// NO.24-076 | 05/2025

# DISCUSSION PAPER

// MAXIMILIAN BACH, THILO KLEIN,
AND SARAH MCNAMARA

# Access, Achievements, and Aspirations: The Impacts of School Tracking on Student Outcomes

Leibniz Association

ZEW

# Access, Achievements, and Aspirations: The Impacts of School Tracking on Student Outcomes*

Maximilian Bach[1], Thilo Klein[2,3] and Sarah McNamara[2,+]

[1] German Federal Statistical Office, Wiesbaden, Germany

[2] ZEW – Leibniz-Center for European Economic Research, Mannheim, Germany

[3] Pforzheim University, Germany

First Version: November 8, 2024

This Version: May 26, 2025

**Abstract**

Though the use of tracking policies to stratify students is commonplace, evidence concerning the effects of ability-based tracking on student performance is mixed. Using rich data from the Hungarian secondary school centralized assignment mechanism and a quasi-experimental framework, we find that attending the highest track noticeably improves standardized test scores and university aspirations two years post-match. Heterogeneity analyses find this effect is independent of socioeconomic status, prior achievement, and parents' educational attainment, and we find only limited evidence of peer spillover effects in terms of academic ability. Given socioeconomic disparities in track placement, tracking may reinforce educational inequality.

**Keywords:** education; school choice; tracking; centralized school admissions; student achievement; inequality of opportunity.
**JEL Codes:** I21, I24, I28, E47, C26

*__Acknowledgments__: We are grateful to participants of (virtual and in-person) seminars in Mannheim, Zürich, Budapest, and Gröningen. We are thankful to Marco Bertoni, Ellen Greaves, Antonio Miralles, Olmo Silva, Bertan Turhan, Ulrich Zierahn-Weilage, Elena Fumagalli, Matthias Parey, and Sergey Popov for comments, as well as audiences at the 2022 Easter Workshop on School Choice and Matching Markets at Queen's University Belfast, the 2023 European Winter Meeting of the Econometric Society in Manchester, the XXXII meeting of the Economics of Education Association (AEDE) in Valencia, and the 2024 Conference on Mechanism and Institution Design in Hungary. We also thank Julia Heigle for her valuable research assistance. Financial support was provided by the Leibniz Association through the project "Improving School Admissions for Diversity and Better Learning Outcomes". The views expressed in this article are ours alone and do not necessarily reflect the views of the Leibniz-Association or the German Federal Statistical Office. We have no conflicts of interest. +Corresponding author: sarah.mcnamara@zew.de.

# 1 Introduction

The degree to which schools track students—separating students based on prior academic performance into different classes, tracks or schools—varies between countries. Some, like Finland, avoid tracking and rely solely on compulsory comprehensive schooling. Others, like Germany, rigorously track students based on ability as early as age ten. Between these two extremes lie countries like the US, where students are typically streamed into ability-based groups or classes within schools, though magnet and selective charter schools are becoming increasingly common.

The main rationale for tracking is that teachers can tailor lessons toward the specific ability levels of their students (Duflo et al., 2011). Yet its critics argue that tracking disproportionately harms students assigned to lower tracks and only benefits students assigned to higher tracks, thus exacerbating educational inequality (Reichelt and Eberl, 2019). This issue is potentially compounded when the school system, in general, is highly segregated along socioeconomic lines. Students from disadvantaged backgrounds are then more likely to attend educational settings that, on average, offer fewer resources, employ less experienced teachers, and maintain lower academic expectations. These disparities can hinder their access to more competitive academic tracks (see, for example, Dräger et al., 2024).

However, the causal evidence on the effects of tracking itself is mixed; some studies find a positive effect (see, e.g., Abdulkadıroğlu et al., 2022; Carrell and Kuka, 2018; Abdulkadıroğlu et al., 2017; Berkowitz and Hoekstra, 2011; Jackson, 2010; Hastings and Weinstein, 2008; Cullen et al., 2006), while others find no effect (see Beuermann and Jackson, 2022; Barrow and de la Torre, 2020; Abdulkadıroğlu et al., 2014; Lucas and Mbiti, 2012). In part, this may be explained by differences in country-specific institutional settings, but it may also be partially driven by the incompleteness of students' observable choice sets—either because preferences and priorities are not observed, which is often the case, or because plausible outside options exist (e.g., Abdulkadıroğlu et al., 2022; Abdulkadıroğlu et al., 2017). Thus far, however, the literature is inconclusive on whether tracking acts as a selective filter for ability or confers a separate benefit on academic achievement.

We examine the effect of tracking on student outcomes in the context of Hungary's between-school tracking system, where most students are tracked into different educational pathways at the end of the 8th grade, aged fourteen. Assignment to one of three tracks is based on a centralized assignment mechanism in which students submit a ranked list of program choices and are, in turn, strictly ranked according to certain criteria. These tracks are hierarchically ordered such that assignment to one of three tracks determines both future education pathways and individual career opportunities. Students apply for school-specific programs, programs can set their own admission criteria, and individual programs are typically small in size (often consisting of only one class). We exploit the structure of the centralized assignment mechanism, applying recent methodological advances by Abdulkadıroğlu et al. (2022) to estimate the causal effect of attending the highest track.

In a system that uses a centralized assignment mechanism to match students and schools, conditioning on a coarse function of students' preferences and schools' priorities allows us to eliminate omitted variable bias arising from the potential correlation between preferences and eventual track assignment. To do so, we first compute the so-called local Deferred Acceptance (DA) propensity score (Abdulkadıroğlu et al., 2017) that comprehensively describes an individual's risk of assignment to a certain track, based on a student's preferences over schools and their local risk of assignment at each preferred school. We then use this propensity score to estimate the marginal value added of assignment to the highest track using a regression discontinuity-type approach. Simultaneously, restricting the analysis to applicants *near* school-specific admissions cut-offs controls for selection bias that occurs when better students are more likely to be admitted to higher-track schools. Combining this approach with a two-stage least squares (2SLS) framework, where we use high-track program offers as an instrument for attendance, we can estimate the effect of high-track attendance while controlling for potential correlation between student preferences and track assignments.

On average, we find that attending the highest track improves average standardized test scores in the 10th grade by approximately 0.11 standard deviations, which can be further decomposed into effects of 0.14 and 0.07 standard deviations for mathematics and reading, respectively. We also find important gender differences, and though both male and female students experience a positive effect on mathematics test scores, this is much larger for males, and the effects on reading accrue to males only. On the other hand, while the university aspirations of 10th-grade students benefit from an average increase of approximately 0.08 standard deviations, this is driven entirely by female students, with no corresponding effect for males. Further, the Hungarian context—in which schools have the autonomy to set their own admissions criteria and the programs are small in size—yields quasi-experimental variation in admission to the highest track for students across the prior achievement distribution. This allows us to estimate local average treatment effects (LATEs) across almost the full distribution of prior academic achievement, not only for students at a specific threshold or universal cut-off. Heterogeneity analyses do not reveal any meaningful differences in effect size by baseline achievement or socioeconomic status (SES), though the effects on university aspirations are higher for those from relatively poorer socioeconomic backgrounds.

To disentangle those factors driving the effect of high-track attendance, we investigate several potential mechanisms to include peer spillovers in academic achievement, peer behavior, and peer diligence as a proxy for grit. We test whether the positive effects on achievement and aspirations are generated primarily through peer effects, given that peer quality in terms of baseline academic achievement differs between the highest and the intermediate track by, on average, 0.17 standard deviations. Conditional on the leave-own-out average of the baseline achievement distribution, we adopt a similar methodology to that used in the main analysis and estimate the effects of attending higher and lower peer-quality programs for inframarginal students who are always assigned to the highest track.

We find that attending a program with higher-quality peers does not have an effect on average standardized achievement, though it does have a weakly significant positive effect on the reading test scores of males in particular. Peer effects in terms of academic ability are therefore unlikely to be driving the main results. We do, however, find evidence that peer behavior is important, particularly for female students' mathematics scores, though we do not find correspondingly positive effects of peer diligence—suggesting that the absence of disruption is more important than conformity with hard-working peers' study habits.

This paper relates to a large body of literature focused on the effects of ability-based tracking in school systems. We are the first to apply this method to a universe of students and schools in which students have limited outside options (e.g., private institutions, or attending out-of-district, etc.), and unlike previous studies in the literature (e.g., Dobbie and Fryer, 2014; Ding and Lehrer, 2007; Hanushek et al., 2003; Hoxby, 2000; see Barrios-Fernandez, 2023 for an overview), we demonstrate that the positive results we obtain are likely not driven by the academic abilities of peers. From an educational policy perspective, these findings shed light on the efficiency-equity trade-off at the individual student level (see, e.g., Colas et al., 2021; Ferraro and Põder, 2018; Barrera-Osorio and Filmer, 2016; Woessmann, 2008). Second, leveraging Hungary's unique institutional context—in which secondary admissions are centralized—we are able to focus on the distributional consequences, a matter that has received comparatively less attention in the literature thus far.

Existing evidence on the consequences of tracking largely stems from two approaches. The first uses variation in tracking policies, or changes in assignment mechanisms, to identify effects. This variation may be generated via de-tracking reforms,[1] differences between regions (see, e.g., Borghans et al., 2020, Matthewes, 2020, and van Elk et al., 2011), or country-level differences (see, e.g., Hanushek and Wössmann, 2006). However, de-tracking reforms typically do not occur in isolation and are often accompanied by other institutional or educational content changes (e.g., curriculum revisions).[2] A second strand identifies effects *within* a tracking system for students at the margin of admission to a higher track, primarily using an RDD-type approach (see, e.g., Borghans et al., 2019; Dustmann et al., 2017). However, this typically limits identification to students at a specific cut-off, and is thus potentially uninformative about the effects for inframarginal students further away from a universally applicable admissions threshold.

Although we also rely on an RDD-type approach, one of our key contributions is that, unlike previous studies that are only able to compute effects at a specific margin, we study the effects of high-track attendance across a wide range of the prior achievement distribution. The context of the Hungarian system, with many small programs in different "markets", gives us greater variation in cut-offs and therefore allows us to use the admissions criteria of individual small programs as localized cut-offs. Unlike prior studies that rely on universal cut-offs, this allows us to address not only the question of "does tracking matter?" but also "for whom does it matter most?" We find that while students with lower prior performance on standardized tests, or from comparatively more deprived socioeconomic backgrounds, are less likely to be admitted to the highest track, on average, these students benefit at least as much from high-track attendance.

Finally, while many of the more recent contributions to the literature on educational differentiation study long-term outcomes, such as earnings or degree attainment, we directly study one of the main mechanisms via which tracking can be expected to affect long-term outcomes. That is, the tailoring of instruction to in-

[1] See, e.g., Canaan (2020), for France; Roller and Steinberg (2020), Piopiunik (2014), and Bach (2023) for Germany; Pekkala Kerr et al. (2013), for Finland; Guyon et al. (2012), for Northern Ireland; Hall (2012), Meghir and Palme (2005), for Sweden; Malamud and Pop-Eleches (2011), for Romania; and Aakvik et al. (2010), for Norway.

[2] One notable exception is Guyon et al. (2011), however, they only study a partial increase in students admitted to the highest track.

dividual students' needs potentially improves learning outcomes. Long-run studies have the drawback that higher track attendance can increase educational attainment or earnings simply because they increase students' eligibility for higher levels of post-secondary education or career fields, rather than directly improving students' competences. In this paper, we demonstrate improvements on standardized measures of mathematical and language abilities, similar to the core components of the PISA test, that are relevant to everyday life and are designed to directly test students' skills in solving labor market-relevant problems; such as extracting information from written text or computing a balance sheet.

This paper proceeds as follows. Section 2 describes the Hungarian institutional context, and Section 3 describes the data. Section 4 introduces our empirical approach and describes how identification can be obtained in this quasi-random setting by exploiting the central assignment mechanism. Section 5 presents our main results, and Section 6 presents the results of robustness tests. Finally, Section 7 concludes.

# 2 Institutional Context

In Hungary, educational tracks are hierarchically ordered, such that the choice of educational pathway from secondary school onward determines future education and employment opportunities. Most students apply for their preferred secondary education program in the 8th grade.[3] Students apply for specific school-course combinations and must decide between three tracks: vocational training schools (lowest track), vocational secondary schools (intermediate track), and grammar schools (highest track). In the highest track, students follow an academically oriented curriculum in preparation for the 12th-grade maturity exam. The intermediate track comprises mixed programs combining academic study with vocational subject options after the 10th grade, although students can opt to take the maturity examination. In the lowest track, however, students specialize in a vocational training pathway. They graduate with a lower-level vocational qualification and are not able to take the maturity exam at the end of 12th grade.

