prestige. A relatively small number of "feeder" middle schools enroll a large number of students who go on to the specialized high schools and the specialized high school share is a publicized metric that conveys elite school status (Lee, 2018).

The second middle school outcome is *value-added* on sixth grade achievement. We estimate value-added from student-level regressions of sixth grade achievement on interactions between year and each of the following: school enrollment indicators, fifth grade demographic characteristics, and cubic functions of fifth grade math and ELA achievement, following Angrist et al. (2017, 2024a). Angrist et al. (2024a) show that value-added estimates for math scores 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.[21]

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.[22]

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

[20]While the deferred acceptance algorithm used in the match is strategy-proof (Gale and Shapley, 1962), at least in settings without restrictions on the number of programs that applicants can rank, a number of empirical studies document deviations from truth telling (see Rees-Jones and Shorrer, 2023). It is worth noting that even in the presence of strategic reporting, our outcome remains important for measuring whether G&T meaningfully changes school access. Ultimately, what matters for the pipeline is realized application behavior rather than true underlying preferences.

[21]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.

[22]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.

much less likely to be FRPL-eligible, have a disability, or be an English language learner. Roughly 22% of all kindergartners applied for G&T. Of the applicants, 26% were deemed eligible for district programs and 11% were eligible for citywide programs.

3 Empirical Strategies

Our objective is to estimate the treatment effects of enrolling in G&T for different student populations. We now describe the RD and lottery IV estimation procedures and how they relate to the estimands of interest.

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, defined as having score $R_i \geq 90$, does not imply enrollment.[23] We therefore 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 dummy variables for demographics, year dummies, and year-demographic interactions. The parameter of interest is β, which is the average treatment effects (LATE) of kindergarten G&T enrollment at the qualifying cutoff.[24] Because the RD design uses an instrument based only on qualification rather than offers, there is no need to control for types or priorities.

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. Column 3 of Table 1 reports summary statistics for the RD estimation sample. Relative to NYC overall, students are much more likely to be White and Asian, and much less likely to

[23] Note that we use the terms "qualification" and "eligibility" interchangeably to indicate attaining the minimum score of 90 for G&T identification. Attaining a score of at least 90 qualifies a student for district G&T. We specify that qualification or eligibility pertains to citywide programs when referring to the higher citywide cutoff of 97.

[24] Applicants may re-take the entrance exam in later years if they fail to qualify, but there remains a substantial first stage of qualification on ever enrolling in a G&T program, as shown in Appendix Figure C1.

be Black, Hispanic, FRPL-eligible, or have a disability. 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 estimates are robust to alternative bandwidths.[25]

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). We demonstrate that the instrument has a strong first stage, discussed further 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, we show in Appendix Table D1 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 which 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 are 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.[26]

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

[25] Our results are also robust to imposing a quadratic instead of linear fit, excluding baseline covariates, or pooling all application grades. We also obtain similar results using the honest RD framework of Kolesár and Rothe (2018).

[26] 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.

for each value of θ_i) impractical as there will be many students who do not share their type with others. Therefore, following Abdulkadiroğlu et al. (2017), our strategy pools across types by replacing 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 omitted variables 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 generates gains in 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. By the law of large numbers, this estimate converges to the true finite-market propensity score as the number of simulations increases (Abdulkadiroğlu et al., 2017). In our application, we simulated the propensity score by repeatedly running the match 100,000 times.[27]

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 conditional LATEs for strata defined by values of the propensity score.[28] Baseline covariates X_i are the same as in our RD approach and similarly serve to increase precision. We restrict the estimation sample to those with a non-degenerate risk of G&T assignment: $p_i \in (0.005, 0.995)$.[29] Note that, 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.[30] We report summary statistics for the lottery sample in column 4 of Table

[27]By contrast, prior applications of this identification strategy mostly use analytic formulas to compute a large-market approximation to the propensity score (e.g., Abdulkadiroğlu et al., 2017; Angrist et al., 2022, 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 propensity scores computed by simulation and formula align closely.

[28]Note that our procedure is tantamount to the recentering approach described by Borusyak and Hull (2023), who study the general problem of using treatments that arise from multiple sources of variation according to a known formula.

[29]The trimming rule drops observations who are essentially guaranteed assignment or non-assignment. Results are similar if we instead truncate to $p_i \in (0, 1)$.

[30]The lottery sample does not exclude applicants who score close to cutoff, so that applicants can appear

1, which reveals that students with non-degenerate assignment risk are more selected on demographics than the RD sample.[31]

Our strategy relies on accurate computation of the propensity score p_i. We note that prior to 2015, we have missing sibling priority data, preventing us from perfectly replicating G&T assignments. For these years, we assume that applicants do not have sibling priority when simulating the match. While this implies that our propensity scores are estimated with error, tests for covariate balance suggest 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.[32]

3.3 Attrition and Retention in NYC Schools

Differential attrition threatens the validity of either research design if the attrition decision 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 treatment effect estimates may be negatively biased. We note that attrition is a particularly outcome in the context of G&T as one rationale for the program is its potential to attract and retain families in the public school system. Prior work provides evidence that there are G&T retention effects in NYC (Cleveland, 2023) and elsewhere (Bui et al., 2014). Other work on tracking, however, has found little evidence of attrition for marginal students at the qualifying cutoff (Card and Giuliano, 2016a).

In the RD sample, we estimate retention effects of receiving a G&T offer by estimating grade-specific effects using 2SLS. In the lottery sample, we estimate the retention effect using the reduced form of the same outcome on G&T assignment Z_i, again estimated separately by grade. See Appendix C.1 for details.

Figure 2 depicts estimates of the effect of G&T assignment on enrollment across grades in NYC public schools. 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

in both the lottery and RD samples.

[31] Panel A of Appendix Figure C2 depicts the distribution of scores among applicants who scored 80 or above and Panel B shows that the sample with non-degenerate G&T assignment probability spans the full score distribution from 90 through 99.

[32] When sibling data is missing, 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.

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. Evidently, for some grades, G&T is effective at retaining lottery winners in the district.

We address attrition in several ways. First, given that the threat to validity arises from selection bias, we report tests for balance in the 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. See Appendix C.5 for details.

3.4 First Stage

Table 2 reports the first stage effects of the RD eligibility and lottery offer instruments on G&T enrollment. In addition to the endogenous variable in our analysis – kindergarten G&T enrollment – we also show the first stage effects on other first-stage measures of interest: ever enrolling in K-5 G&T and years enrolled in G&T.

Column 1 reports the effects of G&T eligibility. At the cutoff, the probability of enrolling in kindergarten G&T the year after application increases by 24 percentage points. The probability of ever enrolling in a G&T program (from kindergarten through 5th grade) increases by 16 percentage points. The difference between kindergarten enrollment and any K-5 enrollment reflects re-taking behavior in later grades. Finally, eligible students at the cutoff enroll in G&T for 0.75 more years than those who just missed the cutoff.

We also 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 are a manifestation of noncompliance. 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 offer. 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 the estimates of G&T effects obtained from the RD and lottery designs. Most of the prior work on G&T and tracking employs the RD design to estimate effects at a qualifying cutoff. A lottery design yields effect estimates that pertain to students both near and away from the cutoff, which might differ from those obtained via RD for several reasons.

First, the lottery design generates estimates that reflect a population with a larger range of baseline ability than those obtained via RD. While the RD design is a local effect for marginal students at the cutoff of 90, the lottery sample reflects a weighted average of LATEs across the score distribution from 90 to 99 (with the mean around 96). Therefore, differences across estimates might reflect heterogeneous treatment effects for students who vary in their baseline ability.

Moreover, 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.[33] It follows that lottery effect estimates might also reflect heterogeneous effectiveness across programs.

Appendix Table D2 compares the characteristics of compliers across the two designs. Relative to RD compliers, lottery compliers are more likely to be White and less likely to be Black, Hispanic, Asian, or FRPL (though the comparison is complicated somewhat by an increase in the share with other or missing race). Zooming in on the lower-scoring lottery compliers (those above the cutoff but within the RD bandwidth), we see that FRPL

[33]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, $p_{is'} = 0.5$.

shares are similar for RD and lottery compliers though differences by race still persist.

4 Causal Effects of G&T Enrollment

4.1 Kindergarten Classroom Environment

Enrolling in G&T significantly changes the characteristics of the classes a student attends, 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 potentially boost achievement, though evidence on this 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 the strong role of selection bias in comparisons of schools with different racial compositions (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 -0.5 students, as reported in column 2 of Table 3. While enrollment causes a 13 percentage point decrease in the FRPL share and a 9 percentage point increase in the White and Asian share, both effects are smaller than their RD analogs. 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 (which is taken in either grade 8 or 9). We stack the samples across grades for the math and ELA outcomes to increase precision.[34] Panels D through F of Figure 4 depict the conditional means underlying the RD design.

The effect estimates are small and insignificant on all of the test score outcomes for both the RD and lottery design. Effects on ELA scores are 0.04 (SE = 0.06) standard deviations using the RD sample and -0.02 (SE = 0.04) standard deviations using the lottery design, while effects on math range from -0.07 (SE = 0.06) standard deviations for RD to -0.04 (SE = 0.05) standard deviations for lottery.[35] The stacked setup is over-identified, as there are grade-specific instruments for a single binary treatment. Therefore, we also report p-values obtained from testing the model's over-identifying restrictions. We uniformly fail to reject the null that the restrictions hold, 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. Finally, we similarly find little impact on Algebra I performance, though the estimates are less precise.[36]

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 similar point estimates of -0.04 for math and 0.02 for ELA, though the magnitude of the standard errors are 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

[34]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.

[35]In all cases, we are able to rule out moderately positive effect sizes based on the benchmarks proposed by Kraft (2020).

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

and research designs. Furthermore, our results suggest that the large correlation between G&T enrollment and test score outcomes, as reported in Appendix Table F1, is explained by selection bias.

Our findings are also 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.[37] While Card and Giuliano (2016a) find positive impacts on high-achieving non-gifted students who enroll in a G&T class, gifted students on the margin of qualification benefit little in their setting (Card and Giuliano, 2014).[38]

It is worth emphasizing that our lottery estimates corroborate the RD-based estimates of prior work. While others have studied how rank effects and negative self-perceptions can hurt students at the margin of G&T and advanced track programs (Urquiola et al., 2024; Barrow et al., 2020), our estimates weigh against this being an important mechanism in our setting. Negative rank effects are less likely to play a role for the inframarginal students who score far 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 programs 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 Middle and High Schools

As described above, we estimate impacts on the characteristics of middle school to which 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 enroll later 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.[39]

[37] Treatment effect bounds cannot rule out zero in their lottery design.

[38] Importantly, however, G&T in this district did have large impacts on college enrollment, AP course-taking, and grades for gifted boys at the cutoff (Card et al., 2024).

[39] See Appendix C.2 for details on the value-added model. Results for ELA value-added are similar. We also show in Appendix Figure B1 that the specialized high school share is correlated with estimated value-

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 or 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 0.03 standard deviations larger than those for the controls. There is a marginally significant 0.02 standard deviation 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.[40] Our results suggest that G&T admissions reforms have the potential 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. Moreover, changes in aspirations or information do not appear to be mediated by changes in test score performance.[41]

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 Table 6 and Figure 6, are uniformly small and insignificant (though testing and offer rates are relatively high for the non-G&T students in the sample). On average, it appears the lottery-driven impacts on school preferences dissipate by high school age.

added on specialized high school enrollment, suggesting that the outcome partly reflects causal effects.

[40]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 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.5 provides more details.

[41]Carlana et al. (2022) also suggest that the change in aspirations is a more important mediator than test score performance in explaining how a tutoring and counseling intervention closed the immigrant-native gap in school track choice (for males).

5 Heterogeneity

5.1 Baseline Ability

We first estimate how effects of kindergarten G&T vary across the distribution of entrance exam scores. Figure 7 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.

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).[42]

We first re-examine the impacts on the classroom environment, as shown in Table 7. In the RD sample, G&T continues to generate class size reductions, and we can reject that the magnitudes are the same for FRPL (-5.7) and non-FRPL students (-3.5). Changes in peer exposure are similar for both groups with 25-29 percentage point reductions in FRPL peer shares and 10-15 percentage point increases in White and Asian peer shares.

[42] A notable exception is Barrow et al. (2020), who find that higher tracks in high school decrease grades and the probability of attending a selective college for low-income students.

Turning to the lottery design, we find class size reductions only for FRPL students (-2.5) but none for non-FRPL students. G&T also has much larger impacts on peer exposure for the FRPL students.

We continue to find little effect on test score outcomes across the subgroups. None of the estimates reported in Table 8 are significant at the 5% level, and we cannot reject that any of the FRPL and non-FRPL effects are the same. 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. Our effects, by contrast, are relevant for gifted students alone.

We find stronger patterns of heterogeneity when estimating impacts on school access. The estimates in Table 9 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 8.1 percentage points for non-FRPL (though we fail to reject that the differences are the same due to smaller samples).[43] Results for value-added in 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 9 motivates an investigation of heterogeneity by FRPL and entrance exam score combined. Figure 8 reproduces Figure 7 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.[44] Effects are also somewhat larger for FRPL students.

Finally, we report estimates for high school outcomes in Table 10. We find that G&T has little impact for either FRPL or non-FRPL students in the RD sample. For FRPL stu-

[43] 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 grade 6 school by $0.169 (SE = 0.045)$, the value-added of the enrolled grade 6 school by $0.124 (SE = 0.038)$, and the probability of receiving an SHSAT offer by $0.394 (SE = 0.164)$ for FRPL boys.

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

dents 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 Program Type

Are G&T effects heterogeneous by program? A natural starting point is to compare district and citywide G&T. The contrast is salient for parents, and the more selective citywide programs appear to be especially desirable. Recall that citywide programs exclusively enroll gifted students, which suggests that the citywide treatment could be more intensive than the district programs. For example, greater homogeneity of student ability could facilitate greater acceleration of curricula.[45] Moreover, it is important to note that all five citywide programs continue through at least 8th grade. We 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.[46] Specifically, we use 2SLS to instrument district and citywide enrollment with district and citywide offers. By the same logic as above, district and citywide offers are ignorable conditional on the propensity scores for each (restricting the sample to applicants with a 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 higher) 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 (Behaghel et al., 2013; Bhuller and Sigstad, 2024; Heinesen et al., 2022; Kirkeboen et al., 2016). Restrictions on effect heterogeneity and behavioral responses to the instruments yield treatment effects that are easier to interpret. For example, constant effects ensure that the coefficients on each treatment correspond to an average treatment effect

[45]The school-wide model for citywide resembles the exam schools in NYC and Chicago studied by (Abdulkadiroğlu et al., 2014; Angrist et al., 2023b; Barrow et al., 2020; Dobbie and Fryer, 2014), albeit with entry at much earlier ages and with different admissions criteria.

[46]Note that it is difficult to estimate citywide effects using its qualifying cutoff in an RD design as there are only three points of support above the citywide cutoff of 97.

on compliers shifted into only that treatment. Alternatively, restrictions on instrument response types (i.e., principal strata as in Frangakis and Rubin, 2002) yields interpretable effects under arbitrary forms of treatment effect heterogeneity.[47] However, as noted by Bhuller and Sigstad (2024) and Heinesen et al. (2022), modest violations of constant effects and the behavioral restrictions likely matter little for bias.