[3]A small number of extended duration high-track programs offer enrollment after the 6th grade. These programs do not participate in the 8th-grade centralized matching process and are thus excluded from our analysis. Since our identification strategy targets students who are empirically "near" the admissions cut-offs of high-track programs at the end of the 8th grade (i.e., those "at risk" of high-track assignment), these early-admitted students (who are effectively guaranteed a spot) should not influence our results.

The structure and progression of the Hungarian system are outlined in Figure 1. In this paper, we focus on students at the margin of admission to the highest track and estimate the value added of attendance on student test scores and university aspirations two years post-assignment. Students in the highest track follow an academic curriculum that aims to prepare them for higher education after the maturity exam (*"érettségi vizsga"*) at the end of 12th grade. In the intermediate track, on the other hand, students are introduced to a vocation-oriented area of study in the 11th and 12th grades. If students choose to continue after attempting the maturity exam at the end of 12th grade, they complete a vocational specialty in their 13th year of schooling, leading to a professional qualification (*"szakmai vizsga"*).

The maturity exam itself is comprised of at least five mandatory examinations in Hungarian literature and language, mathematics, history, a foreign language, and a fifth subject of the student's choosing that can be vocational in nature, or from the general education offer. Students can choose to sit additional examinations. In each subject, these exams can be taken either at the standard level (*"közép"*) or the higher level (*"emelt"*). Entrance to higher education is competitive, and students who pass subjects at the advanced level are more likely to be admitted to oversubscribed courses. Better preparation at the secondary school stage can therefore significantly affect later education and career opportunities.

When applying to secondary schools in the 8th grade, each student submits a strict rank-order preference list of programs consisting of specific school-course combinations. These lists can be of arbitrary length, and choices are also not geographically restricted. Since 2000, students have been allowed to apply for any program nationwide. For those students unwilling to commute or who select a program far away from their usual place of residence, dormitories are available. Given that the capacities of schools in Hungary typically exceed the number of students due to demographic change, schools are often able to offer dormitory places to applicants who require them. On the school side, programs must generate strict priority rankings over applicants for all students who list a particular program (conditional on the school deeming them "acceptable" for admission).

When determining admission priorities, programs may base their criteria on a number of factors, including results from centralized examinations organized at the beginning of the 8th grade, in-school achievement grades, and some particularly popular or selective programs may also request students participate in oral interviews. Additionally, some programs may prioritize students with a particular religious affiliation or student-specific characteristics, such as an enrolled sibling. These criteria must be equally applied to all applicants to a particular program. Some programs rely on all of these criteria, while others rank applicants solely based on in-school performance during the 8th grade. Typically, programs rank applicants according to a weighted average of their primary school grades, entrance exam scores, and interview scores. The relative weights are determined by individual programs, though they are subject to some constraints.

Based on these program priorities and student preferences, the final assignment is determined via a centralized assignment mechanism organized at the national level. The matching is performed by computer software using the student-proposing deferred acceptance (DA) algorithm (Biró, 2008), described in greater detail in Appendix A. The Hungarian mechanism does not use a randomized lottery-type tie-breaker. Schools must generate strict rankings over students, although they retain the right not to rank any students they deem unacceptable for admission. After matching has concluded, schools decide how to form classes. An additional matching round is conducted for unmatched students and unfilled courses, though this is organized at the school level and affects only a small fraction of students.

# 3 Data

Our analysis relies on two key sources of data: administrative data regarding the matching procedure—including students' preference lists and program priority rankings, and student survey data, which includes individual-level data on students and their family background.

**KIFIR.**[4] Our source of administrative data is KIFIR, a dataset containing the preferences, priorities, and outcomes of the national centralized matching scheme for the universe of Hungarian 8th-grade students applying for secondary education. We rely on the 2015 wave of this matching procedure. Overall, 88,401 students applied to 6,181 different school-program combinations offered by 1,035 schools. The average number of schools listed by each student is 4.47, and 94.4% of students were matched to a school in the first round.

**National Assessment of Basic Competencies (NABC).** Our source of student

[4]KIFIR is an acronym for *"középfokú felvételi információs rendszer"*, or "secondary enrollment information system".

survey data comes from three waves of the NABC, conducted in 2013, 2015, and 2017 when students were in the 6th, 8th, and 10th grades, respectively. The NABC data contains the results of standardized tests taken by all students in Hungary at the end of grades 6, 8, and 10, which are designed to measure student ability in reading and mathematical literacy. Similar to the core components of the PISA test, the NABC is not designed to measure student performance according to a specific curriculum, but rather fundamental competences. Specifically, the reading section is comprised of two 45-minute blocks where students retrieve, analyze, and reflect on information obtained from narrative and expository texts. The mathematics section focuses on real-life applications of mathematical skills, including reading tables and graphs or performing financial calculations. In addition to raw test scores for mathematics and reading, the NABC data contains information from student surveys, which were answered voluntarily with relatively high response rates (approximately 80%). The survey component yields a rich set of sociodemographic background controls in addition to details on the students' academic histories (e.g., past GPA, both overall and by subject), classes repeated, family background (e.g., family composition, parental education, parental occupation, and employment status), and career aspirations.

**Estimation Samples.** The KIFIR and NABC data can be merged via unique student identifiers. Test scores from the 2015 wave of the NABC were measured during the 8th grade and provide us with *pre-assignment* information on academic achievement. Test scores from the 2017 wave were measured in the 10th grade and provide us with *post-assignment* information on academic achievement. However, not all students contained in the KIFIR dataset can be linked to the NABC data. This is for two key reasons. First, schools provided a student identifier voluntarily that can be used to link the two data sources. Second, 5.33% of the students in the KIFIR dataset were matched to so-called early-selective high-track programs, which begin in grades 5 or 7, for which we do not have adequate NABC data. In total, we have information on pre-assignment and post-assignment academic achievement for 54,013 students who applied for secondary education in 2015.[5]

[5]Table 1 tests for selective attrition to ensure that the response rate of missing NABC observations in the 10th grade is not a function of a student's high-track assignment probability, previous academic attainment, or individual student characteristics.

# 4 Empirical Strategy and Identification

The fundamental problem when recovering the causal effect of track-assignment on student outcomes several years post-match is the non-random sorting of students. If more competitive programs, or programs of a certain “type”, attract better students, higher test scores cannot be solely attributed to the effect of the track itself. Under centralized assignment, offers made at specific programs are determined by student preferences over programs and the priority rankings by programs over students. These preferences and rankings are, therefore, two key confounding channels. To estimate the causal effect of high-track attendance, our estimation framework leverages randomness embedded in the Hungarian centralized assignment mechanism. We apply recent methodological advances by Abdulkadıroğlu et al. (2022) to estimate the causal effect of attending the highest track for students at the margin of admittance to specific programs.

This method proceeds in two steps. First, we control for a scalar function of student preferences referred to as the *local DA propensity score*. This score comprehensively describes an individual’s risk of assignment to a certain track, based on student preferences over programs and a student’s local risk of assignment at each of their preferred programs. This is more practical than fully conditioning on preferences when the number of preference “types” is very large. Second, by only comparing applicants in a narrow window around program-specific cut-offs, similar to a regression-discontinuity design (RDD), we eliminate bias arising from potential correlation between a student’s outcomes and their relative rank position. A brief description of the underlying theory and the resulting estimation strategy are discussed in the following, though an in-depth discussion of most issues can be found in Abdulkadıroğlu et al. (2017) and Abdulkadıroğlu et al. (2022).

## 4.1 Identification

While Hungary implements a DA algorithm featuring school-specific tie-breakers—to include past test scores, grades, and interview scores—we first abstract from these complications and focus on a market with a single, shared tie-breaker to explain the key elements of the approach by Abdulkadıroğlu et al. (2022).

Assume that students submit rank-ordered preferences over schools and let $\theta_i$ denote student $i$’s list of preferences (henceforth, student type).[6] Let student $i$’s

[6] In contexts where schools disclose only a strict ranking of applicants, and the underlying admissions criteria remain unobservable, every student belongs to a single marginal-priority group,

ability be denoted by $\epsilon_i$. In this abstraction, students are matched to schools by a DA algorithm that takes only student type ($\theta$) and a single, randomly assigned tie-breaker as inputs. The tie-breaker distinguishes between students with the same preferences. The DA algorithm outputs a single school assignment for each student denoted by the indicator $Z_{ij}$, which takes the value one when student $i$ is assigned to school $j$. $Z_i = (Z_{i1}, \ldots, Z_{iJ})$ collects assignment indicators for student $i$. Random tie-breaking ensures that admission offers are randomly assigned, conditional on student preferences. The conditional random assignment (CRA) of $Z_i$ can be summarized as follows:

**Assumption CRA:** *Student ability is independent of high-track program assignment conditional on student type:* $\epsilon_i \perp\!\!\!\perp Z_i|\theta_i$.

The CRA assumption suggests that an estimate of the causal effect of assignment to the highest track amounts to a comparison of student outcomes for students receiving a high-track program assignment within strata defined by $\theta_i$.

As noted in Abdulkadıroğlu et al. (2017), however, in practice, there are typically nearly as many preference combinations as there are students. Consequently, fully conditioning on student "type" is often infeasible or leaves very few degrees of freedom for empirical analysis. Instead, Abdulkadıroğlu et al. (2017) propose reducing the dimensionality of the conditioning set by pooling students of different types in a manner that preserves conditional independence of school assignments and potential outcomes. Pooling relies on the *school assignment propensity score*, which for student $i$ assignment to school $j$ is defined:

$$p_{ij} = Pr(Z_{ij}|\theta_i). \tag{1}$$

Abdulkadıroğlu et al. (2017) show how to compute $p_{ij}$ analytically. The vector $p_i = (p_{i1}, \ldots, p_{iJ})$ collects the propensity scores for student $i$ at all schools. As demonstrated in Rosenbaum and Rubin (1983), random assignment conditional on a vector of controls implies conditional random assignment given the propensity score obtained from these controls. This result can be stated as follows:

**Lemma 1:** *Under Assumption CRA, student ability is independent of school assignments conditional on assignment risk:* $\epsilon_i \perp\!\!\!\perp Z_i|p_i$.

so "student type" is determined solely by the student's preference profile—rather than by the joint combination of preferences and priorities assumed in Abdulkadıroğlu et al. (2022)'s original setting.

In other words, since school assignment is ignorable conditional on type, it is also ignorable conditional on the school assignment propensity score. Moreover, assignment scores are determined by a few key match parameters. Conditioning on low-dimensional propensity scores thus leaves far more degrees of freedom than fully conditioning on student type.

Abdulkadıroğlu et al. (2022) generalize this approach to DA algorithms with non-random, school-specific tie-breakers, to include previous test scores, grades, and interview scores, as is the case in the Hungarian setting. With this form of tie-breaking, assignments are a function of both tie-breakers and student type, and thus, confounding from non-lottery tie-breakers remains even after conditioning on $p_{ij}$. To overcome this challenge, Abdulkadıroğlu et al. (2022) propose focusing on assignment probabilities for applicants with tie-breaker realizations close to key cut-offs, and the inclusion of local controls for tie-breaker values as is typical in a regression discontinuity design (RDD) setting.

Denote by $R_{is}$ student $i$'s tie-breaker value at school $s$, where $R_{is} < R_{js}$ implies school $s$ prefers student $i$ to student $j$. A DA allocation with school-specific, non-lottery tie-breakers is characterized by a set of *tie-breaker cut-offs* denoted $\tau_s$ for school $s$. For any school $s$, $\tau_s$ is determined by the tie-breaker of the last student (highest tie-breaker value) assigned to $s$. For each school-specific tie-breaker cut-off, we define an interval $(\tau_s - \delta, \tau_s + \delta]$ where the parameter $\delta$ is a bandwidth analogous to that used for non-parametric RDD estimation. In the limit, as $\delta$ shrinks to zero, the probability that a student has a tie-breaker value that clears cut-off $\tau_s$ (i.e., $R_{ij} < \tau_s$) inside this interval can be treated as approximately random with a probability equal to $\frac{1}{2}$.