We support the validity of the multisector strategy by investigating the extent to which the necessary behavioral restrictions are violated. Restrictions on principal strata yield testable implications, as noted by Bhuller and Sigstad (2024) and Heinesen et al. (2022). Appendix Table D3 reports the results of first stage regressions of each treatment on the two instruments to test whether there are "cross-effects" of each instrument on the other treatment. A positive effect of, for example, citywide offers on district enrollment while controlling for district offers suggests the presence of disallowed principal strata. While we reject that cross-effects are zero, we find that they are quantitatively small, suggesting only modest violations of the identification assumption. In Appendix C.4, we apply a bounding argument from Heinesen et al. (2022) to show that share of applicants who violate the necessary conditions are small, suggesting little scope for bias.

We report multisector 2SLS results in Table 11. Neither district nor citywide appear to have significant impacts on math and ELA scores (with the possible exception of a negative 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 increases 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 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 in the full

[47] With three values for treatment and three for instruments, as in our setting, Behaghel et al. (2013) were the first to show that the 2SLS estimand reflects an average treatment effect on compliers when there are only six response types such that take-up is a function of one instrument only. In our setting, this restriction implies that students either never enroll in G&T, always enroll in district G&T, always enroll in citywide G&T, or enroll in either G&T sector if and only if they were to have an offer there. Bhuller and Sigstad (2024) show that this sufficient condition is also necessary with unrestricted effect heterogeneity.

sample, effects are concentrated in the FRPL sample where district and citywide increase it by roughly 0.1 standard deviation. 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 effect heterogeneity across district and citywide programs, including in subsamples defined by FRPL status.[48] We also provide evidence that G&T effects on school access are 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.

5.4 Race and Gender

We report impacts on achievement by race and gender in Appendix Table F4 and Appendix Table F5. There is little heterogeneity in RD estimates by race. Lottery estimates are close to zero for Black and Hispanic students and large and negative for White and Asian students (though the estimates are imprecise). There is little heterogeneity in test score impacts by gender in either design.

Likewise, we find little heterogeneity using the RD design for middle school access outcomes. Appendix Table F6 and Appendix Table F7 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 value-added of the enrolled middle school 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

[48] We might also expect heterogeneity across the roughly 80 G&T programs in NYC, especially because schools are granted large levels of autonomy in how they implement their programs. Unfortunately, data constraints prevent a thorough investigation of program-level heterogeneity. We estimated value-added models (VAM) for G&T programs, but have found that the VAMs are significantly biased using lottery-based tests of VAM validity (Angrist et al., 2017, 2016). We suspect that the bias arises from the lack of pre-K test scores. This limitation also hinders the use of VAMs to gauge the quality of fallback options for G&T compliers.

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.

Recent reforms to the G&T program have emphasized the dual goals of equity and efficiency: increasing the diversity of G&T and improving the targeting of the program to those who benefit most.[49] Whether or not an admissions reform can simultaneously accomplish these goals depends on the scope of treatment effect heterogeneity and the extent to which the reform changes the composition of G&T students. We note that a simple comparison using data before and after the 2021 admissions reform is complicated by the COVID pandemic, as well as the inability to observe middle school outcomes for the more recent cohorts. Therefore, we simulate the effects of the reform holding all other factors constant, including the pre-reform treatment effect estimates.

6.1 Estimates from Reweighting on Observables

Our first strategy reweights estimated LATEs using observed demographic shares to obtain the ATT under each admissions regime. Specifically, we estimate the ATT by weighting FRPL-specific LATE estimates by the share of FRPL students in the treated population before and after the reform. To do this, we impose an additional assumption that effect heterogeneity is mediated by FRPL status alone. This permits us to extrapolate LATEs to always-takers in the treated population. Appendix C.6 provides a formal justification, following the argument of Angrist and Fernández-Val (2013). Under this assumption, we estimate the ATT in the testing regime as

$$\text{ATT}_t = \sum_g \beta_g \Pr(G_i^t = g | D_i^t = 1),$$

where G_i^t denotes covariate group membership in the testing period, D_i^t indicates treatment status in the testing period, and β_g is the 2SLS estimate obtained from using those with $G = g$ in the lottery sample. In this exercise, $G \in \{0, 1\}$ is binary and denotes FRPL or non-FRPL status.[50] The estimate of the post-reform ATT replaces pre-reform demographic shares with their post-reform analogs, holding fixed the estimated treatment

[49] For example, when announcing an expansion of G&T, Mayor Adams emphasized the equity goal of "removing inequities in the admissions process," while Chancellor David Banks cited the policy's potential to better "identify the students who will excel with accelerated learning" (NYC Department of Education Press Office, 2022).

[50] The 2SLS regression is estimated on cohorts applying for enrollment in 2017-2018 and earlier (where we observe the middle school outcomes).

effects:

$$\text{ATT}_r = \sum_g \beta_g \Pr(G_i^r = g | D_i^r = 1).$$

Reweighting on FRPL status suggests that the ATT on the characteristics of schools to which students apply and enroll changes minimally in response to the reform. Column 1 of Table 12, for example, shows that the ATT on the specialized high school share of an applicant's first choice middle school barely increases from 0.077 to 0.080 after the reform. The value-added ATT, and the ATT estimates for our enrollment outcomes, similarly change little. Effect heterogeneity and the increase in the post-reform FRPL share are not large enough to generate large changes in the ATT. However, the reform likely changed the composition of G&T students beyond FRPL status. In particular, it seems likely that many students who would have failed to qualify under exam admissions were able to enroll in G&T after the reform. The entry of these students is relevant for the ATT since, as described above, heterogeneity by baseline ability appears substantial for our middle school outcomes. Unfortunately, we cannot reweight on the distribution of entrance exam scores as post-reform G&T students do not take any entrance exam. Therefore, we turn towards a structural model of G&T demand that allows us to simulate the impacts of the reform, holding fixed a sample of students for whom we can observe the entrance exam score.

6.2 Structural Model of G&T Demand

6.2.1 Model Setup

Let $s \in \{0, 1, ..., S\}$ index an outside option and G&T programs. The outside option, denoted $s = 0$, includes enrollment in non-G&T programs in NYC, private schools, and schools outside NYC. The model proceeds in three stages.[51] In the first stage, 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. In the second stage, 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. In the third stage, applicants receive a preference shock then decide where to enroll by accepting or rejecting their G&T offer.[52]

[51]Our approach builds on structural models of school choice employed in prior work (Avery et al., 2025; Kapor et al., 2020; Walters, 2018).

[52]We abstract from the decision of whether to initially enter 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 changes in

The additional shock rationalizes the presence of never-takers. We now describe each stage in more detail.

6.2.2 Stage 1: Applications

In the first stage, families 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 $w_{is} = E[\max\{0, U_{is} - U_{i0}\}|u_{i0}, u_{is}]$. Offers provide option value, so applicants could apply to s even if $u_{is} < u_{i0}$.

Applicants select 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 expected utility for s, $p_{is}(A)$ 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. We describe these components in turn.

Utilities

Applicants initially have utility u_{is} over G&T programs which is additive in observables and an idiosyncratic preference shock. Formally, we have:

$$u_{is} = \delta_s + W_i'\phi + \epsilon_{is},$$

where δ_s is mean utility for program s and W_i is a vector that includes distance and assignment priorities for sibling and district. Distance is measured as the number of miles between the centroid of the applicant's tract of residence and the school. Let the vector $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_S)$ collect the mean utilities across all programs. The preference shocks are distributed as $\epsilon_{is} \sim \text{Gumbel}(0, 1)$ for all s. We normalize the value of the outside option so that $u_{i0} = \epsilon_{i0}$ and assume that the shocks are independent across students.

either the total number of applicants in Appendix Figure A2 or the demographics of applicants in Appendix Figure A3 provides support for this assumption.

After offers are realized, but prior to enrolling, applicants receive additional independent preference shocks $\xi_{is} \sim \text{Gumbel}(0, \kappa)$ 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:

$$U_{is} = u_{is} + \xi_{is} \qquad\qquad U_{i0} = u_{i0} + \xi_{i0}.$$

Finally, by standard Gumbel distribution properties: $w_{is} = \kappa \log(\exp(\frac{u_{is} - u_{i0}}{\kappa}) + 1)$. 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.

Assignment Probabilities

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. Let $p_{is}(A_i)$ denote this probability. Agarwal and Somaini (2018) show that a consistent estimate for p_{is} can be generated by repeatedly running the match with bootstrap samples and computing offer shares across simulations. Each iteration samples from the full population with replacement, holding fixed preferences, priorities, and ROLs. Appendix G.1 provides further details on the procedure.

Costs

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 the commonly observed phenomenon of ROLs that are shorter than the maximum length of 12.

We assume that cost scales linearly in application length such that:

$$C_i(A) = c_i |A|$$
$$c_i = c + \zeta_i,$$

where $c \geq 0$ is a baseline marginal cost and ζ_i is an individual-specific deviation. We assume this deviation is drawn from a truncated normal: $\zeta_i \sim \text{TN}(0, \sigma_\zeta, -c, \infty)$. Idiosyncratic costs are assumed independent of preference shocks ϵ_{is} and ξ_{is}.

6.2.3 Stages 2 and 3: Offers and Enrollment

In stage 2, the DA algorithm generates at most one G&T offer $O_i \in \{0, 1, ..., S\}$ for each applicant. Applicants who do not receive an offer to any program have $O_i = 0$.

Preference shocks ξ_{is} are realized in stage 3 after assignments but before enrollment. Let $S_i \in \{0, 1, ..., S\}$ denote the schools that families choose to enroll in. Families with an offer at s enroll in s ($S_i = s$) if and only if $U_{is} > U_{i0}$. Those who turn down their assignment enroll in the outside option ($S_i = 0$).

6.2.4 Identification

We briefly discuss the intuition for parameter identification in this model. Agarwal and Somaini (2018) demonstrate 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 "special regressor." As discussed in Idoux (2021), cost parameters are identified with variation in assignment probabilities conditional on observables. Variation in ROL length among observationally-equivalent applicants identifies the variance of the costs. Finally, conditional on offers and indirect utilities, the gradient of offer take-up as a function of indirect utilities identifies the scale parameter of the post-offer shock in utilities.

6.2.5 Estimation

An individual's likelihood contribution takes the form:

$$L_i(A_i, O_i, S_i|W_i) = \int\int \Pr(A_i|W_i, \epsilon_i, c_i) \Pr(O_i|A_i, W_i) \Pr(S_i|Z_i, A_i, O_i, W_i, \epsilon_i) dF(\epsilon_i, c_i). \tag{4}$$

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 2,358 applicants (64% of the full sample of applicants who scored 90 or above).[53]

Unfortunately, estimating the likelihood above directly by simulated maximum likelihood is computationally infeasible. The problem arises because the cardinality of $\mathcal{A}$, the

[53] 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.

set of all possible portfolios, is extremely large. With 81 programs and a maximum ROL length of 12, each student has roughly 3^{22} possible ROLs.

We take several steps to make the problem tractable. First, we restrict applicant choices by defining choice sets to include any program within the applicant's district, any other program that is 3 miles away or less, and any citywide program (if eligible). This reduces the number of programs in the choice set to an average of 11.4 programs (which also restricts $\mathcal{A}_i$). For 77% of applicants, this restriction includes their complete ROL and it includes the top choice for 95% of applicants. Second, we estimate the parameters of our model using only the top 4 choices on applicants' ROL (after imposing the choice set restrictions). Note that after dropping programs not in the choice set, the relative ranking of the remaining options remain the same. Only the ordering of the programs, rather than the absolute positions, matters for the likelihood. 62% of applicants submit a ROL of four choices or fewer, and 97% of G&T students enroll in a program that they ranked as a 1st-4th choice.[54]

After imposing our choice set restrictions, 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. See Appendix G for an expression of the likelihood and the optimization procedure.

6.2.6 Parameter Estimates and Model Fit

Our procedure recovers plausible parameter estimates that are consistent with prior work. As reported in Appendix Table G1, we find that utility strongly decreases in distance and increases in priority status. The scale parameter of the stage 3 shock κ equals 0.67, suggesting that many applicants receive large shocks prior to enrolling. The large magnitude is consistent with a sizable share (39%) of G&T offers that are turned down. Application costs are small but non-negligible, with an estimated 8% of applicants having costs smaller than 0.01. The mean $\hat{\delta}_s$ across citywide schools equals 3.8, suggesting that citywide schools are very desirable. The average of $\hat{\delta}_s$ across district programs, however, is negative and equals -2.0. We interpret this as evidence that most applicants prefer their outside option to a typical district G&T program. However, there are a large number of district programs so that applicants like at least one district program enough to justify submitting an application. This finding is consistent with the large number of applicants who list only a small number of schools on their ROL.

[54] Prior work employs alternative approaches for making the optimal portfolio problem tractable with many possible portfolios. For example, Idoux (2021) imposes limited rationality assumptions on applicant behavior. Che et al. (2023) also restricts their sample to a smaller market (Staten Island instead of NYC).

We assess model fit by comparing observed to simulated enrollments. We simulate enrollment as follows. Collect the parameters in $\Omega = (\boldsymbol{\delta}, \phi, \kappa, c, \sigma_\zeta)$. The first step draws parameters Ω from a multivariate normal with mean $\hat{\Omega}$ and covariance matrix obtained using the estimated Hessian matrix. 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. We set program capacities to be equal to the number of observed offers given by each program in 2019 within our estimation sample. We repeat this process 200 times and obtain enrollment patterns that are very similar to the observed distribution. Panel A of Figure 9 shows that the simulated distribution aligns closely with the observed distribution by baseline ability and FRPL status, despite the fact that neither FRPL status nor the entrance exam score was used to model utility. Appendix Table G2 also provides evidence that the observed and simulated demographic shares of G&T students align closely.

6.3 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.[55] 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.[56] 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. Again, we simulate enrollments 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 9 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 as another out-of-sample validation of model fit.[57] There is also a sharp decline in baseline

[55]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.

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

[57]The 95% prediction interval, computed using the 2.5th and 97.5th percentiles of the outcome across

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 distribution. As above, we compute the ATT under the recommendations system as $\sum_g \beta_g \Pr(G_i^r = g|D_i^r = 1)$, though we now index the groups by both FRPL and the entrance exam score. Unlike the model-free reweighting exercise above, the model provides a method of simulating the exam score distribution of post-reform G&T students. Therefore, we allow G to vary by FRPL status and exam score.

This exercise relies on two additional assumptions. First, we assume that treatment effects for students who were ineligible under the pre-reform system are equal to the estimated lottery effects for the lowest-scoring eligible applicants. Treatment effects on the ineligible students are not identified, but we obtain similar results using several different methods for imputing the missing treatment effect. Second, we assume that the effects of G&T are homogeneous by program. This would be violated, for example, if there were match effects between specific students and programs.

We find that the simulated ATT differs markedly from those obtained from re-weighting on FRPL status alone. Columns 2-4 of Table 12 report estimates of the ATT obtained from re-weighting FRPL and score-specific estimates. Column 2 estimates group-specific treatment effects in the pre-reform period using the FRPL-specific linear fits depicted in Appendix Figure F2. We find that the ATT on the specialized high school share of applicants' first choice school declined from $+5$ percentage points to -5 percentage points. The corresponding enrollment ATT declines from $+7$ percentage points to -1. The pattern is explained by the large decline in the baseline ability of new G&T students sharply who also have smaller treatment effects.[58] The result stands in stark contrast to results from an exercise that reweights by FRPL distributions alone, which matters little for the ATT. The ATT changes for value-added outcomes are less pronounced, but still show a decline under the recommendations system.

Columns 3 and 4 report additional results corresponding to alternative ways of estimating the FRPL and exam score-specific conditional average treatment effects. To increase precision, we report results from estimating treatment effects via two FRPL-specific score bins (scores 90-94 and scores 95-99) in column 3 and four FRPL-specific score bins (scores 90-92, 93-95, 96-98, and 99, as as shown in Figure 8) in column 4. The results are

200 simulations, is $[25.7\%, 29.7\%]$. Note that our exercise relies on stability of the preference parameters across time. This could be violated if the pandemic changed underlying preferences.

[58]Those treatment effects are also less slightly precisely estimated, which explains the imprecision of our ATT estimates.

qualitatively similar. Regardless of the specification, we find that the reform substantially decreases the ATT of G&T on applications and enrollment to selective middle schools.

We caution that the counterfactual comparisons rely on the assumption that the gradient in G&T program effects arises from differences in baseline scores, rather than from differences in programs. One challenge to this interpretation is the continuation of citywide programs through middle school grades. Recall that citywide eligibility required a score of 97 or above under the exam-based admissions system. The citywide programs also rank highly on the specialized high school share outcome. Therefore, treatment effects on high scorers could be partly driven by the mechanical effect of staying at the same school. We argue that our results do not solely reflect this margin for the following reasons. First, we find a similar gradient of treatment effects, depicted in Appendix Figure F1, when restricting to those ineligible for citywide (those with exam scores of 90-96). Second, approximately 73% of citywide kindergarten students are still in citywide G&T by grade 6, suggesting that a non-negligible share of students move to other programs. Third, we find positive treatment effects from district G&T in multisector model results reported in 11. Nevertheless, we acknowledge that we might overstate the recommendations-driven decline in the ATT if heterogeneity is driven more by program type than baseline score.

We conclude by asking what alternative policies could possibly achieve given the pattern of treatment effects, and how the existing policies compare to what is theoretically attainable. We visualize this exercise in Figure 10, which traces out a possibilities frontier with ATTs on one axis and FRPL shares on the other.[59] 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 shows a recommendation-based system. We find that neither of the existing policies lies on the frontier. This implies that admissions reforms need not entail an equity-efficiency tradeoff when moving away from either of the existing policies. The optimal policy, which admits high-ability FRPL students, could simultaneously increase the FRPL share and the ATT.

[59]These plots rely on the specification where LATEs are linear in entrance exam scores, as in column 2 of Table 12. We abstract from estimation error in these calculations. An interesting exercise that we leave to future work is the derivation of optimal treatment assignment rules that account for noise (Chernozhukov et al., 2025).

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 that we plan to explore in 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 an interesting question for future work. Universal screening, which NYC implemented in 2022, may prove to be a promising policy option.

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., G. GRAY-LOBE, C. M. IDOUX, AND P. A. PATHAK (2022): "Still Worth the Trip? School Busing Effects in Boston and New York," NBER Working Paper No. 30308, July.

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., P. HULL, AND C. WALTERS (2023a): "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.

ANGRIST, J. D., P. D. HULL, P. A. PATHAK, AND C. R. WALTERS (2016): "Interpreting Tests of School VAM Validity," *American Economic Review: Papers & Proceedings*, 106, 388–392.

——— (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., P. A. PATHAK, AND R. A. ZARATE (2023b): "Choice and Consequence: Assessing Mismatch at Chicago Exam Schools," *Journal of Public Economics*, 223, 104892.

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.

Antonovics, K., S. E. Black, J. B. Cullen, and A. Y. Meiselman (2022): "Patterns, Determinants, and Consequences of Ability Tracking: Evidence from Texas Public Schools," NBER Working Paper No. 30370.

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

Avery, C., G. Kocks, and P. A. Pathak (2025): "The Algorithm Advantage: Ranked Application Systems Outperform Decentralized and Common Applications in Boston and Beyond," NBER Working Paper No. 34207.

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.

Borusyak, K. and P. Hull (2023): "Nonrandom Exposure to Exogenous Shocks," *Econometrica*, 91, 2155–2185.

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.

CHE, Y.-K., D. W. HAHM, AND Y. HE (2023): "Leveraging Uncertainties to Infer Preferences: Robust Analysis of School Choice," ArXiv Preprint 2309.14297.

CHERNOZHUKOV, V., S. LEE, A. M. ROSEN, AND L. SUN (2025): "Policy Learning with Confidence," ArXiv preprint arXiv:2502.10653.

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.

CURRIE, J. AND D. ALMOND (2011): "Human Capital Development before Age Five," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, Elsevier, vol. 4, 1315–1486.

DANGOR, G. (2021): "New York City Schools Will End Controversial Gifted And Talented Classes," *Forbes*, https://www.forbes.com/sites/graisondangor/2021/10/08/new-york-city-schools-will-end-controversial-gifted-and-talented-classes/.

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.

FRANGAKIS, C. E. AND D. B. RUBIN (2002): "Principal Stratification in Causal Inference," *Biometrics*, 58, 21–29.

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

GALE, D. AND L. S. SHAPLEY (1962): "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, 69, 9–15.

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.

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.

HECKMAN, J. J. (2006): "Skill Formation and the Economics of Investing in Disadvantaged Children," *Science*, 312, 1900–1902.

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.

KAPOR, A. J., C. A. NEILSON, AND S. D. ZIMMERMAN (2020): "Heterogeneous Beliefs and School Choice Mechanisms," *American Economic Review*, 110, 1274–1315.

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

KRAFT, M. A. (2020): "Interpreting Effect Sizes of Education Interventions," *Educational Researcher*, 49, 241–253.

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.

LU, Y. AND S. L. WEINBERG (2016): "Public Pre-K and Test Taking for the NYC Gifted-and-Talented Programs: Forging a Path to Equity," *Educational Researcher*, 45, 36–47.

LU, Y., S. L. WEINBERG, AND M. MCCORMICK (2020): "Test-Taking for Gifted and Talented Kindergarten: Underscoring the Importance of Outreach," *Gifted Child Quarterly*, 64, 259–274.

NAPOLITANO, J. (2024): "San Francisco Voters Overwhelmingly Support Algebra's Return to 8th Grade," *The 74*, https://www.the74million.org/article/san-fran-voters-overwhelmingly-support-algebras-return-to-8th-grade/.

NYC DEPARTMENT OF EDUCATION PRESS OFFICE (2022): "Mayor Adams, Chancellor Banks Announce Expansion of Gifted and Talented Programs Citywide," Press release.

PLUCKER, J. A., N. BURROUGHS, AND R. SONG (2010): "Mind the (Other) Gap! The Growing Excellence Gap in K-12 Education." *Center for Evaluation and Education Policy, Indiana University*.

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

REES-JONES, A. AND R. SHORRER (2023): "Behavioral Economics in Education Market Design: A Forward-Looking Review," *Journal of Political Economy Microeconomics*, 1, 557–613.

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.

SIFFERLEN, J. AND D. FARRELL (2023): "After Delays and Pushback, Cambridge Schools are Bringing Back 8th Grade Algebra I," *GBH*, https://www.wgbh.org/news/education-news/2023-08-01/after-delays-and-pushback-cambridge-schools-are-bringing-back-8th-grade-algebra-1.

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. (2018): "The Demand for Effective Charter Schools," *Journal of Political Economy*, 126, 2179–2223.

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

Tables

Table 1. Summary Statistics

	All kindergarten students (1)	G&T kindergarten applicants (2)	RD sample (3)	Lottery sample (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	0.18	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	0.10	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	171,767	27,870	5,004

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	0.240***	0.412***
	(0.007)	(0.014)
N	27,870	5,004
Control mean	0.001	0.188
Ever enroll	0.155***	0.414***
	(0.011)	(0.016)
N	20,406	3,791
Control mean	0.129	0.256
Years enrolled	0.748***	2.294***
	(0.050)	(0.083)
N	20,406	3,791
Control mean	0.468	1.072

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 track 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 grade 5 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***	–0.521
	(0.352)	(0.454)
N	20,854	4,049
First stage F	1517	1020
Non-G&T Mean	22.332	23.507
FRPL Share	–0.261***	–0.125***
	(0.024)	(0.016)
N	20,854	4,049
First stage F	1517	1020
Non-G&T Mean	0.452	0.355
White/Asian Share	0.138***	0.090***
	(0.021)	(0.013)
N	20,854	4,049
First stage F	1517	1020
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 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)
ELA	0.042	–0.023	0.003
	(0.059)	(0.041)	(0.034)
N (estimation)	50,106	9,712	59,818
N (students)	17,254	3,281	19,587
First stage F	192	172	179
Non-G&T Mean	1.02	1.21	1.04
Over-id p-value	[0.703]	[0.300]	[0.530]
Math	–0.070	–0.039	–0.050
	(0.063)	(0.045)	(0.037)
N (estimation)	41,124	8,097	49,221
N (students)	17,118	3,266	19,441
First stage F	284	236	254
Non-G&T Mean	1.07	1.30	1.10
Over-id p-value	[0.143]	[0.340]	[0.264]
Algebra Regents	0.083	–0.003	0.018
	(0.100)	(0.049)	(0.045)
N (estimation)	4,307	918	5,225
N (students)	4,307	918	5,009
First stage F	370	693	530
Non-G&T Mean	1.02	1.15	1.03
Over-id p-value			[0.440]

Notes: This table reports 2SLS estimates of the effect of G&T enrollment 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. We omit grades for math because many students take high school Regents exams in 7th and 8th grade and are exempt from the middle school assessments. The specifications for ELA and math are over-identified with grade-specific instruments. The estimation procedure is described in Appendix C.3. The Algebra I outcome are 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 specifications are over-identified with lottery and RD instruments that are also grade-specific for ELA and math. The table reports Kleibergen-Paap F-statistics and outcome means for students who do not enroll in G&T. * significant at 10%; ** significant at 5%; ***

Table 5. Effects on Grade 6 Applications and Enrollment

	RD (1)	Lottery (2)
	A: First Choice Grade 6 School	
Specialized HS Share	–0.002	0.071***
	(0.019)	(0.013)
N	13,100	2,582
First stage F	980.7	799.8
Non-G&T Mean	0.198	0.248
Value-Added	–0.001	0.033***
	(0.018)	(0.013)
N	13,100	2,582
First stage F	980.7	799.8
Non-G&T Mean	0.065	0.088
	B: Enrolled Grade 6 School	
Specialized HS Share	–0.026	0.090***
	(0.018)	(0.013)
N	10,402	2,100
First stage F	805.4	782.5
Non-G&T Mean	0.149	0.202
Value-Added	–0.003	0.021*
	(0.020)	(0.012)
N	10,402	2,100
First stage F	805.4	782.5
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 Specialized High School Outcomes

	RD (1)	Lottery (2)
SHSAT Tested	–0.039	0.016
	(0.062)	(0.033)
N	7,827	1,491
First stage F	615.1	838.6
Non-G&T Mean	0.731	0.802
Specialized HS Offer	0.012	–0.015
	(0.064)	(0.044)
N	7,827	1,491
First stage F	615.1	838.6
Non-G&T Mean	0.302	0.452

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 on an indicator 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%.

Table 7. Effects on Kindergarten Classroom Characteristics by FRPL Status

	RD		Lottery	
	FRPL	Non-FRPL	FRPL	Non-FRPL
	(1)	(2)	(3)	(4)
Class Size	–5.658***	–3.502***	–2.522**	–0.100
	(0.640)	(0.426)	(1.246)	(0.484)
N	4,945	15,909	651	3,398
First stage F	483	1021	130	882
Non-G&T Mean	22.163	22.382	23.161	23.573
p-value	[0.005]		[0.192]	
FRPL Share	–0.288***	–0.249***	–0.271***	–0.101***
	(0.036)	(0.030)	(0.046)	(0.016)
N	4,945	15,909	651	3,398
First stage F	483	1021	130	882
Non-G&T Mean	0.663	0.390	0.615	0.306
p-value	[0.397]		[0.000]	
White/Asian Share	0.104***	0.149***	0.159***	0.083***
	(0.038)	(0.025)	(0.043)	(0.013)
N	4,945	15,909	651	3,398
First stage F	483	1021	130	882
Non-G&T Mean	0.544	0.652	0.575	0.687
p-value	[0.320]		[0.071]	

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%.

Table 8. Effects on Achievement by FRPL Status

	RD		Lottery	
	FRPL	Non-FRPL	FRPL	Non-FRPL
	(1)	(2)	(3)	(4)
ELA	–0.087	0.106	–0.079	–0.011
	(0.100)	(0.073)	(0.115)	(0.044)
N (estimation)	13,281	36,805	1,764	7,948
N (students)	4,408	12,830	556	2,725
First stage F	63.2	128.3	22.3	149.7
Non-G&T Mean	0.891	1.060	1.084	1.234
p-value	[0.168]		[0.585]	
Math	–0.194*	–0.010	0.025	–0.047
	(0.109)	(0.077)	(0.119)	(0.048)
N (estimation)	10,766	30,338	1,449	6,648
N (students)	4,379	12,723	552	2,714
First stage F	90.2	190.4	28.9	206.6
Non-G&T Mean	0.964	1.107	1.173	1.326
p-value	[0.193]		[0.372]	
Algebra Regents	–0.081	0.139	0.222	–0.043
	(0.184)	(0.121)	(0.145)	(0.053)
N (estimation)	1,080	3,227	139	779
First stage F	104.0	258.3	105.4	573.9
Non-G&T Mean	0.880	1.060	0.899	1.197
p-value	[0.771]		[0.710]	

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%.

Table 9. Effects on Grade 6 Applications and Enrollment by FRPL Status

	RD		Lottery	
	FRPL	Non-FRPL	FRPL	Non-FRPL
	(1)	(2)	(3)	(4)
		A: First Choice Grade 6 School		
Specialized HS Share	0.002	–0.003	0.119***	0.064***
	(0.032)	(0.024)	(0.040)	(0.013)
N	3,317	9,756	459	2,121
First stage F	324.9	645.7	111.0	688.8
Non-G&T Mean	0.184	0.203	0.233	0.252
p-value	[0.865]		[0.201]	
Value-Added	0.028	–0.013	0.082**	0.026*
	(0.030)	(0.023)	(0.033)	(0.013)
N	3,317	9,756	459	2,121
First stage F	324.9	645.7	111.0	688.8
Non-G&T Mean	0.071	0.063	0.082	0.089
p-value	[0.258]		[0.129]	
		B: Enrolled Grade 6 School		
Specialized HS Share	–0.073**	–0.007	0.120***	0.081***
	(0.029)	(0.022)	(0.031)	(0.014)
N	2,750	7,644	390	1,710
First stage F	264.1	523.8	119.6	667.5
Non-G&T Mean	0.125	0.158	0.163	0.211
p-value	[0.069]		[0.261]	
Value-Added	–0.029	0.007	0.074**	0.009
	(0.034)	(0.025)	(0.030)	(0.013)
N	2,750	7,644	390	1,710
First stage F	264.1	523.8	119.6	667.5
Non-G&T Mean	0.044	0.054	0.055	0.086
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 10. Effects on SHSAT Outcomes by FRPL Status

	RD		Lottery	
	FRPL	Non-FRPL	FRPL	Non-FRPL
	(1)	(2)	(3)	(4)
SHSAT Tested	–0.142	0.022	0.183**	–0.017
	(0.098)	(0.080)	(0.088)	(0.035)
N	2,205	5,621	290	1,201
First stage F	232.7	377.8	100.9	734.0
Non-G&T Mean	0.691	0.746	0.742	0.816
p-value	[0.193]		[0.038]	
Specialized HS Offer	–0.041	0.044	0.119	–0.040
	(0.087)	(0.086)	(0.112)	(0.048)
N	2,205	5,621	290	1,201
First stage F	232.7	377.8	100.9	734.0
Non-G&T Mean	0.213	0.335	0.314	0.485
p-value	[0.485]		[0.201]	

Notes: This table reports 2SLS estimates of the effect of G&T enrollment on high school outcomes, separately by FRPL status. See Table 6 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 11. Multisector Model Estimates