Using this insight, Abdulkadıroğlu et al. (2022) propose estimating the probability that student $i$ is assigned to school $s$, the *local DA propensity score* denoted by $\psi_s(\theta_i, R_i)$, as follows:

$$\hat{\psi}_s(\theta_i, R_i, \delta) = \begin{cases} 0 & \text{if } R_{is} > \tau_s + \delta \\ 0 & \text{if } R_{ib} \leq \tau_b - \delta \text{ for some } \ b \in B_{\theta_i s} \\ 0.5^{\hat{m}_s(\theta_i, R_i)} & \text{if } R_{is} < \tau_s - \delta \\ 0.5^{1+\hat{m}_s(\theta_i, R_i)} & \text{if } R_{is} \in (\tau_s - \delta, \tau_s + \delta]. \end{cases} \tag{2}$$

where $B_{\theta_i s}$ is the set of schools that student $i$ prefers to $s$ and

$$\hat{m}_s(\theta_i, R_i, \delta) = \left| \left\{ b : R_{ib} \in (\tau_s - \delta, \tau_s + \delta] \text{ for } b \in B_{\theta_i s} \right\} \right|$$

The first two lines in (2) refer to scenarios where applicants to $s$ are disqualified at $s$ (line 1), or are assigned to some preferred school for sure (line 2). In both cases, the probability of assignment to school $s$ is treated as zero. In a scenario where applicants are surely qualified at $s$, the probability of assignment to $s$ is determined entirely by the probability of not being assigned to some preferred school, where the tie-breaker falls within the narrow interval around that school's cut-off. This probability is given by $0.5^{\hat{m}_s(\theta, R_i)}$ (line 3). Finally, line 4 describes the assignment probability at school $s$, where student $i$ does not surely qualify, as the product of the disqualification rate at the applicants' preferred schools (line 3) and the qualification rate at $s$, where the latter is 0.5 conditional on the student falling in the narrow interval around the cut-off.

Result (2) implies that the causal effect of high-track attendance is identified in the Hungarian setting. To see this, let $S_G$ denote the set of high-track programs. Because DA assigns students to at most one school, the local propensity score for assignment to any high-track program, denoted $\psi_G(\theta, R)$, can be estimated as the sum of the estimated scores for all high-track programs:

$$\hat{\psi}_G(\theta, R, \delta) = \sum_{s \in S_G} \hat{\psi}_s(\theta, R, \delta) \tag{3}$$

Now let $D_i$ denote program assignment and $C_i$ actual program enrollment. Further, let the causal effect of enrollment be given by a constant $\beta$, so that observed outcomes are determined by $Y_i = Y_{0i} + \beta C_i$, and $D_i$ satisfies the exclusion restriction that it affects $Y_i$ solely by changing $C_i$. In this case, Abdulkadıroğlu et al. (2022) show that asymptotically (as $\delta$ shrinks to zero), $D_i$ is independent of potential outcomes conditional on an estimate of the local high-track propensity score.[7] These results suggest a 2SLS procedure with second- and first-stage equations that can be written in stylized form as:

[7]See Section 4 of Abdulkadıroğlu et al. (2022), where some regularity conditions are invoked.

$$
\begin{aligned}
Y_i &= \beta C_i + \sum_x \alpha_2(x) d_i(x) + g_2(R_i) + \eta_i \\
C_i &= \gamma D_i + \sum_x \alpha_1(x) d_i(x) + g_1(R_i) + \nu_i,
\end{aligned} \tag{4}
$$

where $\beta$ is the causal effect of interest, $d_i(x) = 1\{\hat{\psi}_G(\theta_i, R_i, \delta) = x\}$, and the set of parameters denoted $\alpha_2(x)$ and $\alpha_1(x)$ provide saturated controls for the local propensity score. The functions $g_2(R_i)$ and $g_1(R_i)$ implement local linear controls for school-specific tie-breakers for the set of applicants inside the narrow interval around the school-specific cut-offs determined by the chosen bandwidth $\delta$.

These local linear controls serve to control for imbalances between students just above and below the cut-offs, as in RDD settings, and are parameterized as:

$$
g_2(R_i) = \sum_s \omega_{1s} a_{is} + k_{is}\big[\omega_{2s} + \omega_{3s}(R_{is} - \tau_s) + \omega_{4s}(R_{is} - \tau_s)1(R_{is} > \tau_s)\big] \tag{5}
$$

where $a_{is}$ indicates whether applicant $i$ applied to school $s$, and $k = 1\big[R_{is} \in (\tau_s - \delta, \tau + \delta]\big]$. $g_1(R_i)$ is parameterized analogously to (5). The sample used to estimate (4) is limited to applicants with high-track assignment risk; thus, $g_1(Ri)$ is defined analogously to $g_2(Ri)$. Note that saturated regression conditioning on the local propensity score eliminates always-assigned and never-assigned students with a score of 0 or 1, respectively, because track assignment is constant for these applicants. The identifying variation comes from those students with a non-degenerate risk of assignment to the highest track. An applicant has *high-track assignment risk* when $\hat{\psi}_G(\theta_i, R_i, \delta) \in (0, 1)$.

Next, we define a narrow window around the admission thresholds (i.e., cut-offs) of individual programs. In this setting, the cut-off at a given program is determined by its capacity and thus is the rank of the last student seated. When there are fewer than two in-bandwidth observations on either side of the relevant cut-off, the bandwidth for that program is set to zero. Given the relatively small size of individual programs in terms of number of seats, for the main analysis, we do not compute local bandwidths at individual programs. Rather, we use a common bandwidth of 0.25 based on two criteria: as demonstrated in Figure 2, it is the bandwidth beyond which estimates are stable, and for which a placebo test indicates a null effect on 8th-grade test scores.[8] In an additional robustness test,

[8]The placebo tests on 8th-grade outcomes ensure that the effects we measure are not spurious, arising due to poor balance. When conducting placebo tests, we control for the high-track propensity score, RDD controls, and main control variables described in Table 4.

we alternatively construct the selected sample (and corresponding local propensity score) by computing locally optimal bandwidths, following the extensions to Imbens and Kalyanaraman (2012) proposed in Calonico et al. (2017) and Calonico et al. (2019) to determine the mean square error (MSE) optimal bandwidth choice at individual programs.

One benefit of our approach is that by allowing a flexible definition of the admissions cut-off, which varies by program, we can study the effects of high-track accessions at different points of the distribution of prior achievement. Figure 3 illustrates the distribution of standardized test scores, as measured in the 8th-grade pre-tracking, for the lowest-scoring student admitted to each high-track program relative to the overall population. As there is substantial variation in the cut-off at individual programs evident, the approach described above leads to substantial variation among "at-risk" students in the selected sample—as Figure 4 shows. While, on average, high-track students have higher 8th-grade test scores compared to the overall student population (see Panel (a)), for our sample of "at-risk" students, there is substantial common support relative to the universe of Hungarian students (see Panel (b)). This allows us to overcome a typical limitation associated with an RDD-type approach.

When estimates are localized around a universal admissions cut-off, and the Local Average Treatment Effect (LATE) is estimated only for marginally accepted/rejected students at this threshold, it is not *a priori* clear whether applicants not in the neighborhood of this cut-off would also benefit from high-track attendance. However, in our setting, the large degree of variation in school-specific cut-offs allows us to use the admission criteria of individual small programs as localized cut-offs, and estimate the LATE for different points on the baseline achievement distribution.

## 4.2 Validation of the Empirical Design

Our empirical strategy approximates randomized assignment to the highest track by limiting the sample to students with non-degenerate "risk" of assignment to individual high-track programs. This approach requires that, after conditioning on the local propensity score and running variable controls, high-track program offers are as good as random. In this section, we examine the diagnostics needed to validate this design.

First, Table 2 provides descriptive statistics for the overall sample and those at risk of high-track assignment. Here, it is already evident that naïve estimates of the effect of high-track attendance are likely to be affected by student selection. Those holding an offer from a high-track program, on average, listed almost 50% more high-track programs as a fraction of their overall list and were almost twice as likely to list a high-track program first. Similarly, the first column of Table 3 reveals large differences in student characteristics between those who hold a high-track program offer and those who do not. Students with an offer are positively selected on several key characteristics, including baseline achievement measures and parental background. Baseline test score gaps in reading, for example, are approximately 0.84 standard deviations higher for students receiving an offer.

In the second column of Table 3, we test whether receiving a high-track program offer predicts student characteristics conditional on saturated propensity score controls. The second column restricts the sample to students with non-degenerate high-track assignment risk (i.e., those with a propensity strictly between 0 and 1), conditional on the common empirical bandwidth ($\delta_N = 0.25$). Although this reduces the number of observations from $54,631$ in the full sample to $2,518$ students with a non-degenerate assignment risk, there is a substantial reduction in both the size of the coefficients and statistical significance. This implies that conditioning on the local propensity score significantly reduces the risk of selection effects and omitted variable bias for the estimates presented in the next section.

# 5 Results

In the following, the fully saturated specifications include propensity score and running variable controls (i.e., distance to the cutoff at individual programs), individual student controls, such as age, gender, SES, among others, and baseline test scores measured in the 8th grade. See Table 4 for further details.

As described in Section 4, we instrument enrollment in the highest track with an offer of a place in a high-track program. Table 5 presents the results from this first stage. Similar across all specifications, the results suggest that even with a fully saturated model high-track enrollment is probabilistic. This is because, on the one hand, not all students accept their offer, and, on the other, some applicants who do not hold an offer still ultimately enroll in a high-track program. In column (3) of Table 5, the first stage coefficient suggests that an offer from a high-track program increases the likelihood of high-track enrollment by 66 percentage points.

Table 6 reports the main results from our pseudo-RDD estimation procedure, and shows a substantial positive effect of high-track attendance on average 10th-grade test scores. The estimated test score gain is statistically significant across all specifications and remains robust with the inclusion of controls. Column (3) shows that, on average, attending the highest track leads to an improvement in test scores of approximately 0.11 standard deviations in the fully saturated specification. When disaggregating this effect by subject-specific standardized test scores, we see that this improvement in student performance is largely driven by improvements in mathematics (0.14 standard deviations), for which the effect is almost twice as large as for reading (the latter of which is not statistically significant). University aspirations are also positively affected, at 0.08 standard deviations, though not statistically significant in aggregate.

The fundamental problem in estimating the causal effect of attending a certain school, or school track, is the non-random sorting of students based on student and parent preferences. If certain schools attract better students, higher test score attainment at these schools cannot be attributed to the quality of the school in a value-added sense. Naïve estimation strategies will therefore lead to an overestimation of the value added. This is further exacerbated by omitted variable bias when other factors, i.e., soft skills, ability, persistence, and motivation, are unobserved, but are likely correlated with student preferences. By exploiting the structure of the centralized assignment mechanism, we obtain causal estimates that are not biased by these channels.

Following Abdulkadıroğlu et al., 2022 and comparing the results of the fully specified model in column (3) of Table 6 with OLS results in column (4), we are able to characterize the size of the bias that would otherwise affect these results. The OLS estimates in Table 6 are computed based on a sample that includes the universe of 8th-grade students and are obtained by omitting propensity score controls without considering a student's local assignment risk. The OLS results indicate that approximately two-fifths of the performance differential between students who are offered a place in a high-track program and those who are not can be explained by better students selecting into the highest track.

## 5.1 Who benefits most from tracking?

More closely examining heterogeneity in the effects of high-track attendance demonstrates several key findings. Figure 5 shows that the main results are not driven by baseline differences between students in terms of SES or prior achievement.

Rather, conditional marginal effects on 10th-grade average test scores demonstrate very similar effects on average standardized test scores for low, middle, and high SES students, as well as those with and without college-educated parents. We also find similar effects across the distribution of baseline achievement, as measured in 8th grade. This suggests that it is not only high-ability students who benefit from high-track attendance and that, when it comes to test scores, attending the highest track has a more universal effect on achievement.

Disaggregating the results by subject test type, Figure 6 also suggests that students with low baseline achievement in particular may actually benefit the most from high-track attendance in terms of mathematics scores, although there is some evidence that males in particular benefit from improvements in reading test scores, while female students do not. Finally, Figure 7 suggests that the effects of high-track attendance on university aspirations are similar, independent of parental higher education attendance, though there is some suggestive evidence that females, students with low baseline academic achievement, and those from a low SES benefit more.