	Achievement		Middle school enrollment		High school	
	ELA	Math	Specialized HS Share	Value-Added	Took SHSAT	Specialized HS Offer
	(1)	(2)	(3)	(4)	(5)	(6)
A: All Applicants						
District	0.016	–0.028	0.039**	–0.005	0.036	0.003
	(0.052)	(0.055)	(0.016)	(0.016)	(0.041)	(0.058)
Citywide	–0.050	–0.119***	0.159***	0.015	0.067**	0.076*
	(0.042)	(0.044)	(0.011)	(0.012)	(0.033)	(0.045)
N (estimation)	14,967	12,554	3,292	3,292	2,488	2,488
N (students)	4,828	4,801	3,292	3,292	2,488	2,488
First stage F	63.0	93.0	303	303	282	282
B: FRPL						
District	–0.101	–0.086	0.138***	0.101**	0.092	0.124
	(0.162)	(0.151)	(0.040)	(0.043)	(0.093)	(0.138)
Citywide	–0.210	–0.210	0.268***	0.110***	0.199**	0.231*
	(0.144)	(0.136)	(0.033)	(0.037)	(0.090)	(0.127)
N (estimation)	2,533	2,107	568	568	481	481
N (students)	758	753	568	568	481	481
First stage F	9.814	13.0	52.0	52.0	66.0	66.0
C: Non-FRPL						
District	0.028	–0.022	0.021	–0.021	0.028	–0.017
	(0.055)	(0.059)	(0.017)	(0.017)	(0.046)	(0.065)
Citywide	–0.035	–0.106**	0.143***	0.003	0.036	0.041
	(0.043)	(0.045)	(0.012)	(0.013)	(0.035)	(0.048)
N (estimation)	12,430	10,443	2,724	2,724	2,007	2,007
N (students)	4,068	4,046	2,724	2,724	2,007	2,007
First stage F	54.0	80.0	251	251	220	220

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. Robust standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 12. Estimates of ATT for Grade 6 Outcomes

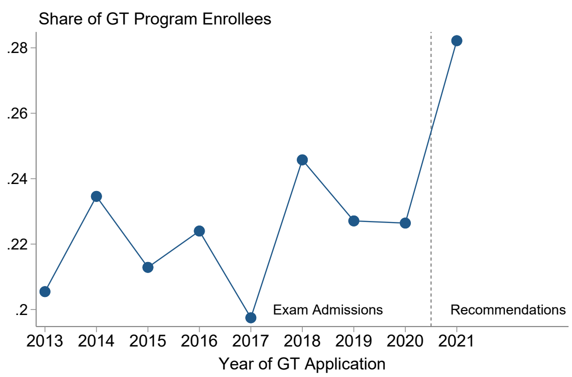
	FRPL Weighting (1)	FRPL and Exam Score Weighting: Linear Score Effect (2)	FRPL and Exam Score Weighting: Two Score Bins (3)	FRPL and Exam Score Weighting: Four Score Bins (4)
A: Application, Specialized HS Share				
Exam	0.077***	0.046***	0.062***	0.044***
	(0.014)	(0.008)	(0.008)	(0.013)
Recommendation	0.080***	–0.050	–0.008	–0.115
	(0.015)	(0.045)	(0.037)	(0.070)
B: Application, Value-Added				
Exam	0.039***	0.017**	0.029***	0.008
	(0.013)	(0.008)	(0.007)	(0.012)
Recommendation	0.042***	–0.052	–0.010	–0.072
	(0.013)	(0.041)	(0.034)	(0.058)
C: Enroll, Specialized HS Share				
Exam	0.090***	0.068***	0.081***	0.049***
	(0.013)	(0.007)	(0.007)	(0.010)
Recommendation	0.092***	–0.009	0.013	–0.003
	(0.013)	(0.034)	(0.029)	(0.042)
D: Enroll, Value-Added				
Exam	0.024*	0.012	0.018***	–0.000
	(0.012)	(0.007)	(0.007)	(0.011)
Recommendation	0.028**	–0.032	–0.023	–0.007
	(0.013)	(0.040)	(0.032)	(0.052)

Notes: This table reports estimates of the average treatment effect on the treated (ATT) for G&T. Panels A and B report ATTs on the characteristics of schools that students ranked first on their middle school application. Panels C and D report ATTs on the characteristics of schools that students enrolled in for 6th grade. See Table 5 for a description of the outcomes. Each panel reports estimates of the ATT under exam- and teacher recommendation-based admissions. Column 1 reports estimates obtained by reweighting FRPL-specific treatment effects, and columns 2-4 report estimates obtained by reweighting conditional LATEs defined by FRPL and entrance exam scores. The procedure is described in Section 6. Results in column 2 rely on effect estimates obtained from a 2SLS procedure with two endogenous variables, enrollment and the interaction of enrollment with the exam score, and two instruments, offers and the interaction of offers with the exam score. The models are estimated separately by FRPL status. Specifications used for columns 3-4 estimate LATEs by FRPL status and bins of the entrance exam score. The two bin specification groups students by scores of 94 and below and 95 and above. The four bin specification groups students by scores 92 and below, 93-95, 96-98, and 99. G&T students in the recommendations period who scored below 90 are assigned the treatment effect estimate belonging to the lowest scoring group. * significant at 10%; ** significant at 5%; *** significant at 1%

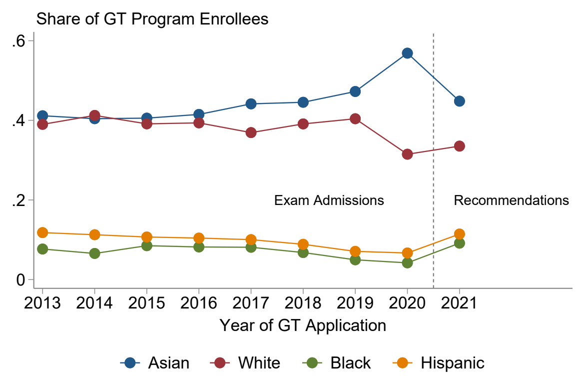
Figures

Figure 1. Trends in G&T Enrollment Shares

A. FRPL

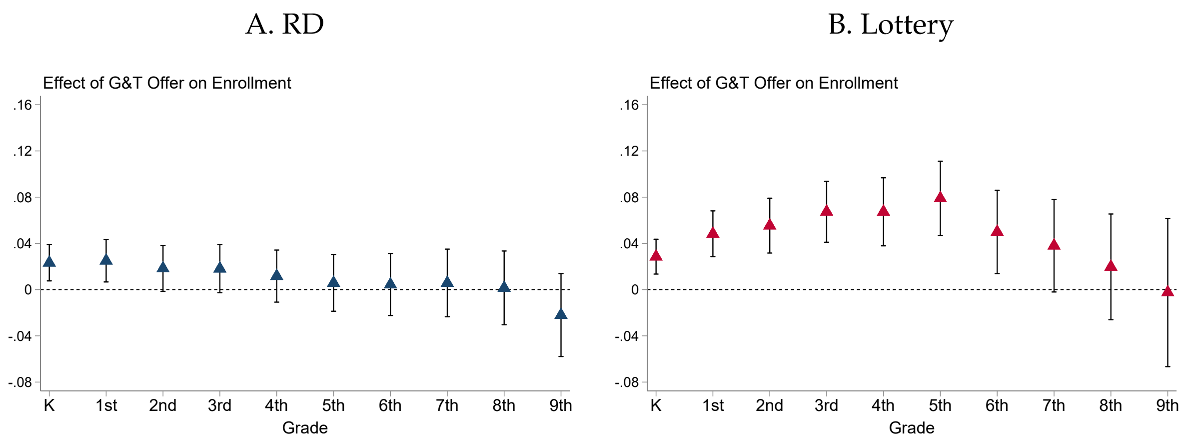


B. Race and Ethnicity



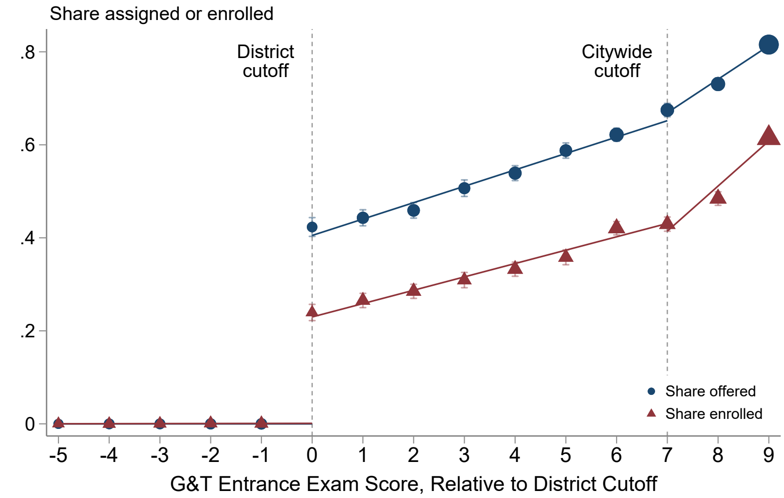
Notes: This figures depicts demographic shares of kindergarten G&T students over time. Panel A plots the share of students who FRPL-eligible. Panel B plots the shares of students by race and ethnicity. Shares are computed in the sample of students with non-missing demographic data.

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



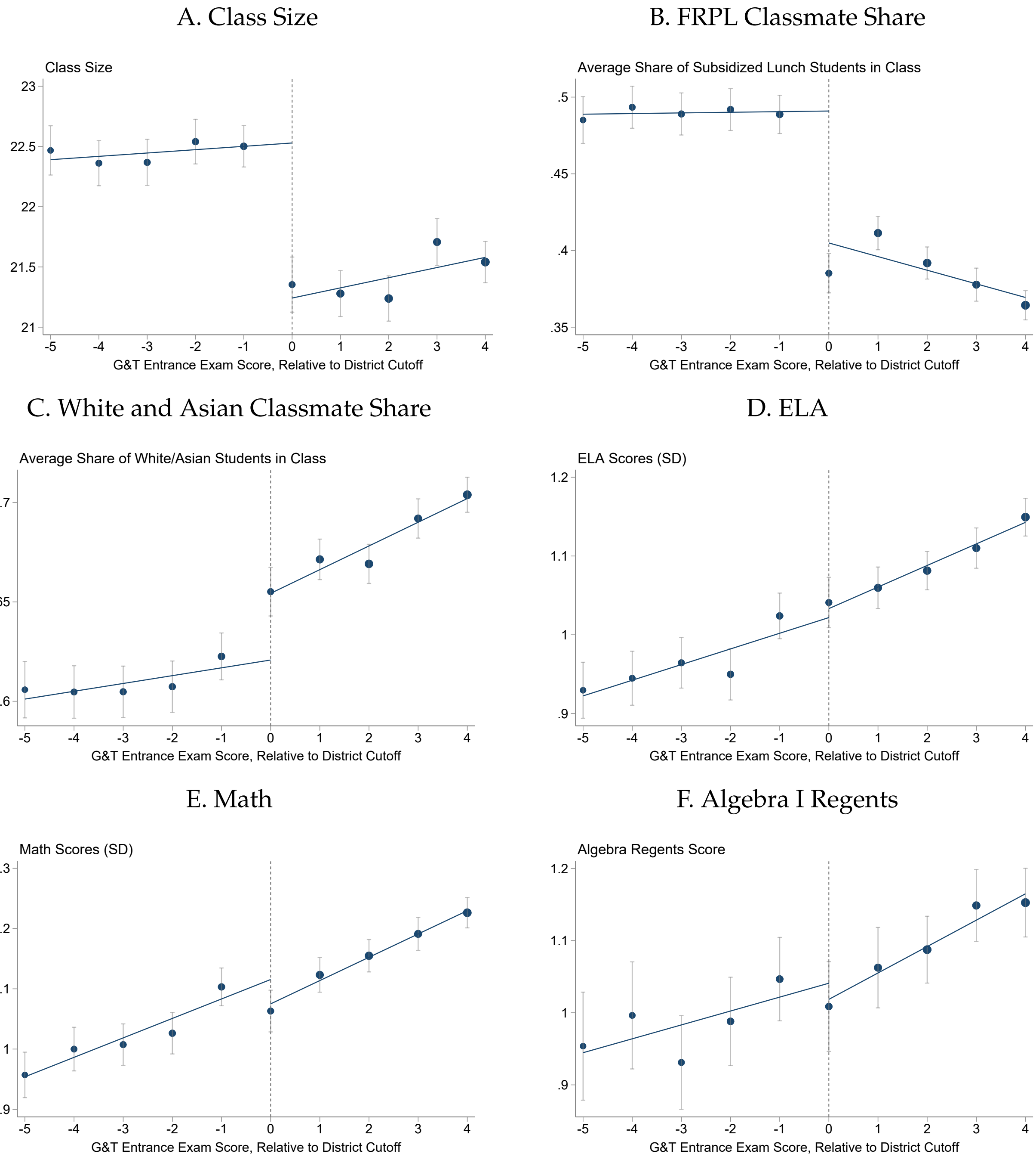
Notes: This figure depicts effect estimates of the impact of G&T offers on enrollment in a non-charter NYC public school. Panel A plots estimates obtained from a fuzzy RD procedure that instruments offer with eligibility. Panel B plots estimates from the reduced form effect of offers in the lottery IV 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.