In general, there are notable differences by gender for both test scores and university aspirations. We further explore these in Table 7, which reveals that while the average effect on mathematics is large and highly statistically significant, this is particularly true for male students at 0.18 standard deviations. For female students, the effect is much smaller at 0.11 standard deviations, and not statistically significant. Similarly, there are large effects on reading test scores evident for males (0.16 standard deviations) but no effect for females. Conversely, the effect on university aspirations is both large and statistically significant for female students, at 0.13 standard deviations, but nonexistent for males. Overall, heterogeneity analyses reveal estimates that are broadly similar to the main effects; estimates are positive, statistically significant, and affect both test scores and future aspirations. That low SES and lower ability students benefit at least as much from high-track attendance is of particular interest, and consistent with earlier literature that suggests students not admitted to the highest track are disproportionately harmed by tracking (e.g., Reichelt and Eberl, 2019), given we observe differences in track accessions by socioeconomic background characteristics.

## 5.2 Mechanisms

In Table 8, we examine potential moderating factors to disentangle the mechanisms that could be driving the observed effects. The table displays the results of a series of regressions of 10th-grade student characteristics on an indicator of whether a student received an offer for a high-track program. First, using the peer leave-own-out mean calculated at the class level, columns (1–3) show differences in baseline 8th-grade test scores, gender, and SES, respectively. The results suggest that there are only minor differences in terms of gender composition, though female students attend the highest track at a slightly higher rate. There are differences in terms of SES, however, indicating that students who attend high-track programs tend to have a higher level of SES, on average. Further, column (1) suggests that average peer quality in terms of baseline achievement differs between high-track programs and intermediate-track programs by 0.17 standard deviations.

Existing evidence from the literature suggests that peer effects may lead to educational spillovers (see, e.g., Denning et al., 2023; Carrell and Kuka, 2018; Dobbie and Fryer, 2014; Ding and Lehrer, 2007; among others), and in part, the results obtained thus far may be driven by classroom or cohort-composition effects. Educational spillovers occur when the characteristics of peers in the same classroom, program, school, or broader social network affect students' own outcomes, though the direction of the effect may differ conditional on exposure to peers with certain characteristics. Some peers may generate positive spillovers, improving the outcomes of their peers following sustained exposure, while others may generate negative spillovers. The specific channel we test in the following is whether students assigned to the highest track benefit more from the presence of so-called "high-quality peers".

That is, we test whether peer composition at the program level has a causal effect on students' own performance by estimating the effect of attending a high-track program with an above-median peer quality, measured using the leave-own-out average of the baseline achievement distribution. We employ a similar empirical framework to the main results, but rather than the margin of admission to the highest track we estimate effects at the margin of admission to an above-median peer quality high-track program, conditional on common empirical bandwidths of $\delta_N = 0.38$, $\delta_N = 0.42$ and $\delta_N = 0.38$ for peer quality measured in terms of average standardized test scores, mathematics, and reading, respectively.[9] Simultaneously, we restrict the sample only to those students who are sure to attend the highest

[9]See Figure 8 for sample bandwidth sensitivity analyses. We apply the same selection criteria used to determine the bandwidth for the main results, and, in addition to the placebo test reported in Figure 8 (see also Panel C of Table 9), we ensure balance on students' own 8th-grade test scores pre-assignment.

track. This allows us to exploit quasi-experimental variation to disentangle peer effects, given, as mentioned in Section 4, there is a large degree of variation in program-specific admissions cutoffs for high-track programs.

Table 9 reports the results of this analysis with fully saturated propensity score controls. Estimates from the fully saturated specification in column (3) show that attending a school with higher-quality peers does not significantly improve student test scores in a causal sense, and the effect is close to zero in magnitude despite a relatively strong potential mechanism (Panel A). In Table 10, similar to the main specification, we disaggregate these effects by subject and repeat the analysis while distinguishing between male and female students. For mathematics and reading, respectively, the "at-risk" samples of students exposed to high-quality peers were constructed with respect to peer performance in the relevant subject. Panels A and D utilize samples based on average peer test scores. Although overall peer effects are either small or not particularly statistically significant, we find some suggestive evidence that males benefit from high-quality peers in terms of reading (0.16 of a standard deviation). In general, however, it is unlikely that the effects of high-track attendance estimated at the margin of high-track admissions in the main results are driven by peer quality effects in terms of academic ability, given the main effects predominantly operate via the channel of mathematics test scores (see Table 6).

We then test two alternative channels of peer effects, given that a more recent literature suggests peer effort and behavior matter at the classroom level (see, e.g., Adamopoulou et al., 2024; Dong et al., 2023; Bietenbeck, 2020), and may even have long-run consequences (e.g., Carrell and Kuka, 2018). Using pre-tracking behavior grades, as well as diligence grades as a proxy for grit, we again employ a similar empirical framework to the main results while restricting the sample only to those who are sure to attend the highest track, conditional on common empirical bandwidths of $\delta_N = 0.49$ and $\delta_N = 0.5$ for behavior and diligence, respectively.[10] Tables 11 and 12 report the effects of attending a program with above-median peer behavior grades on average standardized test scores and subject-specific test scores, respectively, while Tables 13 and 14 report those for above-median peer diligence.

The results suggest that peer behavior, in particular, is important not only on average (0.11 standard deviations), but that this is especially true for female

[10]See Figure 9 for sample bandwidth sensitivity analyses. We again apply the same selection criteria used to determine the bandwidth for the main results, and, in addition to the placebo test reported in Figure 9 (see also Panel C of Tables 11 and 13 for behavior and diligence, respectively), we ensure balance on students' own 8th-grade test scores pre-assignment.

students' mathematics scores, with an effect that is both large in magnitude (0.15 standard deviations) and highly statistically significant. Effects are also positive for reading, though much smaller in magnitude and not statistically significant for males. Given that heterogeneity analyses conducted as part of the main results reveal the effects of high-track attendance operate primarily through the channel of mathematics (see Panel B of Table 6) and are much larger for males (see Panel B of Table 7), the positive effect of peer behavior on female mathematics scores in particular is unlikely to be wholly driving the main results. It is, nevertheless, an important finding, given that on the one hand it may imply a reconsideration of priorities when deciding who is offered a place in the higher track, and on the other has important implications for addressing the STEM achievement gap between male and female students.

Finally, diligence appears to be unimportant in general despite a reasonably large potential mechanism (see Panel A of Table 13), and we do not find statistically significant effects for either gender across both test subjects (see Table 14). This suggests that the absence of peer disruption potentially has a greater effect on own test scores than conformity with the study habits of hardworking peers. The exception is the effect of peer diligence on the university aspirations of males, which are both large ($-0.17$ standard deviations) and statistically significant. Though we are unable to explicitly test for this, one potential explanation is that the presence of hard-working peers exerts a discouragement effect on male students in particular, given the competitive nature of university admissions in Hungary, if they are more likely to compare themselves to their peers when thinking about their future plans.

Overall, we find only limited evidence that the main results are driven by the presence of higher-quality peers in the highest track, at least in terms of academic ability. The subject and gender disaggregation in the peer effect analysis of those always seated at high-track programs shows that academic spillovers in terms of peer ability occur mostly via the channel of male reading scores, though in the main results, the measured effect occurs primarily via the channel of mathematics test scores. However, peer behavior does appear to be generally important, especially for female students' mathematics scores.

# 6 Robustness Tests

In the following, we perform two separate robustness tests for the results obtained thus far. In the first, we re-compute the main results for an alternative sample constructed using a local bandwidth permitted to vary flexibly across individual programs. In the second, we rule out that estimates obtained via heterogeneity analyses in the main results are driven mechanistically by systematic differences in program quality by, e.g., tercile of SES or baseline achievement, among others, for students who "lose" the assignment lottery.

## 6.1 Local Bandwidth Selection

First, we alternatively construct the selected sample (and corresponding local propensity score) by computing locally optimal bandwidths, following the extensions to Imbens and Kalyanaraman (2012) proposed in Calonico et al. (2017) and Calonico et al. (2019) to determine the mean square error (MSE) optimal bandwidth choice. The results of this exercise are presented in Table 16. In general, results are similar to the main results in terms of direction and significance, though they are larger in magnitude, e.g., the effects on mathematics test scores are 0.24 and 0.20 standard deviations for male and female students, respectively, versus 0.18 and 0.11 in the main results. Similarly, the effects on reading test scores are 0.18 and 0.03 standard deviations for male and female students, respectively, versus 0.16 and 0.003 in the main results (though the estimate for female students is not statistically significant, as in the main specification). In general, the results computed using the common bandwidth remain our preferred specification. The sample of "at-risk" students constructed using a local bandwidth is not only approximately 10% smaller than the main sample (leading to reduced power for the heterogeneity analyses), but the balancing is comparatively worse (see Table 15). However, it is nevertheless a reassuring exercise given the broad similarity of the estimated coefficients.

## 6.2 Testing for Mechanistic Effects

Finally, we verify that the heterogeneous effects reported in Section 5.1 are not an artifact arising from different groups of students encountering a larger quality gap between the high- and low-track programs available to them. In general, differences by, e.g., prior academic attainment, do not *necessarily* imply that lower-ability students benefit more from attending the highest track because of some feature of the high-track program. Rather, the positive effect of high-track attendance for lower-baseline-ability students could potentially be explained by the fact that the gap in quality between the high-track program they got into and their closest non-high-track alternative is larger than it is for higher-ability students. To explicitly rule this out, we test for school quality differences conditional on baseline achievement, among other factors.

For every student in the analytical sample we generate a counterfactual assignment to the "nearest" alternative program—defined by the smallest absolute difference in the student's preference rank relative to their actual matched program—subject to two conditions: (i) the student must be admissible at that alternative program, and (ii) the alternative must belong to the opposite track type (high-track vs. non-high-track) than the actual match. When two equidistant alternatives exist (one ranked higher and one lower than the actual assignment), ties are broken at random. We then proxy program quality with the counterfactual class-average leave-own-out baseline test score as a proxy for program quality.

The results of this exercise are presented in Figures 10 and 11, which illustrate, for mathematics and reading, respectively, the distribution of differences in program quality between an individual's matched program and their counterfactually assigned program by terciled 8th-grade baseline achievement. On average, there does not appear to be any notable divergence between the distributions of differences in program quality by achievement tercile, though the regional disaggregation suggests small variations in some regions. Similarly, Figure 12 shows no notable divergence in the distributions of differences in program quality for matched vs. counterfactually assigned programs by SES, gender, or settlement type (i.e., it is unlikely that in more rural areas, the quality of the alternative program should a student "lose" the assignment lottery is significantly worse than in urban areas). We are therefore reasonably confident that systematic differences in school quality should a student "lose" the assignment lottery are not driving the observed results.

# 7 Discussion

In this paper, we exploit a unique institutional context and employ a novel identification strategy to demonstrate that high-track program attendance significantly improves student performance two years post-enrollment. By the 10th grade, students in high-track programs exhibit markedly higher mathematics proficiency than comparable students attending intermediate-track programs, and for males, we also find positive effects on reading proficiency. While the previous literature often attributes such gains to academic peer effects, our findings challenge this narrative. Specifically, we find only limited evidence that peer ability affects 10th-grade average test scores in a causal sense. However, we do find evidence that peer behavior exerts positive spillover effects on individual performance, although these effects are unlikely to be wholly driving the main results.

These findings are particularly salient in the Hungarian educational context. We find that high-track placement plays an important role in shaping long-term educational attainment independently of prior academic performance, gender, or socioeconomic status, wherein those from more deprived backgrounds or with comparably worse prior achievement benefit from high-track assignment at least as much as high-performing or relatively well-off peers. Nevertheless, there is unequal accession to the highest track conditional on socioeconomic background, among other factors. This is important as, notably, both high- and intermediate-track students are eligible to take the 12th-grade maturity exam, a prerequisite for university admission. Moreover, vocational specialization in the intermediate track only begins after 10th grade. During the two-year period under consideration here, instruction in both tracks is ostensibly oriented toward general education, and curricular differences should be minimal. The findings here thus carry important implications for educational policy design.