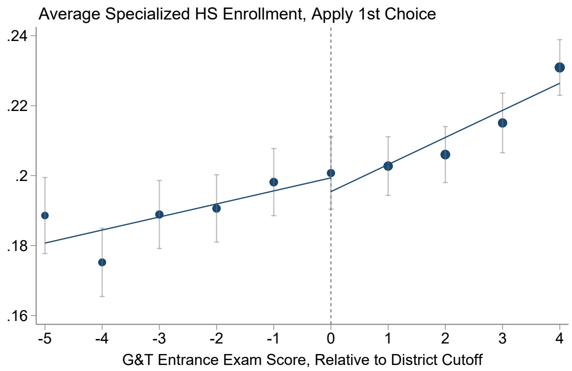
Figure 4. RD Effects on Kindergarten Classrooms and Achievement



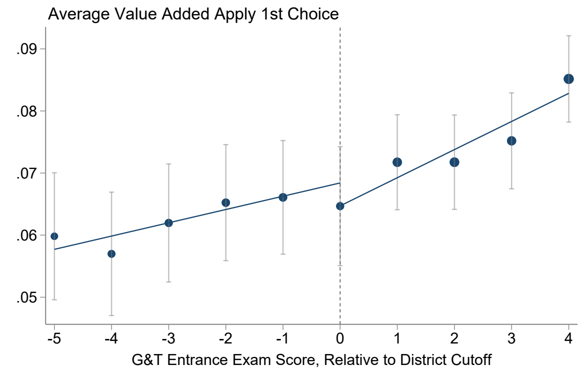
Notes: This figure depicts the reduced form impacts of G&T eligibility on kindergarten classroom characteristics and test scores in grades 3-8. Panel A depicts the impact on kindergarten class size. Panels B and C depict the impact on the share of classmates who are FRPL-eligible and who are White and 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

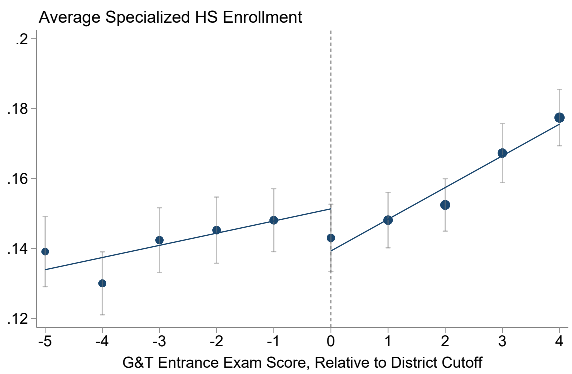
A. Specialized HS Share of First Choice Middle School



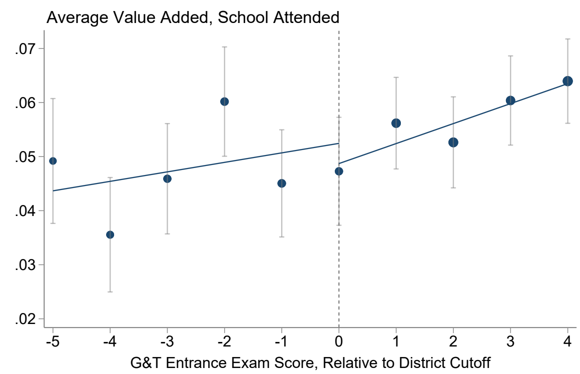
B. Value-Added of First Choice Middle School



C. Specialized HS Share of Enrolled Middle School



D. Value-Added of Enrolled Middle School

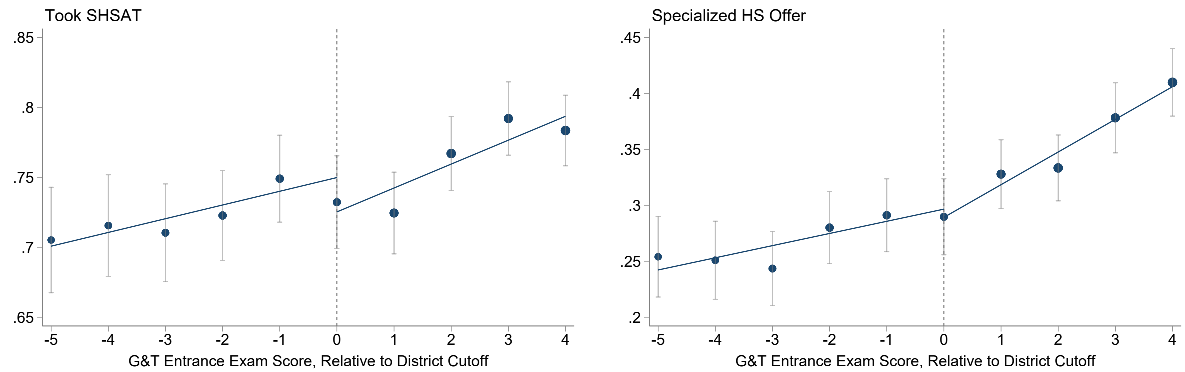


Notes: This figure depicts the reduced form impact of kindergarten G&T eligibility on middle school application and enrollment outcomes. The outcome in Panel A is the specialized high school share of the school ranked as a student's first choice on their middle school application. The outcome in Panel B is estimated value-added of the first choice school on 6th grade math scores. 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. RD Effects on High School Outcomes

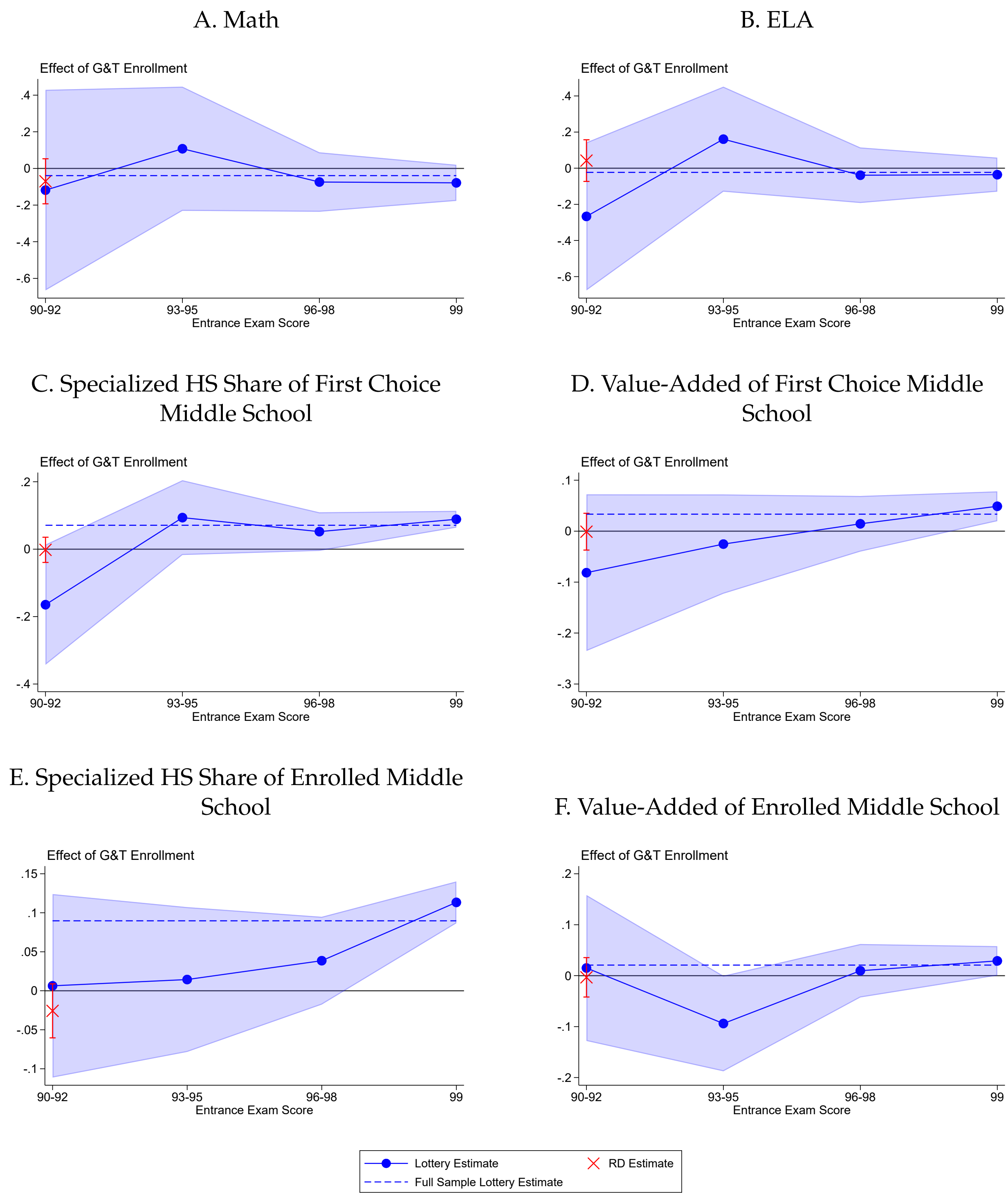
A. Took SHSAT

B. Specialized HS Offer



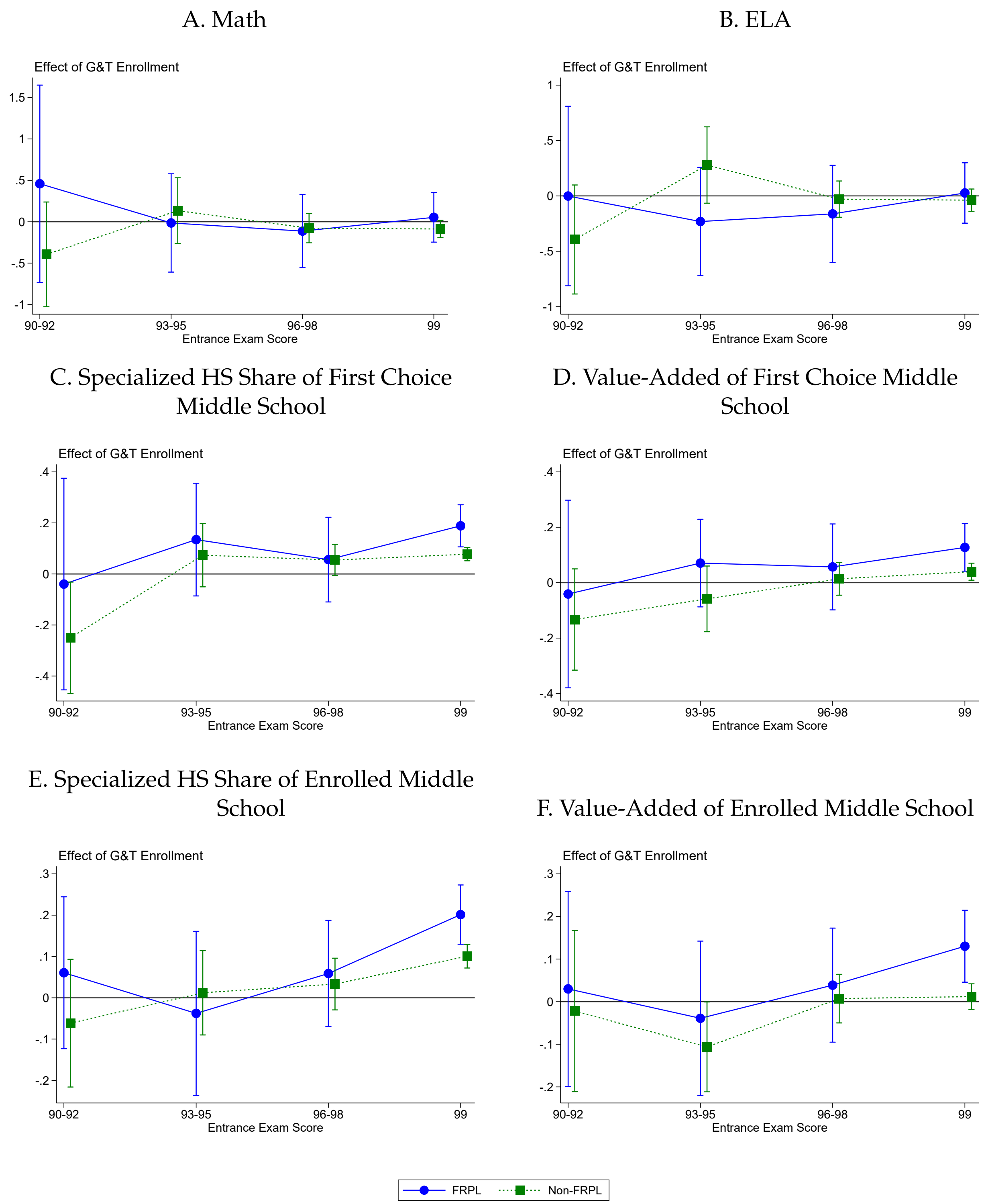
Notes: This figure depicts the reduced form impact of kindergarten G&T eligibility on high school outcomes. The outcome in Panel A is an indicator for taking the Specialized High School Admissions Test (SHSAT). The outcome in Panel B is an indicator for receiving an offer to a specialized high school. Marker sizes are proportional to the number of students attaining each score value. Whiskers depict 95-percent confidence intervals for the conditional means.

Figure 7. 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.

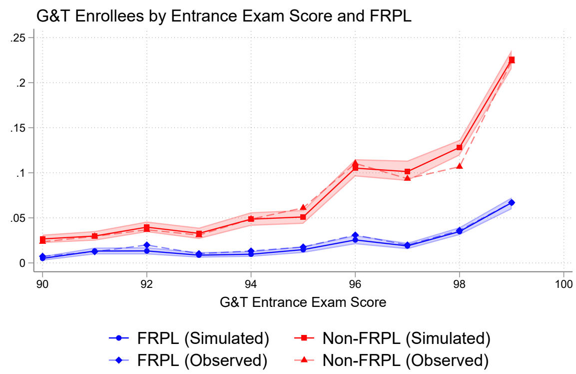
Figure 8. Effects by Score and FRPL Status



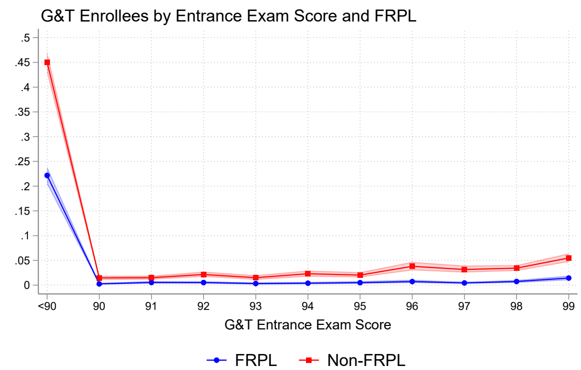
Notes: These figures depict lottery 2SLS estimates of the effects of G&T enrollment by FRPL status and bins of the entrance exam score. The blue line corresponds to FRPL students, and the green line corresponds to non-FRPL students. The whiskers show the 95-percent confidence intervals.

Figure 9. Simulated G&T Enrollment

A. Test Admissions, 2019

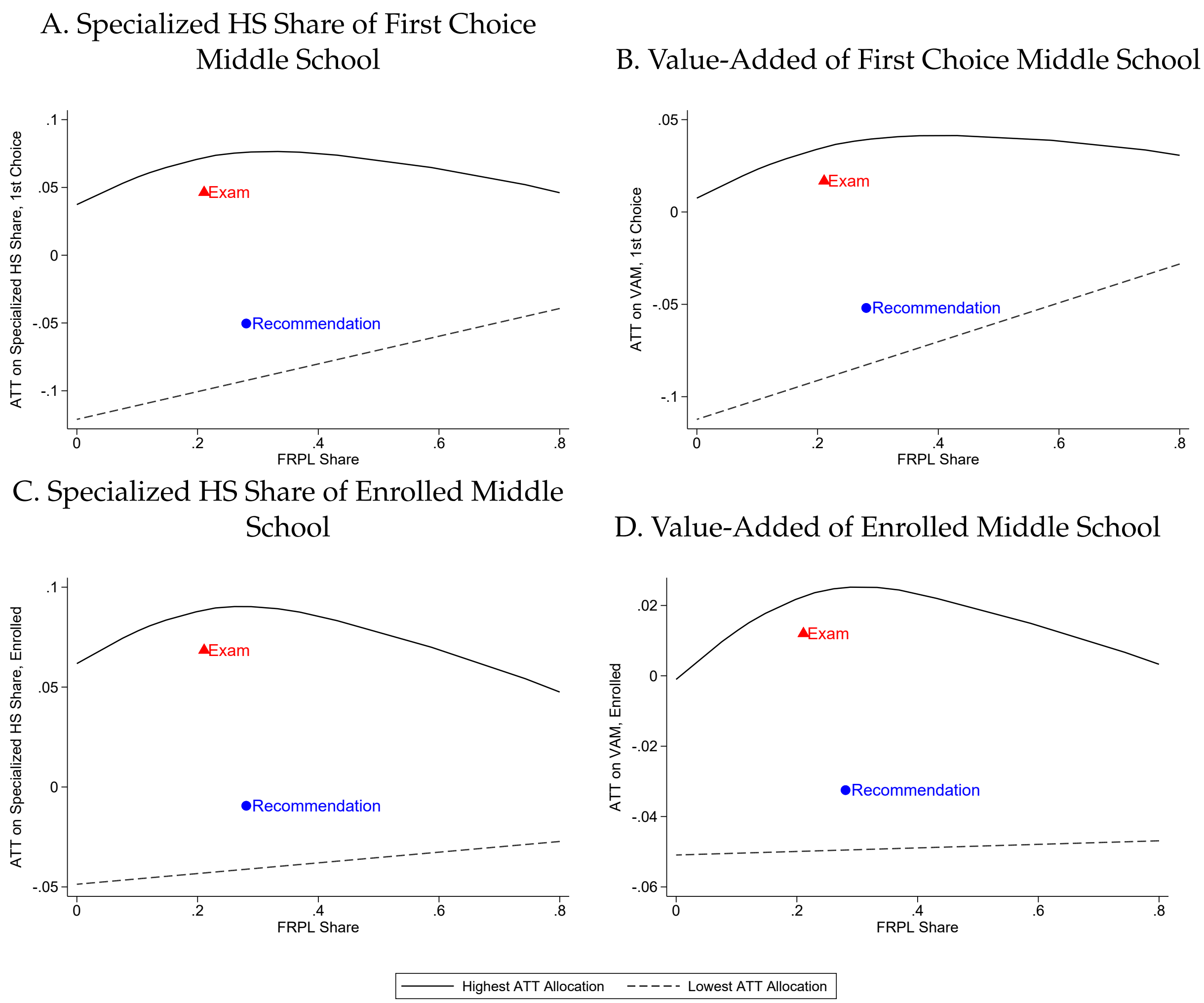


B. Simulated Recommendations, 2019



Notes: Panel A depicts the observed and simulated share of G&T students by entrance exam score and FRPL status in 2019. 95 percent confidence intervals for the simulated shares are shown by the shaded area. Prediction intervals are computed by simulating parameters from their sampling distributions. Panel B depicts simulated enrollment shares under a recommendation system. See Section 6 for details.