Critics of tracking argue that it exacerbates educational inequality by disadvantaging students assigned to lower tracks. Our results support this interpretation, and we find no evidence of the common narrative that more able students benefit most from high-track attendance. Furthermore, given that peer behavior, rather than peer ability, appears to exert positive spillovers on individual outcomes, expanding access to high-track programs for lower-achieving students is, on the one hand, unlikely to negatively affect the performance of their more capable peers, and on the other hand students with strong behavior records may positively influence classroom dynamics, even in the absence of high prior academic performance. Taken together, these findings imply a potential rethinking of assignment priorities in tracked systems.

## References


**Aakvik, Arild, Kjell G. Salvanes, and Kjell Vaage.** 2010. "Measuring heterogeneity in the returns to education using an education reform." *European Economic Review* 54 (4): 483–500.

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

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

**Abdulkadıroğlu, Atila, Joshua Angrist, and Parag Pathak.** 2014. "The elite illusion: Achievement effects at Boston and New York exam schools." *Econometrica* 82 (1): 137–196.

**Abdulkadıroğlu, Atıla, Joshua D. Angrist, Yusuke Narita, and Parag Pathak.** 2022. "Breaking Ties: Regression Discontinuity Design Meets Market Design." *Econometrica* 90 (1): 117–151.

**Adamopoulou, Effrosyni, Yaming Cao, and Ezgi Kaya.** 2024. "Gritty peers." *IZA Discussion Paper No. 17446.*

**Bach, Maximilian.** 2023. "Heterogeneous responses to school track choice: Evidence from the repeal of binding track recommendations." *Economics of Education Review* 95 : 102412.

**Barrera-Osorio, Felipe, and Deon Filmer.** 2016. "Incentivizing schooling for learning: Evidence on the impact of alternative targeting approaches." *Journal of Human Resources* 51 461–499.

**Barrios-Fernandez, Andrés.** 2023. "Peer effects in education." *Oxford Research Encyclopedia of Economics and Finance.*

**Barrow, Sartain Lauren, Lisa, and Marisa de la Torre.** 2020. "Increasing access to selective high schools through place-based affirmative action: Unintended consequences." *American Economic Journal: Applied Economics* 12 (4): 135–16.

**Berkowitz, Daniel, and Mark Hoekstra.** 2011. "Does high school quality matter? Evidence from admissions data." *Economics of Education Review* 30 (2): 280–288.

**Beuermann, Diether, and Kirabo Jackson.** 2022. "The short and long-run effects of attending the schools that parents prefer." *The Journal of Human Resources* 57 (3): 725–746.

**Bietenbeck, J.** 2020. "Own motivation, peer motivation, and educational success." *IZA DP No. 13872.*

**Biró, Péter.** 2008. "Student admissions in Hungary as Gale and Shapley envisaged."Technical report, University of Glasgow.

**Borghans, Lex, Ron Diris, Wendy Smits, and Jannes de Vries.** 2020. "Should we sort it out later? The effect of tracking Age on long-run outcomes." *Economics of Education Review* 75 101973.

**Borghans, Lex, Ron Diris, Wendy Smits, and Jannes de Vries.** 2019. "The long-run effects of secondary school track assignment." *PLOS ONE* 14 (10): 1–29.

**Bukodi, Erzsébet, Péter Róbert, and Szilvia Altorjai.** 2008. "The Hungarian educational system and the implementation of the ISCED-97." *The International*

*Standard Classification of Education* 200–215.

**Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocío Titiunik.** 2017. "rdrobust: Software for regression discontinuity designs." *Stata Journal* 17 372–404.

**Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocío Titiunik.** 2019. "Regression discontinuity designs using covariates." *The Review of Economics and Statistics* 101 442–451.

**Canaan, Serena.** 2020. "The long-run effects of reducing early school tracking." *Journal of Public Economics* 187 104206.

**Carrell, Mark Hoekstra, Scott, and Elira Kuka.** 2018. "The long-run effects of disruptive peers." *American Economic Review* 108 (11): 3377–3415.

**Colas, Mark, Sebastian Findeisen, and Dominik Sachs.** 2021. "Optimal need-based financial aid." *Journal of Political Economy* 129 492–533.

**Cullen, Julie Berry, Brian A Jacob, and Steven Levitt.** 2006. "The effect of school choice on participants: Evidence from randomized lotteries." *Econometrica* 74 (5): 1191–1230.

**Denning, Jeffrey T., Richard Murphy, and Felix Weinhardt.** 2023. "Class rank and long-run outcomes." *Review of Economics and Statistics* 105 (6): 1426–1441.

**Ding, Weili, and Steven Lehrer.** 2007. "Do peers affect student achievement in China's secondary schools?" *The Review of Economics and Statistics* 89 (2): 300–312.

**Dobbie, Will, and Jr. Fryer, Roland G.** 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 (3): 58–75.

**Dong, Xiaoqi, Yinhe Liang, and Shuang Yu.** 2023. "Middle-achieving students are also my peers: The impact of peer effort on academic performance." *Labour Economics* 80 102310.

**Dräger, Jascha, Thorsten Schneider, Melanie Olczyk et al.** 2024. "The relevance of tracking and social school composition for growing achievement gaps by parental education in lower secondary school: A longitudinal analysis in France, Germany, the United States, and England." *European Sociological Review* 40 964–980.

**Duflo, Esther, Pascaline Dupas, and Michael Kremer.** 2011. "Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya." *American Economic Review* 101 (5): 1739–1774.

**Dustmann, Christian, Patrick A. Puhani, and Uta Schönberg.** 2017. "The long-term effects of early track choice." *The Economic Journal* 127 (603): 1348–1380.

**van Elk, Roel, Marc van der Steeg, and Dinand Webbink.** 2011. "Does the timing of tracking affect higher education completion?" *Economics of Education Review* 30 (5): 1009–1021.

**Ferraro, Simona, and Kaire Põder.** 2018. "School-level policies and the efficiency and equity trade-off in education." *Journal of Policy Modeling* 40 1022–1037.

**Guyon, Nina, Eric Maurin, and Sandra McNally.** 2011. "The effect of tracking students by ability into different schools: A natural experiment." *Journal of Human Resources* 47.

**Guyon, Nina, Eric Maurin, and Sandra McNally.** 2012. "The effect of tracking students by ability into different schools a natural experiment." *Journal of Human resources* 47 (3): 684–721.

**Hall, C.** 2012. "The effects of reducing tracking in upper secondary school evidence from a large-scale pilot scheme." *Journal of Human Resources* 47 (1): 237–269.

**Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rivkin.** 2003. "Does peer ability affect student achievement?." *Journal of Applied Econometrics* 18 (5): 527–544.

**Hanushek, Eric A., and Ludger Wössmann.** 2006. "Does educational tracking affect performance and inequality? Differences- in-Differences evidence across countries." *The Economic Journal* 116 (510): 63–76.

**Hastings, Justine, and Jeffrey Weinstein.** 2008. "Information, school choice, and academic achievement: Evidence from two experiments." *The Quarterly Journal of Economics* 123 (4): 1373–1414.

**Hoxby, Caroline.** 2000. "Peer effects in the lassroom: Learning from gender and race variation." *NBER Working Paper No. 7867.*

**Imbens, Guido, and Karthik Kalyanaraman.** 2012. "Optimal bandwidth choice for the regression discontinuity estimator." *Review of Economic Studies* 79 (3): 933–959.

**Jackson, C. Kirabo.** 2010. "Do students benefit from attending better schools? Evidence from rule-based student assignments in Trinidad and Tobago." *The Economic Journal* 120 1399–1429.

**Lucas, Adrienne M., and Isaac M. Mbiti.** 2012. "Access, sorting, and achievement: The short-run effects of free primary education in Kenya." *American Economic Journal: Applied Economics* 4 (4): 226–53.

**Malamud, Ofer, and Cristian Pop-Eleches.** 2011. "School tracking and access to higher education among disadvantaged groups." *Journal of Public Economics* 95 (11): 1538–1549, Special Issue: International Seminar for Public Economics on Normative Tax Theory.

**Matthewes, Sönke Hendrik.** 2020. "Better together? Heterogeneous effects of tracking on student achievement." *The Economic Journal* 131 (635): 1269–1307.

**Meghir, Costas, and Mårten Palme.** 2005. "Educational reform, ability, and family background." *American Economic Review* 95 (1): 414–424.

**Pekkala Kerr, Sari, Tuomas Pekkarinen, and Roope Uusitalo.** 2013. "School tracking and development of cognitive skills." *Journal of Labor Economics* 31 (3): 577–602.

**Piopiunik, Marc.** 2014. "Intergenerational Transmission of Education and Mediating Channels: Evidence from a Compulsory Schooling Reform in Germany." *The Scandinavian Journal of Economics* 116 (3): 878–907.

**Reichelt, Matthias Collischon, Malte, and Andreas Eberl.** 2019. "School tracking and its role in social reproduction: Reinforcing educational inheritance and the direct effects of social origin." *The British Journal of Sociology* 70 (4): 1323–1348.

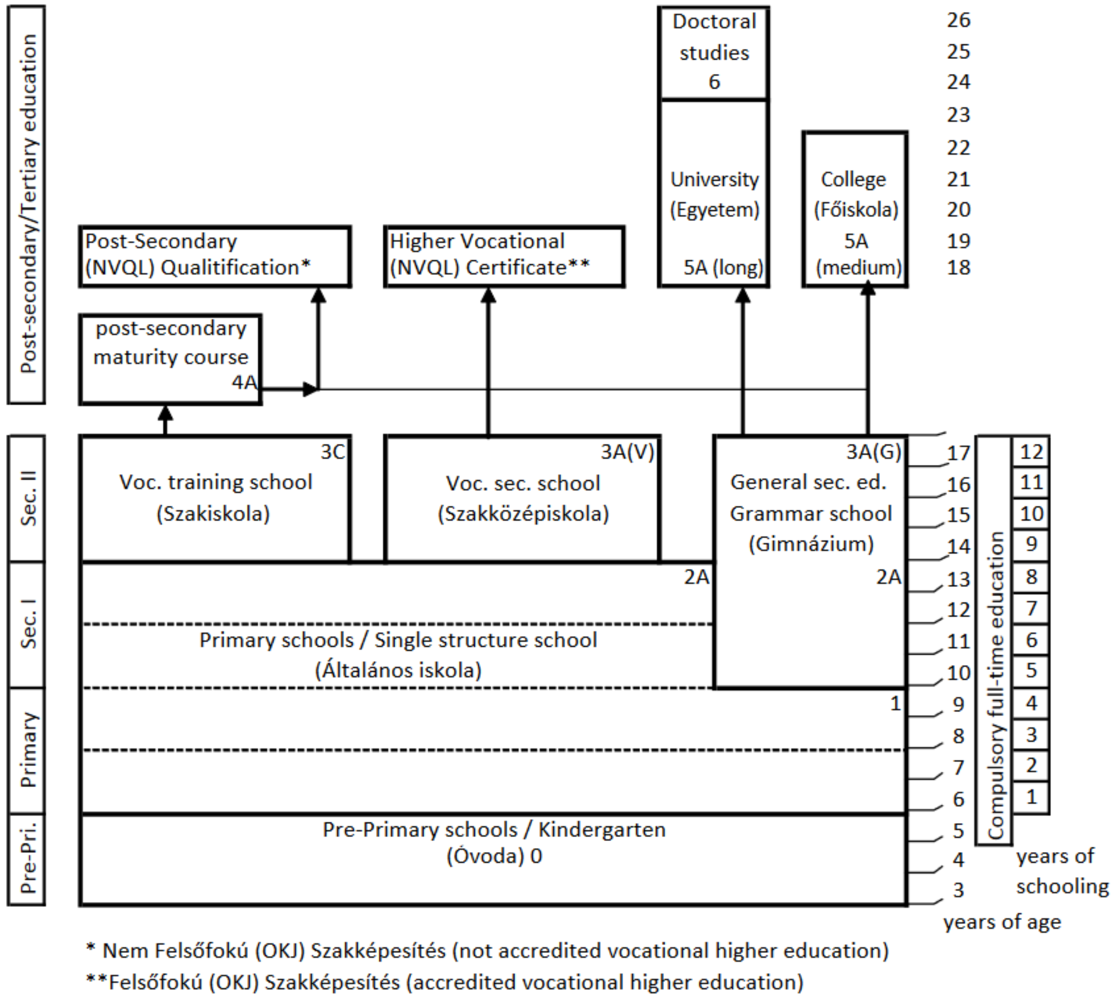
**Roller, Marcus, and Daniel Steinberg.** 2020. "The distributional effects of early school stratification - non-parametric evidence from Germany." *European Economic Review* 125 103422.