Figure 10. Policy Possibility Frontiers



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 12. The red triangle and circle correspond to simulated outcomes under exam- and teacher recommendation-based admissions.

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 want to apply to both G&T and non-G&T submit separate ROLs over their program preferences, 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 grade 5, while all of the citywide programs except for NEST+M end in grade 8. NEST+M continues through grade 12.

Interviews with NYC G&T teachers suggest that while the G&T curriculum is largely the same as the standard curriculum, more of the 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, and teachers may have a Gifted Education Extension Certificate, but this is not a requirement for teaching in the program. Most G&T programs have kindergarten as the main entry point, but some programs do not have entry points until grade 3.

A.3 G&T 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.[60]

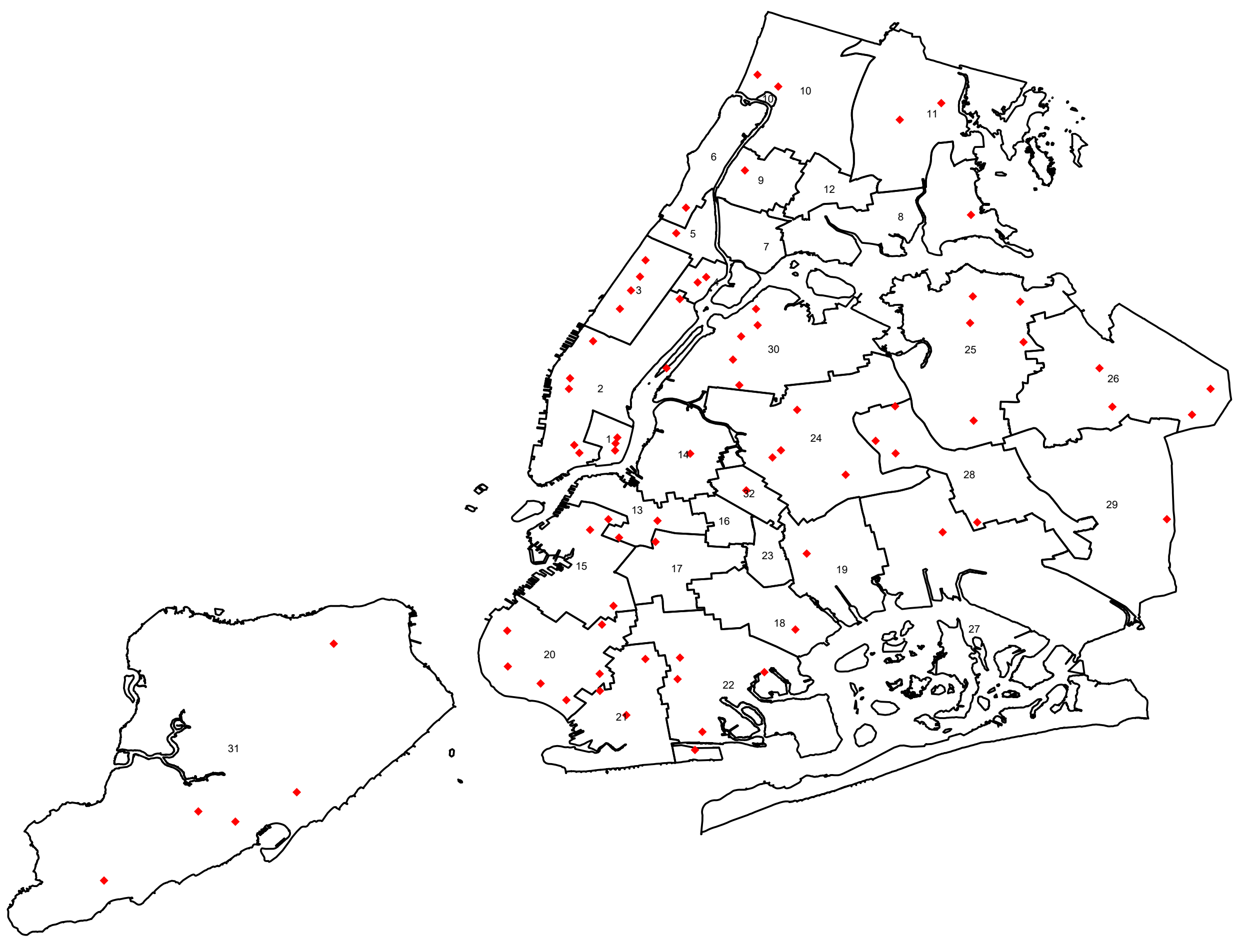
A.4 G&T Recommendation Admissions

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.

[60] For LAUSD, see the Gifted Identification instructions here: `https://gate.lausd.org/`.

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.")

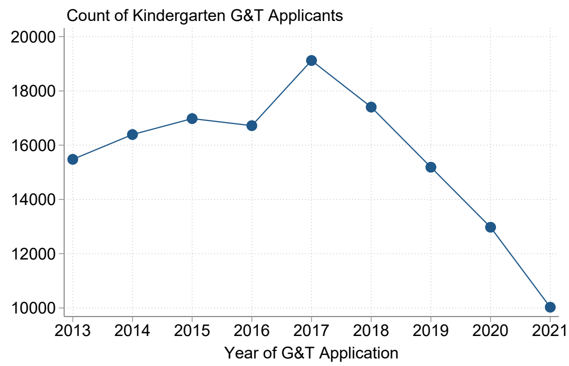
Appendix Figure A1. Kindergarten G&T Locations, 2020-21



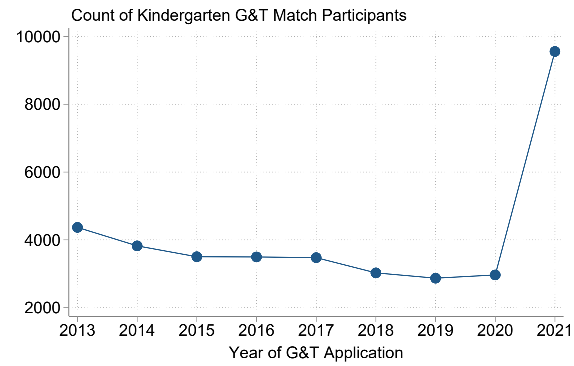
Notes: This figure depicts the locations of kindergarten G&T programs, including both district and citywide programs, during the 2020-2021 school year. The numbers indicate each of the 32 school districts comprising the NYC public school system.

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

A. Applied



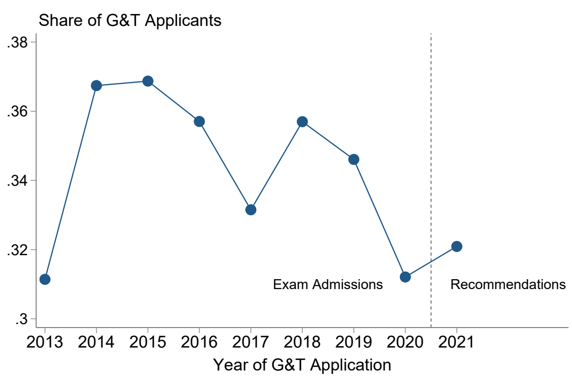
B. Participated in G&T Match



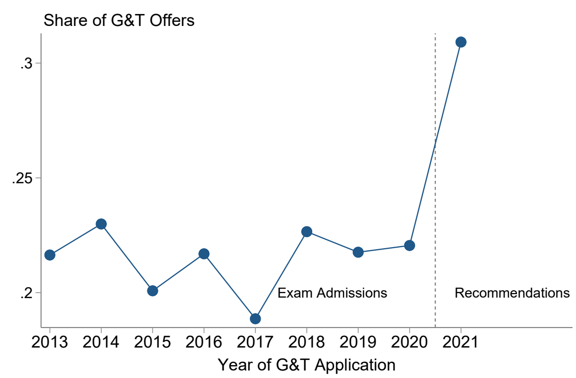
Notes: Panel A shows the number of students who submitted an application for kindergarten G&T by year. Panel B shows the number of gifted-eligible students who participated in the match. In 2020 and earlier, applicants who scored 90 or higher on an entrance exam were eligible for G&T. In 2021 and later, eligibility was determined by teacher recommendations and staff interviews.

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

A. Share of Applicants



B. Share of 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.

Appendix Table A1. Changes in G&T Demographics

	2016-2020 Mean (1)	2021 Mean (2)	2021 Difference from 2016-2020 Mean (3)
FRPL	0.224	0.282	0.058
Female	0.505	0.513	0.008
Black	0.065	0.091	0.027
Hispanic	0.086	0.114	0.028
Asian	0.469	0.448	–0.020
White	0.375	0.335	–0.039

Notes: This table reports means of demographic characteristics among kindergarten G&T students. Column 1 reports means computed across years 2016–2020. Column 2 reports means for 2021, the first year of recommendation admissions. Column 3 reports the difference between columns 1 and 2. Means are computed only using students who have non-missing demographics.

B Data Appendix

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.

From cohorts starting kindergarten in 2015-16 onward, we have complete sibling priority data on the schools where an applicant has priority. For the 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 (OVB). 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

The Office of Student Performance and Accountability 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 (scored 90 or above on the entrance exam). We have data on the classrooms officially identified as G&T beginning in 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 use files on school and student residence location to compute distance. For each student with an active address record, we have data on their census tract of residence from the district's "Zoned DBN" dataset. Our measure of distance is the great circle distance between the school and the centroid of the student's residential tract (in miles). We use the latitude and longitude of each and compute the distance using the Stata `geodist` function.

Achievement: Grades 3-8 State Assessments and Algebra I Regents

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.

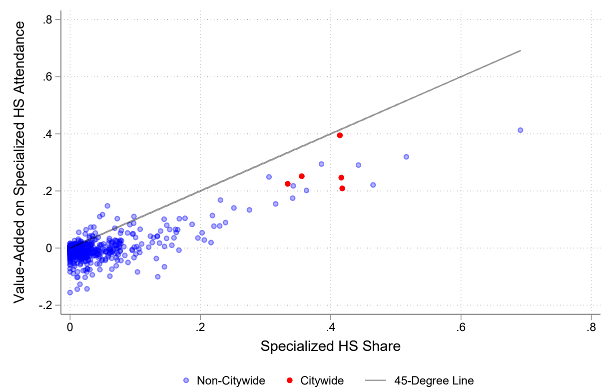
Middle School Applications and Assignments

We obtain data on middle school applications and assignments for school years 2016-2017 through 2024-2025 from the Student Enrollment Office. These data are generated by centralized assignment for middle school admissions and record students' rank-ordered preferences over middle school 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.

Our middle school outcomes focus primary on two characteristics. The first, which we call the specialized high school share, is the share of 6th grade students at the school who end up enrolling in a specialized high school in 9th grade, taken as an average across all students who were in grade 6 between 2016 and 2019; we start with 2016 because this is the first year that students in our application sample would begin grade 6. The second characteristic is the estimated value-added, based on 6th grade state math exams during the same years; Appendix C.2 includes more details on the value-added construction. We end the average in 2019 because state exams were not conducted during the COVID-19 pandemic in spring 2020 or spring 2021.

While the specialized high school share on its own may not represent a causal effect of school attendance on future specialized high school attendance, we show suggestive evidence that it is correlated with causal school effects. To do so, we use the same cohorts of students (those entering grade 6 in 2016-2019) and regress an indicator for future specialized high school attendance, middle school fixed effects for grade 6, grade 5 state test scores, and interactions between year and demographics. Figure Appendix Figure B1 below shows the relationship between this "value-added" fixed effect of each middle school and the observed specialized high school share level. The line of best fit has a slope of 0.550 (SE $= 0.016$), suggesting a strong relationship between causal estimates and observed levels.

Appendix Figure B1. Specialized HS Shares and Value-Added



Notes: This scatterplot depicts the relationship between the specialized high school share outcome and estimated value-added on specialized high school enrollment. Each dot corresponds to a G&T program. The 45-degree line is shown in black.

Specialized High School Admissions Test (SHSAT)

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 Grade 8, are coded as not receiving an offer to a specialized high school.

C Methodological Appendix

C.1 Attrition Specifications

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

$$Z_i = \pi 1(R_i \geq 90) + \lambda_1 R_i + \gamma_1 R_i 1(R_i \geq 90) + \nu_i$$
$$Y_i = \beta Z_i + \lambda_2 R_i + \gamma_2 R_i 1(R_i \geq 90) + \varepsilon_i,$$

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:

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

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. (2024a) shows 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 F3, 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. (2024a) show that estimates obtained in this way 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. (2023a) 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 $K \times 1$ 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

$$
Y_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{iG} \end{bmatrix}, \quad D_i = \begin{bmatrix} D_{i1} \\ D_{i2} \\ \vdots \\ D_{iG} \end{bmatrix}, \quad X_i = \begin{bmatrix} X_{i1}' & 0 & \cdots & 0 \\ 0 & X_{i2}' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_{iG}' \end{bmatrix}, \quad \varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iG} \end{bmatrix}.
$$

We also have a matrix of instruments given by

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

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

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

where Y, D, ε, and ν are $NG \times 1$ vectors, Π is a $G \times 1$ vector, X is a $NG \times KG$ matrix, and Z is a $NG \times G$ matrix. The control coefficients Γ and Ψ form $KG \times 1$ vectors, but we restrict the causal effect τ to be common across grades.

We estimate this system by 2SLS, clustering standard errors at the student level. 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 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 $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$. In addition to the standard IV assumptions, we require H_i to be conditionally ignorable.