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

**Woessmann, Ludger.** 2008. "Efficiency and equity of European education and training policies." *International Tax and Public Finance* 15 199–230.

# Figures

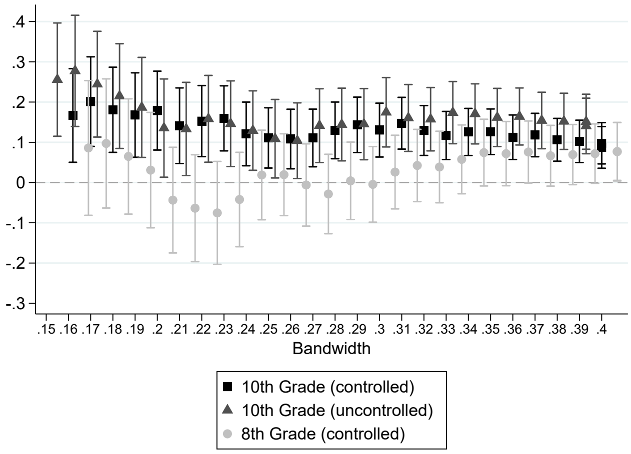
Figure 1: The Structure of the Hungarian Education System



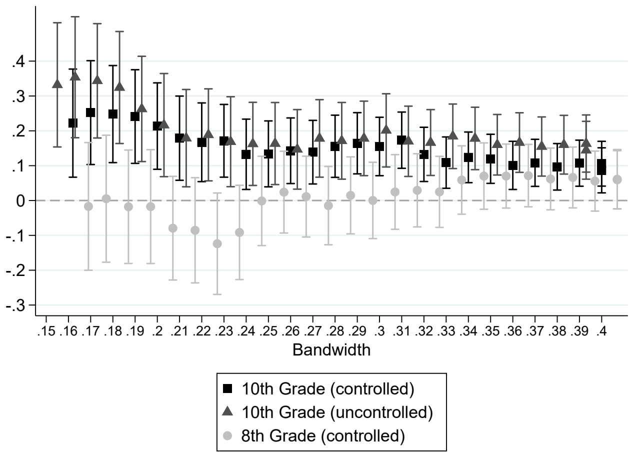

*Notes:* The figure illustrates the structure and potential progression pathways for the Hungarian educational system from pre-primary to tertiary education. *Source:* Bukodi et al. (2008).

Figure 2: Bandwidth Sensitivity Analysis

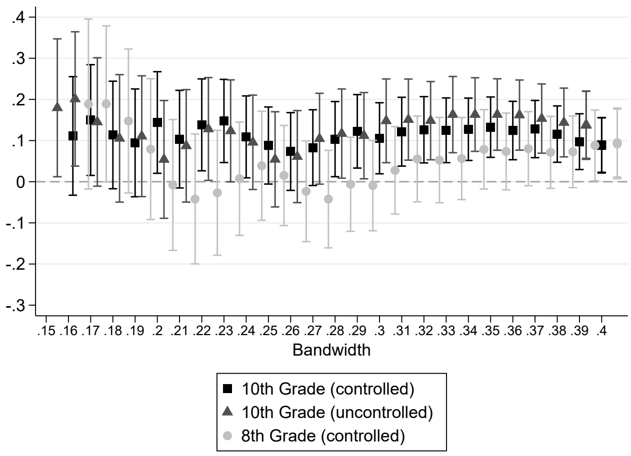
Panel (a): Average Test Score




Panel (b): Mathematics




Panel (c): Reading

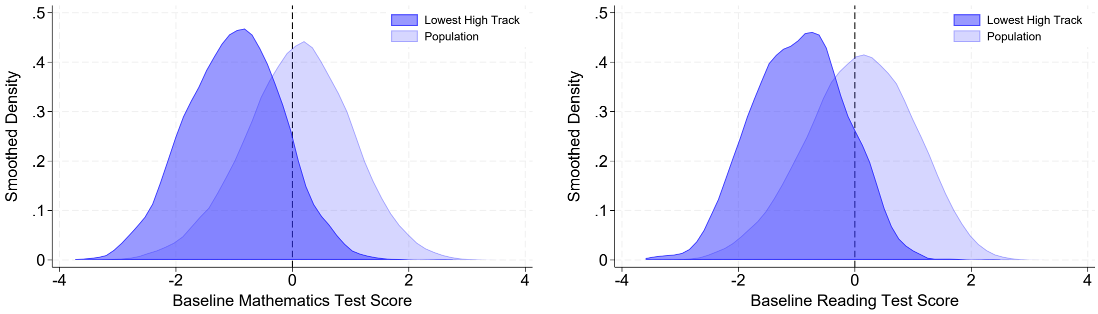



*Notes:* Figure 2 demonstrates the sensitivity of the estimation procedure used to compute the main results to the choice of bandwidth. It depicts the estimated coefficient for a) 10th-grade outcomes in the fully saturated model, b) uncontrolled 10th-grade outcomes, and c) a placebo test using 8th-grade outcomes. Controls include the high-track propensity score, RDD controls, and the main controls described in Table 4.

Figure 3: Distribution of 8th-Grade Standardized Test Scores for the Lowest Scoring Students Admitted to the Highest Track

Panel (a): Mathematics

Panel (b): Reading

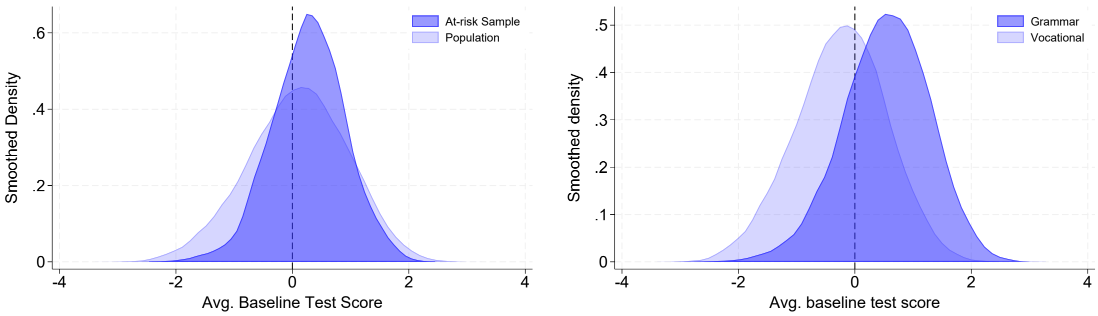



*Notes:* For individual programs, Figure 3 plots the distribution of kernel density estimates for the minimum baseline standardized test score needed to be admitted to the highest track. To contextualize this, the distribution of test scores for the overall student population is also plotted. Baseline test scores are measured in the 8th-grade, for (a) mathematics and (b) reading, respectively.

Figure 4: Distribution of 8th-Grade Average Standardized Test Scores

Panel (a): Full and At-Risk Samples

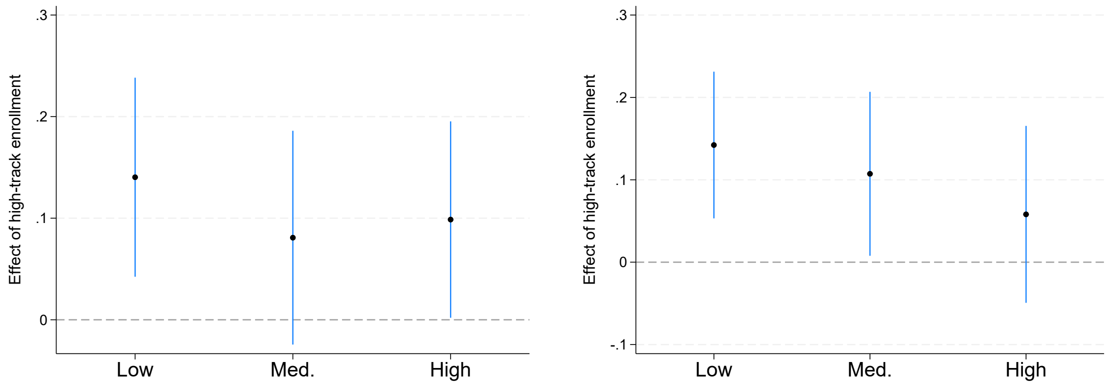
Panel (b): School Type



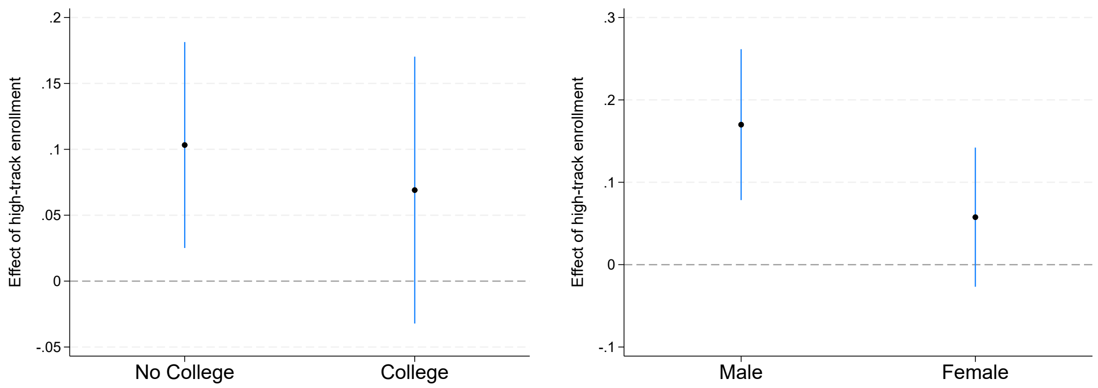

*Notes:* Figure 4 plots kernel density estimates of average baseline standardized test scores measured in the 8th-grade pre-track assignment by (a) track type, and (b) for the full and at-risk samples.

Figure 5: Heterogeneous Effects on 10th-Grade Average Test Scores

Panel (a): Baseline Test Scores Panel (b): SES (Index)




Panel (c): Parent's Education Panel (d): Gender

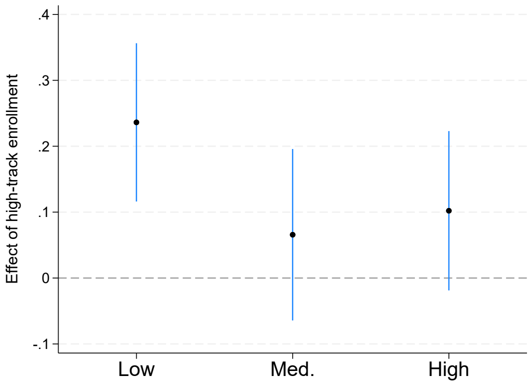



*Notes:* Figure 5 plots conditional marginal effects of high-track enrollment on 10th-grade average test scores by (a) 8th-grade baseline average achievement terciles, (b) socioeconomic status (SES), (c) whether the student's highest educated parent attended some form of college education, and (d) gender. In addition to saturated propensity score and running variable controls, coefficients were estimated controlling for student characteristics and baseline test scores in the 8th grade.