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

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

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 $H_i = 1$, converges to 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

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

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

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

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.6 Reweighting on Observables for the ATT

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:

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

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)**.**

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

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)**.**

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

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

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

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

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

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

Then by Bayes' theorem,

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

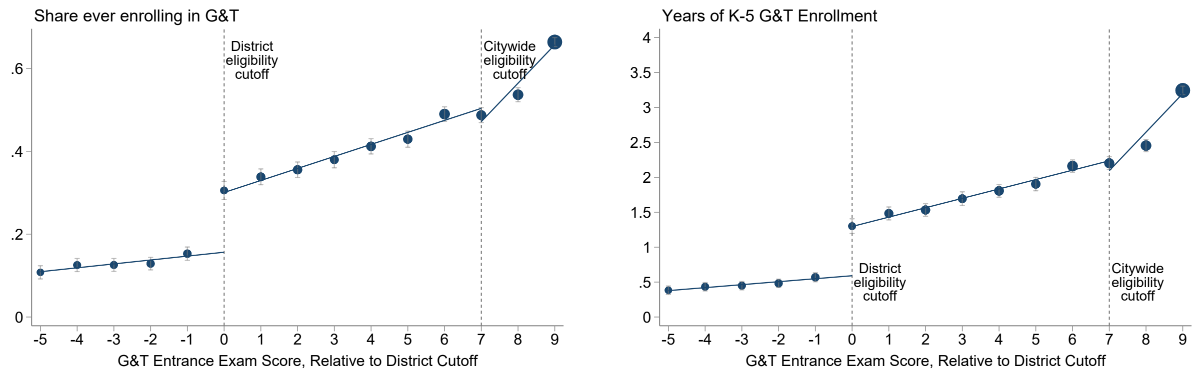
□

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

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

A. Effect on Any K-5 G&T Enrollment

B. Effect on Years of K-5 G&T Enrollment

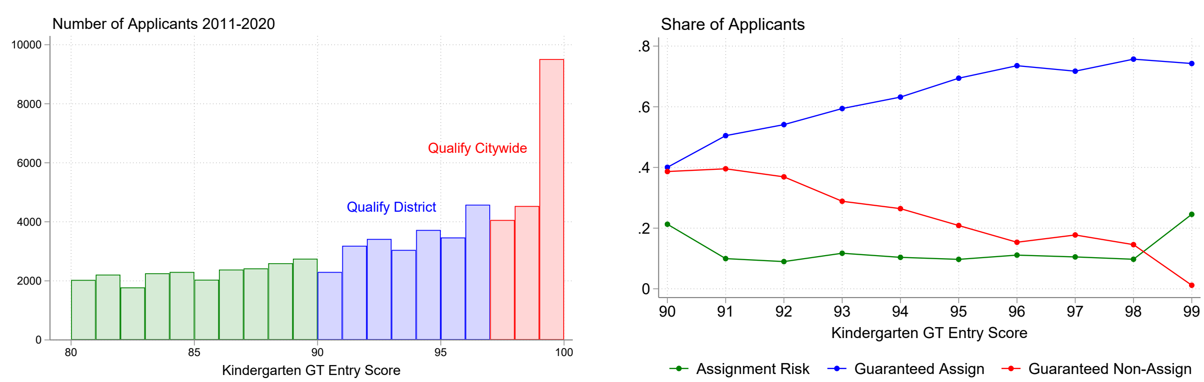


Notes: The figures depict the effect of kindergarten G&T eligibility on alternate first-stage outcomes. Panel A depicts the effect on ever enrolling in a G&T program between grades K-5. Panel B depicts the effect on the number of years enrolling in a G&T program between grades K-5. The sample only includes 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. Distribution of Kindergarten G&T Scores

B. Distribution of Kindergarten G&T 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

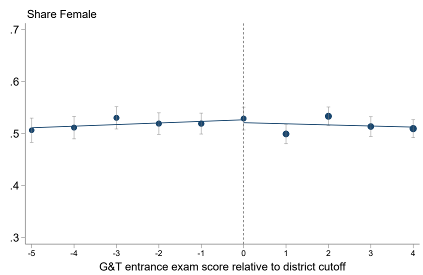
Appendix Table D1. Covariate Balance

	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	–0.107***	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	0.006	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	–0.016	–0.007	–0.148***	–0.002	0.009
	(0.012)	(0.018)	(0.003)	(0.013)	(0.018)
Disability	–0.004	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	24,194	10,570	141,909	4,322	2,127
Joint test					
$\chi^2(8)$	14.4	10.1	7,392	11.4	6.09
p-value	[0.071]	[0.260]	[0.000]	[0.178]	[0.637]

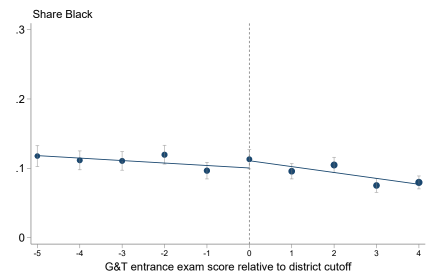
Notes: This table reports statistical tests for balance. Column 1 reports estimates from regressions of baseline demographic variables on the RD district eligibility instrument with RD controls. The RD model imposes linear fits for the conditional means and a bandwidth of length five. Column 2 reports 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. 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

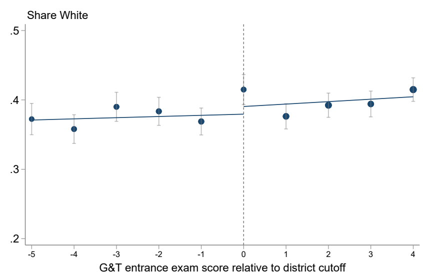
A. Female



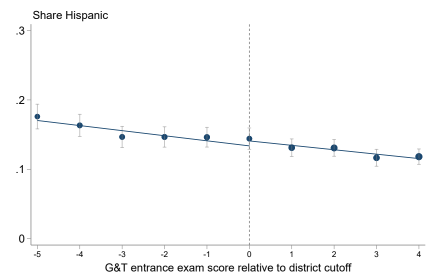
B. Black



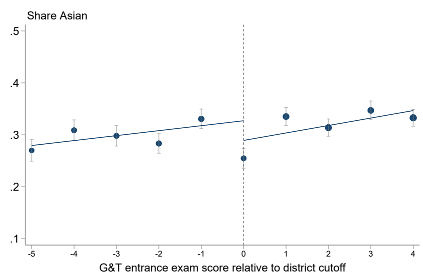
C. White



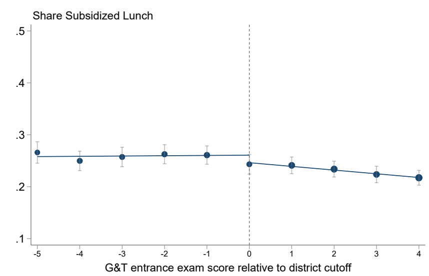
D. Hispanic



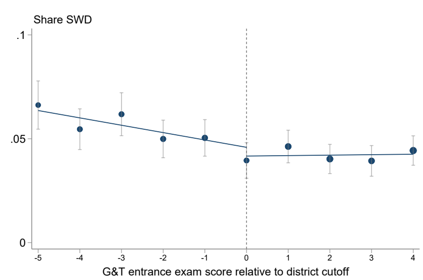
E. Asian



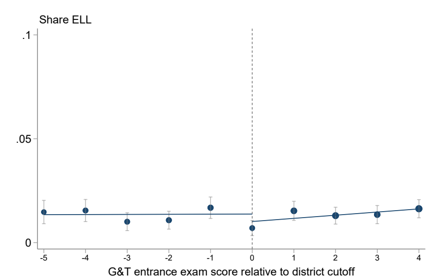
F. FRPL



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 (Pooled Treated and Untreated)

	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	0.564	0.496
	(0.025)	(0.019)	(0.101)	(0.018)
Black	0.140	0.030	–0.012	0.034
	(0.015)	(0.007)	(0.054)	(0.006)
White	0.271	0.502	0.506	0.502
	(0.024)	(0.019)	(0.100)	(0.018)
Hispanic	0.164	0.081	0.214	0.068
	(0.017)	(0.009)	(0.064)	(0.008)
Asian	0.402	0.259	0.428	0.244
	(0.023)	(0.017)	(0.095)	(0.015)
Other/Missing Race	0.023	0.128	–0.136	0.152
	(0.020)	(0.014)	(0.079)	(0.014)
FRPL	0.353	0.131	0.411	0.106
	(0.022)	(0.013)	(0.091)	(0.011)
N	37286	5004	1395	3609

Notes: This table reports estimates of complier covariate means. Sample sizes show the total number of observations in the estimation, including both compliers and non-compliers. Column 1 reports estimates for RD compliers. Column 2 reports estimates for compliers in the lottery sample. 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 instrumenting an indicator for untreated status with either the eligibility or lottery instrument, with interactions of the untreated dummy and the baseline covariate as the outcome. Robust standard errors are shown in parentheses. The lottery sample within the RD bandwidth includes lottery sample applicants with entrance exam scores 90-94. The lottery sample above the RD bandwidth includes lottery sample applicants with entrance exam scores 95-99.

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

Appendix Table E1. Lottery Estimates with Inverse Probability Weighting

	Full sample (1)	FRPL (2)	Non-FRPL (3)
Math	–0.048	0.020	–0.055
	(0.047)	(0.129)	(0.049)
First stage F	198	28.4	201
Over-id p-value	0.626	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	0.420	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	0.089***	0.124***	0.082***
	(0.013)	(0.034)	(0.016)
First stage F	672	112	598
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
	(0.013)	(0.035)	(0.015)
First stage F	672	112	598
Non-G&T Mean	0.080	0.055	0.086
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.5 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. Regression Discontinuity Robustness, Grade 6 Outcomes

	Baseline	Quadratic, BW7	BW7	BW6	BW4	BW3	Uncon-trolled	Honest RD	K-3 Pooled	K-3 Over-ID
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ELA, Grade 6	0.118	0.100	0.171*	0.188**	0.132	0.072	0.112	0.088	0.010	0.099
	(0.105)	(0.141)	(0.089)	(0.094)	(0.119)	(0.140)	(0.109)	(0.173)	(0.118)	(0.099)
N	7,235	10,361	10,361	8,711	5,748	4,333	7,235	5,030	15,780	15,780
Math, Grade 6	0.036	–0.073	0.147	0.108	0.066	–0.044	–0.013	–0.275	–0.036	0.023
	(0.112)	(0.150)	(0.094)	(0.099)	(0.126)	(0.149)	(0.121)	(0.291)	(0.119)	(0.104)
N	7,190	10,301	10,301	8,656	5,709	4,307	7,190	3,193	15,710	15,710
Specialized HS Share, 1st Choice	–0.002	–0.013	0.011	–0.001	–0.009	0.010	–0.014	–0.016	–0.006	0.005
	(0.019)	(0.026)	(0.016)	(0.017)	(0.022)	(0.026)	(0.020)	(0.082)	(0.020)	(0.018)
N	13,100	18,978	18,978	15,800	10,376	7,842	13,100	6,450	25,073	25,073
Value-Added, 1st Choice	–0.001	–0.007	0.011	0.000	–0.001	0.008	–0.011	–0.007	0.002	0.005
	(0.018)	(0.025)	(0.016)	(0.017)	(0.021)	(0.025)	(0.019)	(0.072)	(0.019)	(0.017)
N	13,100	18,978	18,978	15,800	10,376	7,842	13,100	6,458	25,073	25,073
Specialized HS Share, Enrolled	–0.026	–0.035	–0.007	–0.020	–0.031	–0.010	–0.036*	–0.024	–0.033*	–0.022
	(0.018)	(0.024)	(0.015)	(0.016)	(0.020)	(0.024)	(0.018)	(0.100)	(0.019)	(0.016)
N	10,402	15,004	15,004	12,506	8,264	6,251	10,402	5,338	20,974	20,974
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)	(0.022)	(0.027)	(0.020)	(0.062)	(0.021)	(0.018)
N	10,402	15,004	15,004	12,506	8,264	6,251	10,402	5,160	20,974	20,974

Notes: This table reports RD estimates obtained from a variety of alternative specifications. Column 1 reproduces the baseline results. Column 2 adds quadratic running variable controls to the model 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 and Stata package of 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 a linearity in the first-stage. Column 9 reports a pooled specification using applicants to all 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	–0.079	–0.188*	–0.042	0.351***	0.401***	0.328***
	(0.061)	(0.109)	(0.074)	(0.042)	(0.108)	(0.044)
N	10,402	2,750	7,644	2,100	390	1,710
First stage F	805.4	264.1	523.8	782.5	119.6	667.5
Non-G&T Mean	0.561	0.418	0.611	0.749	0.581	0.788
p-value		[0.254]			[0.540]	
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	6,844	1,904	4,939	1,367	264	1,103
First stage F	582.5	214.8	362.2	824.2	100.1	712.8
Non-G&T Mean	0.147	0.121	0.156	0.199	0.161	0.208
p-value		[0.219]			[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 reports 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 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

Appendix Table F1. OLS Effects

	Uncontrolled (1)	Controlled (2)
Math	0.634***	0.161***
	(0.006)	(0.007)
N (estimation)	265,040	265,040
N (students)	97,102	97,102
ELA	0.563***	0.043***
	(0.006)	(0.007)
N (estimation)	278,534	278,534
N (students)	97,358	97,358
Specialized HS Share, Enrolled Middle School	0.110***	0.023***
	(0.002)	(0.002)
N	57,406	57,406
Value-Added, Enrolled Middle School	0.055***	0.009***
	(0.002)	(0.002)
N	57,406	57,406

Notes: This table reports OLS estimates of the effect of G&T enrollment. The first column reports estimates from a regression on the outcome, 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 having a missing value for the G&T exam. See Tables 4, 5, and 6 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	ELA	Math	ELA
	(1)	(2)	(3)	(4)
Grade 3	–0.013	0.043	–0.036	0.040
	(0.074)	(0.076)	(0.052)	(0.052)
	[1.034]	[1.008]	[1.258]	[1.218]
Grade 4	–0.162**	–0.028	–0.079	–0.026
	(0.078)	(0.081)	(0.053)	(0.058)
	[1.082]	[1.034]	[1.323]	[1.234]
Grade 5	–0.113	0.033	0.029	0.010
	(0.091)	(0.094)	(0.068)	(0.070)
	[1.055]	[1.012]	[1.305]	[1.213]
Grade 6	0.036	0.118	–0.073	–0.017
	(0.112)	(0.105)	(0.097)	(0.099)
	[1.138]	[1.024]	[1.328]	[1.188]
Grade 7		0.172		–0.117
		(0.117)		(0.092)
		[1.040]		[1.199]
Grade 8		0.040		–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 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	4,917	1,489	3,428	1,038	206	832
First stage F	412.7	154.1	253.0	649.9	69.0	571.6
Non-G&T Mean	0.303	0.214	0.340	0.444	0.282	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	4,949	1,501	3,447	1,042	205	837
First stage F	418.7	158.1	256.8	663.7	69.5	586.4
Non-G&T Mean	0.329	0.254	0.361	0.451	0.343	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 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 F4. Effects on Achievement by Race/Ethnicity

	RD		Lottery	
	Black/Hispanic (1)	White/Asian (2)	Black/Hispanic (3)	White/Asian (4)
ELA	–0.008	0.006	0.013	–0.385**
	(0.071)	(0.109)	(0.044)	(0.160)
N (estimation)	35,921	11,195	7,747	1,285
N (students)	12,292	3,901	2,607	431
First stage F	126	61.0	145	15.9
Non-G&T Mean	1.079	0.805	1.234	1.016
p-value	[0.689]		[0.007]	
Math	–0.122*	–0.065	0.003	–0.464***
	(0.074)	(0.122)	(0.048)	(0.162)
N (estimation)	29,422	9,187	6,466	1,051
N (students)	12,218	3,851	2,596	427
First stage F	187	87.8	198	23.9
Non-G&T Mean	1.164	0.776	1.346	0.991
p-value	[0.824]		[0.010]	
Algebra Regents	0.105	–0.044	0.004	–0.052
	(0.122)	(0.178)	(0.052)	(0.182)
N (estimation)	3,075	1,046	738	128
First stage F	231	139	578	69.4
Non-G&T Mean	1.132	0.680	1.213	0.858
p-value	[0.522]		[0.333]	

Notes: This table reports 2SLS estimates of the effect of G&T enrollment on test scores, by student race/ethnicity. Column 1 reports fuzzy RD estimates and column 2 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. 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 F5. Effects on Achievement by Gender

	RD		Lottery	
	Female	Male	Female	Male
	(1)	(2)	(3)	(4)
ELA	0.052	0.047	–0.055	–0.007
	(0.082)	(0.084)	(0.056)	(0.061)
N (estimation)	26,248	23,838	5,245	4,467
N (students)	8,993	8,245	1,752	1,529
First stage F	101	91.2	89.1	79.4
Non-G&T Mean	1.107	0.918	1.296	1.098
p-value	[0.670]		[0.302]	
Math	–0.091	–0.034	–0.102	0.025
	(0.091)	(0.085)	(0.064)	(0.065)
N (estimation)	21,541	19,563	4,440	3,657
N (students)	8,901	8,201	1,744	1,522
First stage F	148	135	118	111
Non-G&T Mean	1.030	1.114	1.278	1.324
p-value	[0.641]		[0.172]	
Algebra Regents	0.081	0.106	0.062	–0.051
	(0.138)	(0.142)	(0.071)	(0.067)
N (estimation)	2,249	2,058	483	435
First stage F	198	175	350	300
Non-G&T Mean	1.007	1.026	1.138	1.171
p-value	[0.898]		[0.266]	

Notes: This table reports 2SLS estimates of the effect of G&T enrollment on test scores, by student gender. Column 1 reports fuzzy RD estimates and column 2 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. 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 F6. Effects on Grade 6 Applications and Enrollment by Race/Ethnicity

	RD		Lottery	
	Black/Hispanic	White/Asian	Black/Hispanic	White/Asian
	(1)	(2)	(3)	(4)
A: First Choice Grade 6 School				
Specialized HS Share	0.026	–0.014	–0.001	0.075***
	(0.028)	(0.025)	(0.042)	(0.014)
N	2,950	9,364	343	2,056
First stage F	321.8	631.6	91.5	657.9
Non-G&T Mean	0.127	0.221	0.183	0.257
p-value	[0.288]		[0.092]	
Value-Added	–0.019	0.007	0.035	0.025*
	(0.030)	(0.023)	(0.033)	(0.014)
N	2,950	9,364	343	2,056
First stage F	321.8	631.6	91.5	657.9
Non-G&T Mean	0.029	0.077	0.063	0.094
p-value	[0.500]		[0.800]	
B: Enrolled Grade 6 School				
Specialized HS Share	0.017	–0.050**	0.049	0.090***
	(0.021)	(0.024)	(0.039)	(0.014)
N	2,343	7,491	284	1,678
First stage F	278.8	502.7	83.7	635.1
Non-G&T Mean	0.081	0.170	0.124	0.212
p-value	[0.036]		[0.331]	
Value-Added	–0.030	–0.002	0.052	0.016
	(0.033)	(0.025)	(0.037)	(0.014)
N	2,343	7,491	284	1,678
First stage F	278.8	502.7	83.7	635.1
Non-G&T Mean	0.011	0.064	0.040	0.088
p-value	[0.504]		[0.375]	

Notes: This table reports 2SLS estimates of the effect of G&T enrollment on characteristics of the grade 6 schools where students apply as their first choice and enroll, by race/ethnicity. Column 1 reports fuzzy RD estimates and column 2 reports 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%.

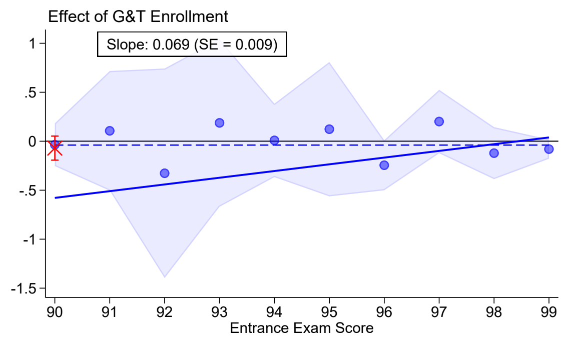
Appendix Table F7. Effects on Grade 6 Applications and Enrollment by Gender

	RD		Lottery	
	Female (1)	Male (2)	Female (3)	Male (4)
A: First Choice Grade 6 School				
Specialized HS Share	–0.028	0.029	0.056***	0.084***
	(0.027)	(0.028)	(0.018)	(0.019)
N	6,867	6,206	1,399	1,181
First stage F	528.8	444.6	400.0	360.5
Non-G&T Mean	0.198	0.199	0.246	0.252
p-value	[0.150]		[0.289]	
Value-Added	–0.013	0.013	0.032*	0.038**
	(0.026)	(0.026)	(0.018)	(0.017)
N	6,867	6,206	1,399	1,181
First stage F	528.8	444.6	400.0	360.5
Non-G&T Mean	0.064	0.066	0.081	0.096
p-value	[0.510]		[0.796]	
B: Enrolled Grade 6 School				
Specialized HS Share	–0.055**	0.004	0.063***	0.118***
	(0.025)	(0.025)	(0.017)	(0.020)
N	5,499	4,895	1,136	964
First stage F	417.1	384.1	422.3	334.9
Non-G&T Mean	0.150	0.149	0.205	0.198
p-value	[0.090]		[0.040]	
Value-Added	–0.016	0.011	0.008	0.041**
	(0.029)	(0.027)	(0.017)	(0.019)
N	5,499	4,895	1,136	964
First stage F	417.1	384.1	422.3	334.9
Non-G&T Mean	0.051	0.051	0.079	0.081
p-value	[0.526]		[0.206]	

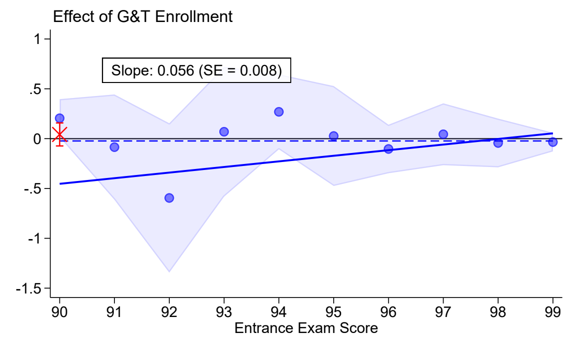
Notes: This table reports 2SLS estimates of the effect of G&T enrollment on characteristics of the grade 6 schools where students apply as their first choice and enroll, by gender. Column 1 reports fuzzy RD estimates and column 2 reports estimates using the lottery design. See Table 5 for a description 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 Figure F1. G&T Effects Gradient, Linear Fit

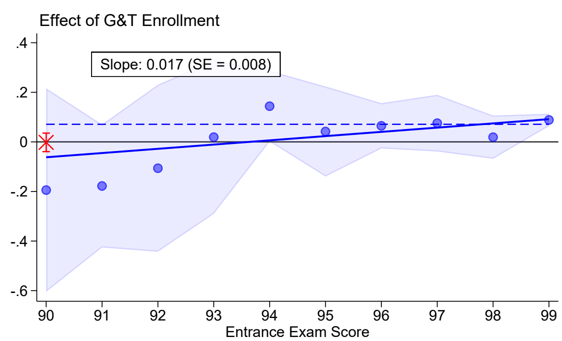
A. Math



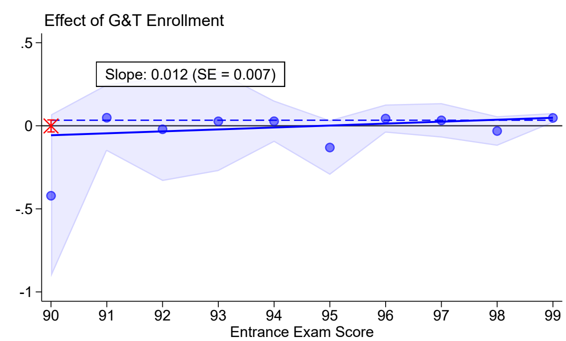
B. ELA



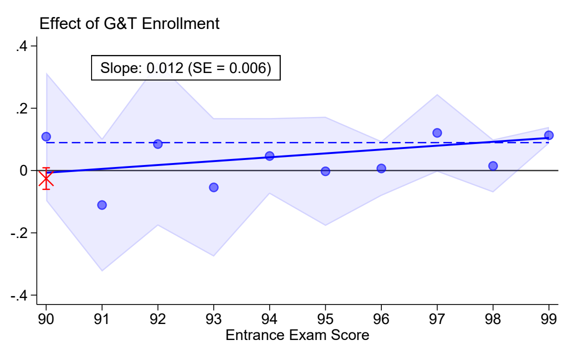
C. Specialized HS Share of First Choice Middle School



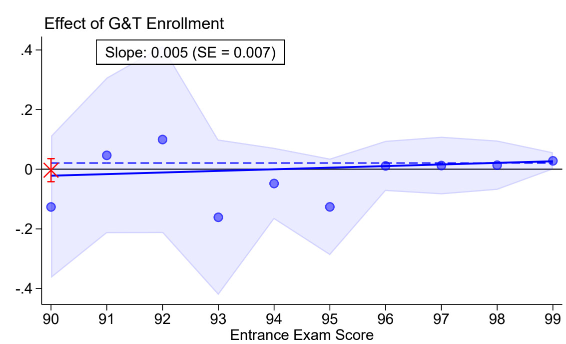
D. Value-Added of First Choice Middle School

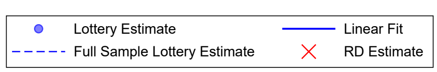


E. Specialized HS Share of Enrolled Middle School



F. Value-Added of Enrolled Middle School

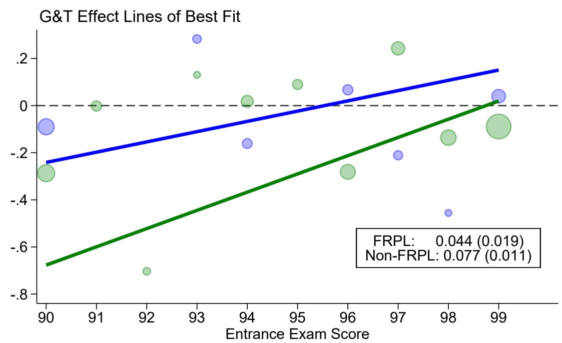




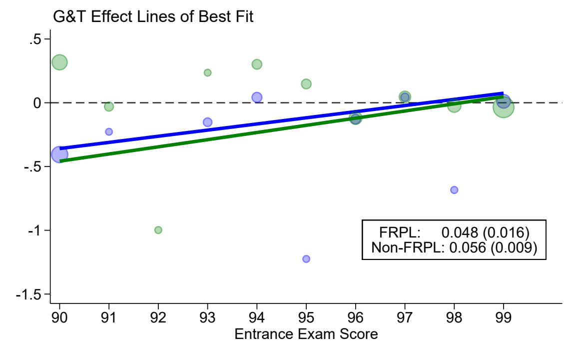
Notes: The plots in this figure depict how G&T effects vary by entrance exam score. The blue markers correspond to 2SLS estimates obtained via estimation in separate samples for each value of the score. Marker size is proportional to the inverse of the standard error on the estimate. The blue line is 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. FRPL G&T Effects Gradient, Linear Fit

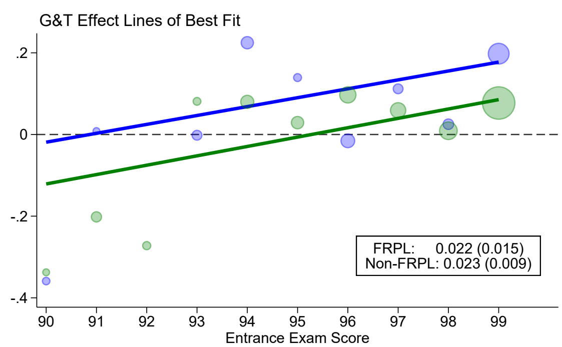
A. Math



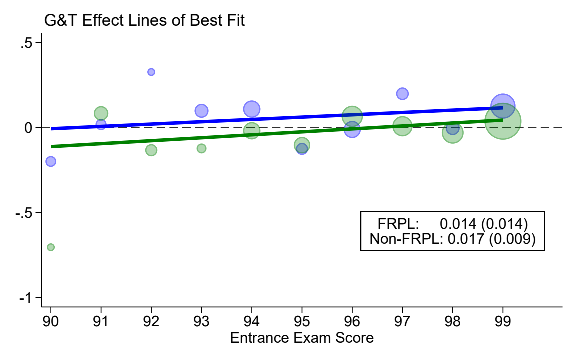
B. ELA



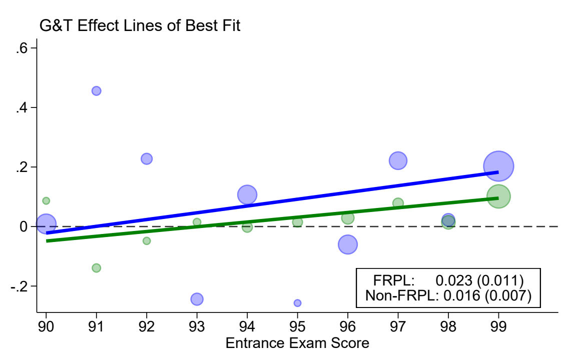
C. Specialized HS Share of First Choice Middle School



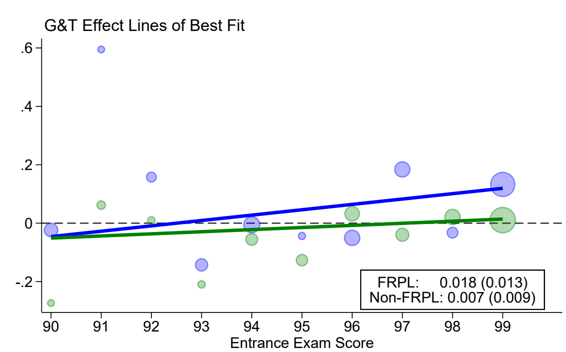
D. Value-Added of First Choice Middle School

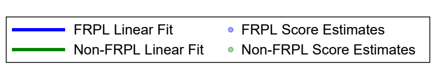


E. Specialized HS Share of Enrolled Middle School



F. Value-Added of Enrolled Middle School





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.

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 status 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

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

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

Offer probabilities, given by the second part of the integrant 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

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

The expression for an individual's likelihood contribution across all stages is given by Equation 4. We approximate the likelihood via simulation using 100 realizations of ϵ_i and c_i drawn using a Sobol sequence. We use the BFGS solver from the `scipy.optimize` package in Python, setting the tolerance such that convergence is attained when the largest absolute component of the gradient is smaller than 10^{-4} (`gtol` $< 10^{-4}$).

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

Parameter	Estimate
Distance Coefficient	–0.568
	(0.000)
District Priority Coefficient	1.441
	(0.005)
Sibling Priority Coefficient	3.721
	(0.008)
Scale Parameter on Stage 3 Preference Shock (κ)	0.666
	(0.001)
Baseline Marginal Cost (c)	0.000
	(0.000)
Marginal Cost Scale Parameter (σ_c)	0.095
	(0.000)
Mean Citywide Program Fixed Effect	3.845
Mean District Program Fixed Effect	–1.985
Share with cost below 0.01	0.084
N	2,358

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 for more details on the procedure.

Table G2. Model Fit, Shares of G&T Enrollees

	Observed (1)	Simulated (2)	Simulated 95% Prediction Interval (3)
FRPL	0.212	0.211	[0.199, 0.221]
Black	0.033	0.031	[0.025, 0.037]
Asian	0.485	0.484	[0.471, 0.500]
Hispanic	0.072	0.067	[0.059, 0.074]
White	0.408	0.415	[0.400, 0.429]

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.