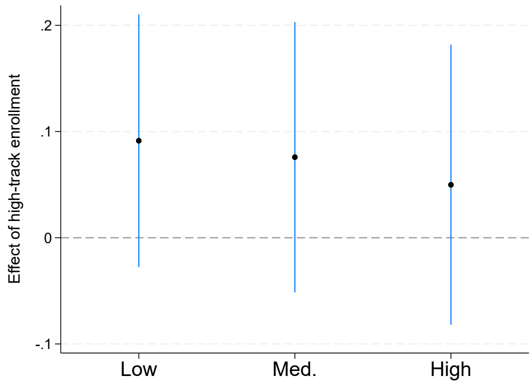
Figure 6: Heterogeneous Effects on 10th-Grade Test Scores by Subject

Panel (a): Baseline Test Scores

i. Math

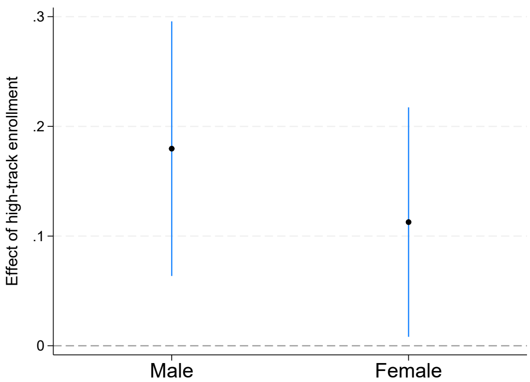



ii. Reading

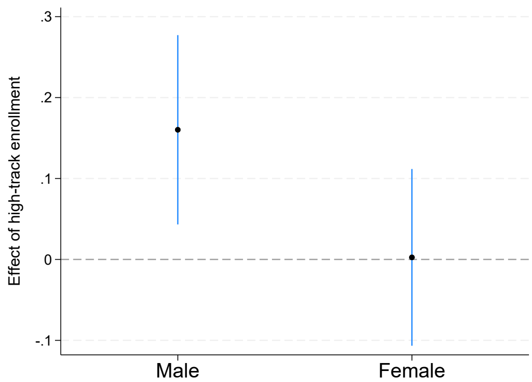



Panel (b): Gender

i. Math



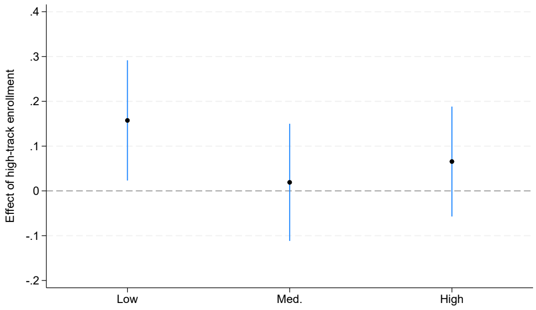

ii. Reading



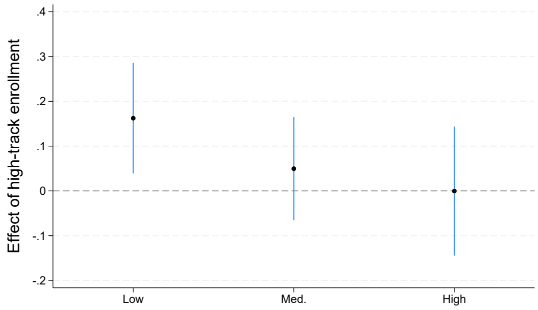

*Notes:* Figure 6 plots conditional marginal effects of high-track enrollment on 10th-grade subject-specific test scores by (a) average 8th-grade baseline achievement terciles, and (b) gender.

Figure 7: Heterogeneous Effects on 10th-Grade University Aspirations

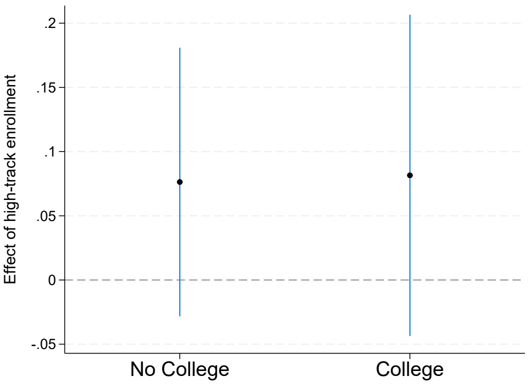
Panel (a): Baseline Test Scores



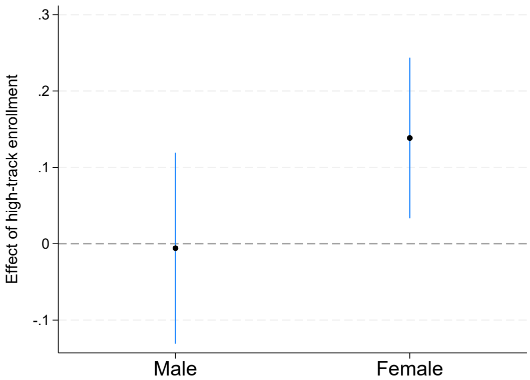

Panel (b): SES (Index)




Panel (c): Parent's Education



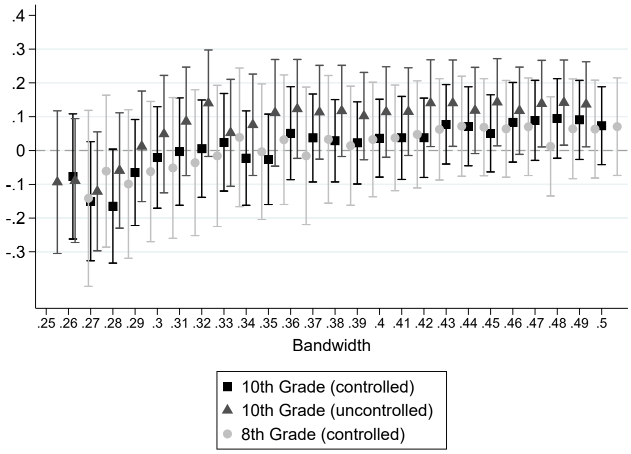

Panel (d): Gender



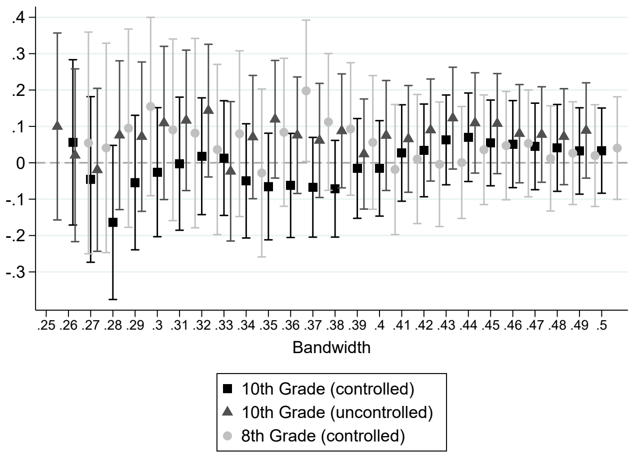

*Notes:* Figure 7 plots conditional marginal effects of high-track enrollment on 10th-grade university aspirations by (a) 8th-grade baseline average achievement terciles, (b) socioeconomic status (SES), (c) whether the student's highest educated parent attended some form of college education, and (d) gender.

Figure 8: Peer Sample Bandwidth Sensitivity Analyses

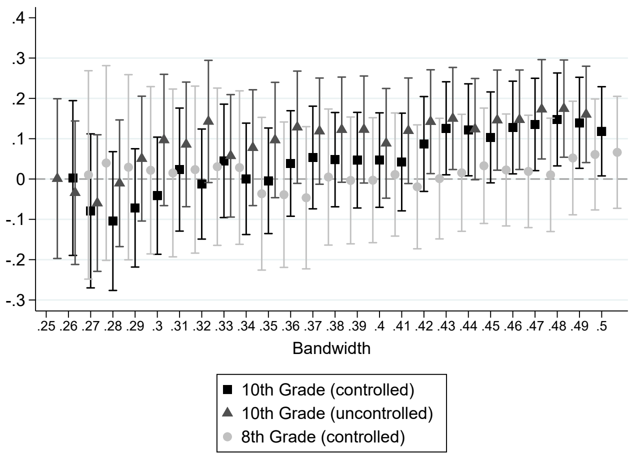
Panel (a): Above Median Average Ability




Panel (b): Above Median Mathematics Ability



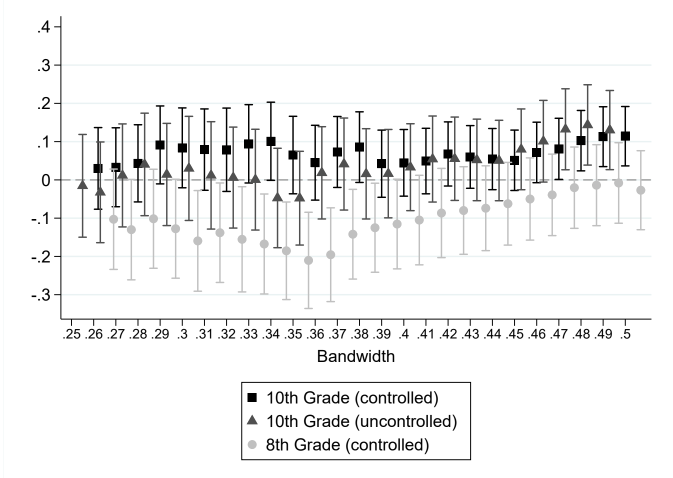

Panel (c): Above Median Reading Ability



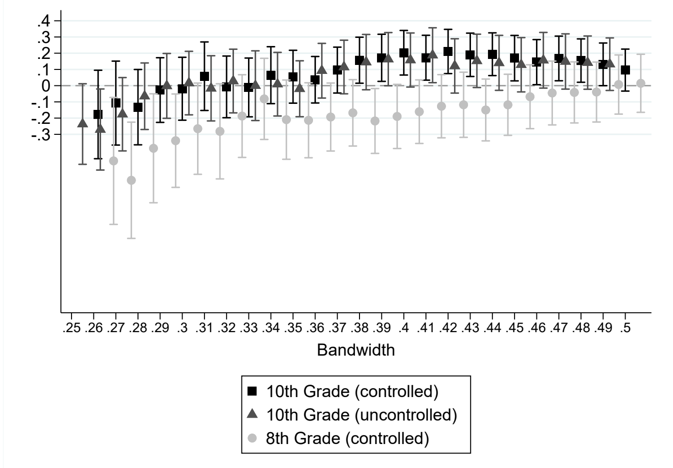

*Notes:* Figure 8 demonstrates the sensitivity of the estimation procedure used to compute the peer analyses to the choice of bandwidth for the three peer ability-based samples, for a) 10th-grade outcomes in the fully saturated model, b) uncontrolled 10th-grade outcomes, and c) a placebo test using 8th-grade outcomes. Controls include the high-track propensity score, RDD controls, and the main controls described in Table 4.

Figure 9: Alternative Peer Sample Bandwidth Sensitivity Analyses

Panel (a): Above Median Behavior

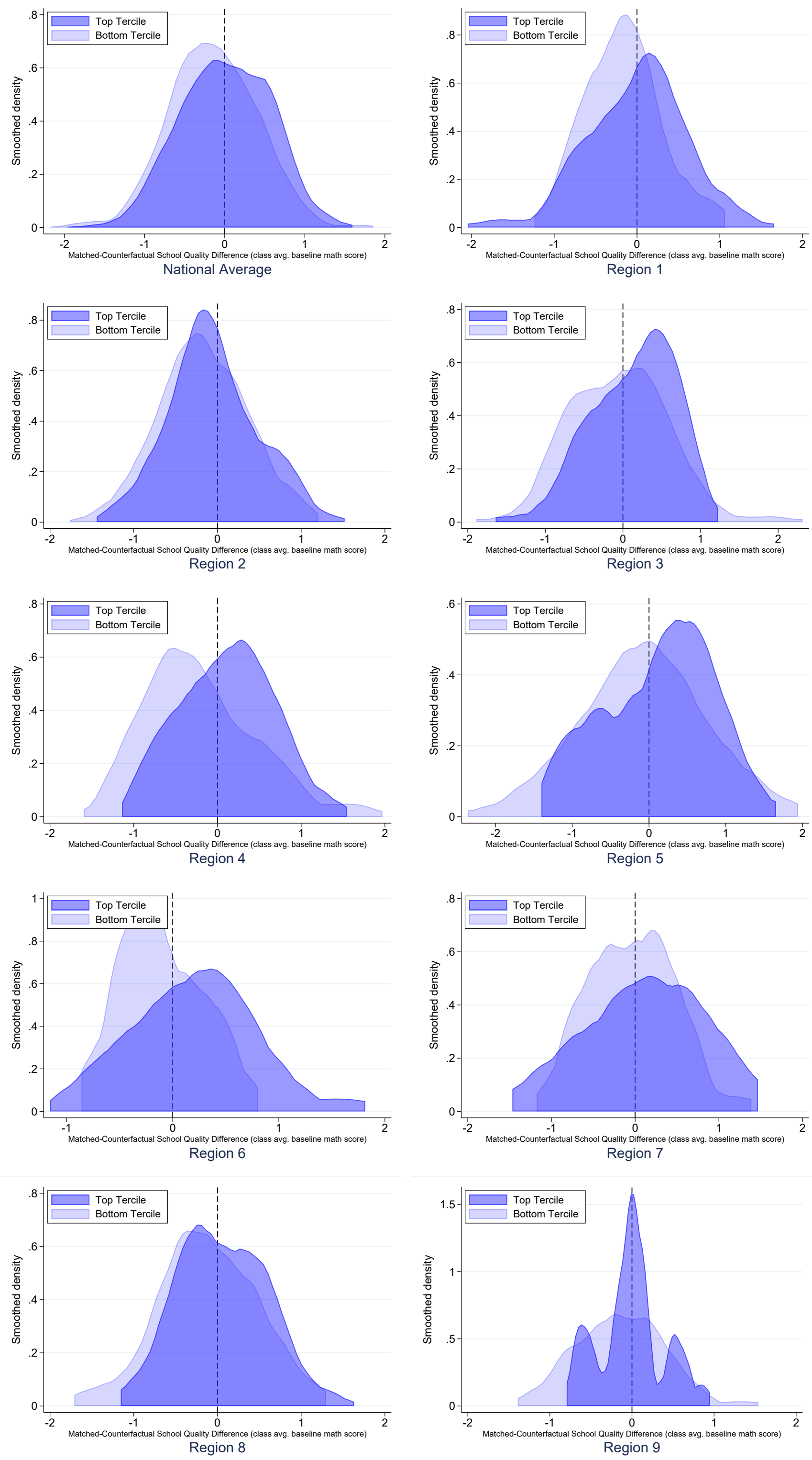



Panel (b): Above Median Diligence



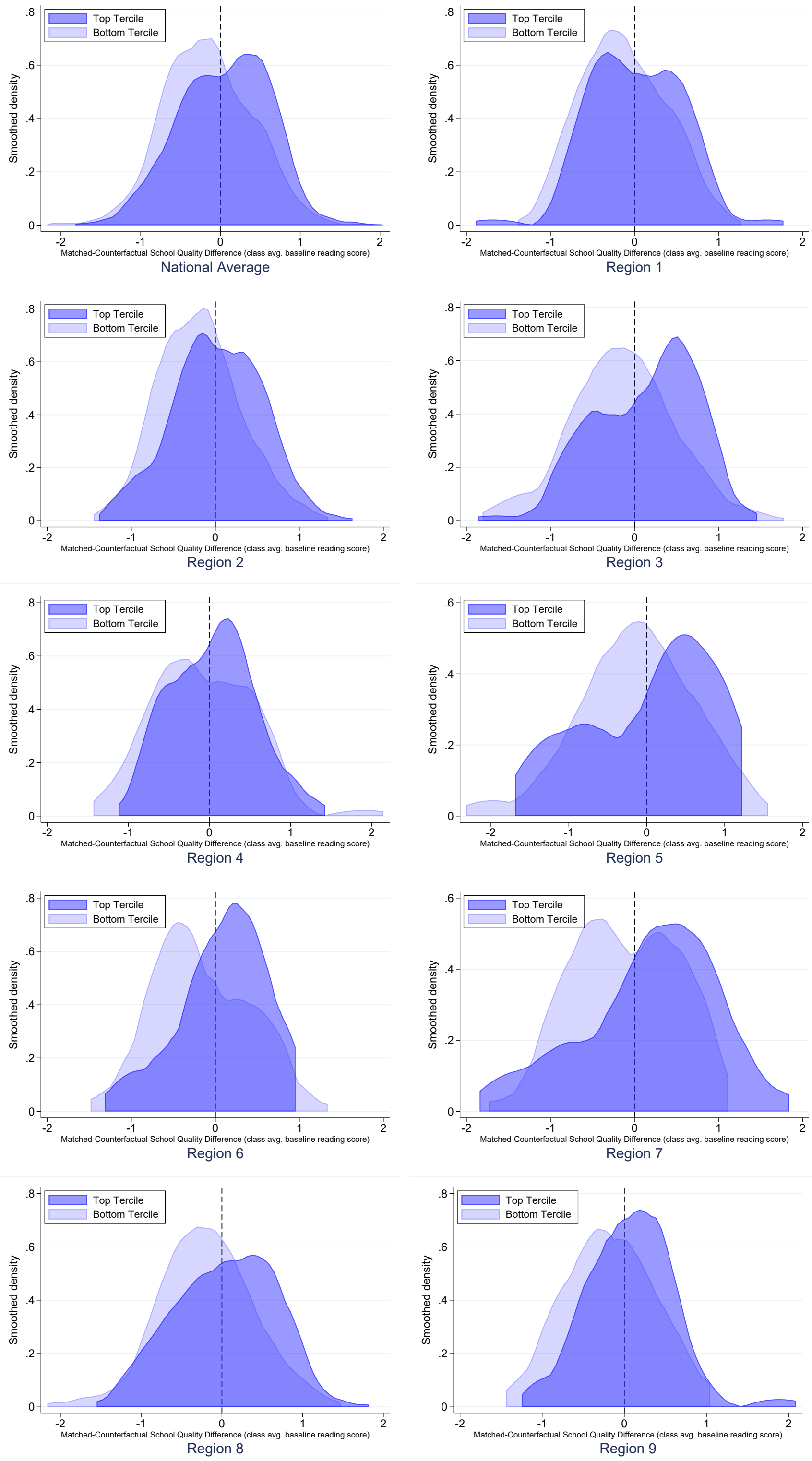

*Notes:* Figure 9 demonstrates the sensitivity of the estimation procedure used to compute the peer analyses to the choice of bandwidth for the two alternative peer samples, for a) 10th-grade outcomes in the fully saturated model, b) uncontrolled 10th-grade outcomes, and c) a placebo test using 8th-grade outcomes. Controls include the high-track propensity score, RDD controls, and the main controls described in Table 4.

Figure 10: Quality Differences Between Actual and Counterfactual Assignment by Terciled Baseline Test Scores (Mathematics)




*Notes:* The figure plots program quality differences (proxied with class-average 8th-grade baseline mathematics test scores) between a student's matched program and their counterfactual assignment for the selected sample of at-risk students. Regions (1–9) are, respectively, Budapest, Szentendre, Hatvan, Debrecen, Szolnok, Kecskemét, Sárbogárd, Székesfehérvár, and Győr.

Figure 11: Quality Differences Between Actual and Counterfactual Assignment by Terciled Baseline Test Scores (Reading)

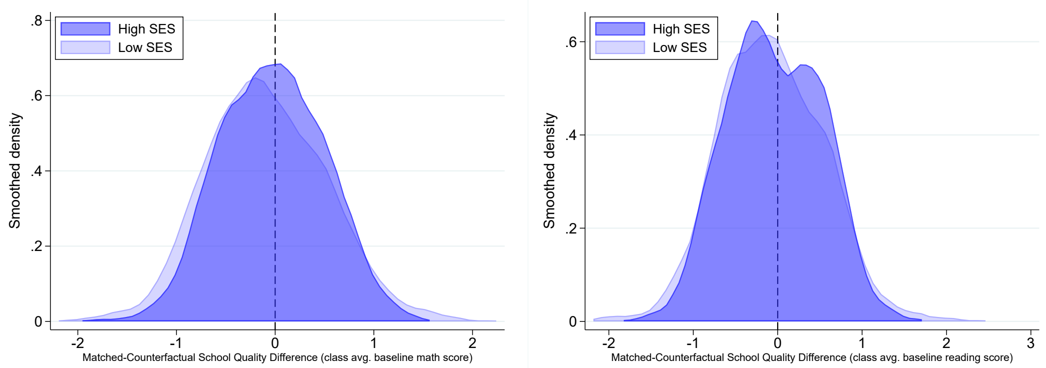



*Notes:* The figure plots program quality differences (proxied with class-average 8th-grade baseline achievement) between a student's matched program and their counterfactual assignment for the selected sample of at-risk students. Regions (1–9) are, respectively, Budapest, Szentendre, Hatvan, Debrecen, Szolnok, Kecskemét, Sárbogárd, Székesfehérvár, and Győr.

Figure 12: Quality Differences Between Actual and Counterfactual Assignment

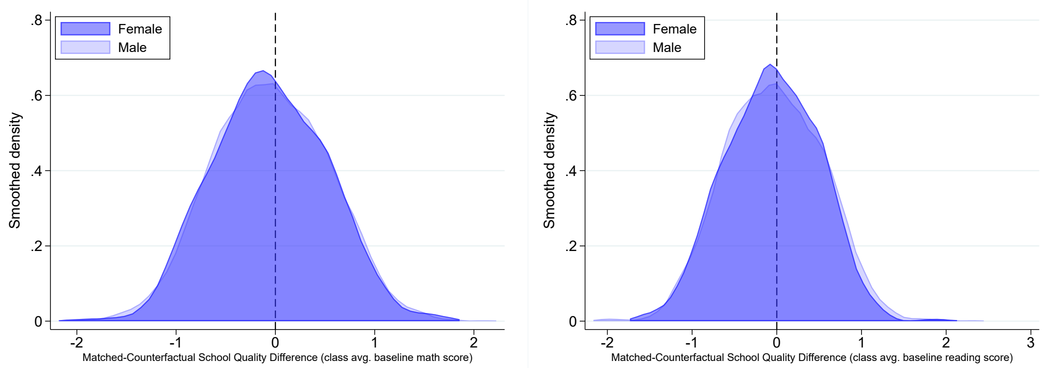
Panel (a) SES

i. Mathematics ii. Reading



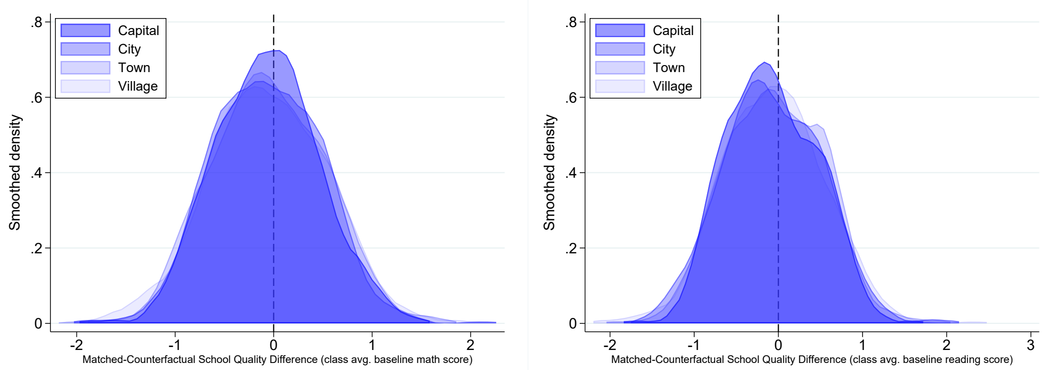

Panel (b) Gender

i. Mathematics ii. Reading




Panel (c) Settlement Type

i. Mathematics ii. Reading




*Notes:* The figure plots program quality differences (proxied with class-average 8th-grade baseline test scores) between a student's matched program and their counterfactual assignment for the selected sample of at-risk students for (a) socioeconomic status (SES), (b) gender, and (c) settlement type.

# Tables

Table 1: Testing for Selective Attrition

| | (1) | (2) | (3) |
|---|---|---|---|
| High Track | -0.027 | -0.019 | -0.021 |
| | (0.027) | (0.027) | (0.027) |
| $R^2$ | 0.414 | 0.436 | 0.440 |
| $N$ | 3,175 | 3,175 | 3,175 |
| Propensity score FE | ✓ | ✓ | ✓ |
| RDD controls | ✓ | ✓ | ✓ |
| Student controls | | ✓ | ✓ |
| Baseline test scores | | | ✓ |

*Notes:* Table 1 presents estimates from 2SLS models of high track attendance on sample attrition, where enrollment in a high track program is instrumented with receipt of a high track offer. In the fully saturated model, we control for the propensity score, running variable controls, individual student characteristics and baseline test scores in the 8th grade. The sample is limited to applicants with non-missing baseline test scores. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Sample Descriptive Statistics

| | All Applicants | At-Risk Sample |
|---|---|---|
| | (1) | (2) |
| Baseline 8th-grade test scores | | |
| Math (mean) | 0.09 | 0.25 |
| Reading (mean) | 0.09 | 0.29 |
| Demographics | | |
| Female (%) | 49.20 | 54.69 |
| Age (mean) | 16.55 | 16.52 |
| Social benefits (%) | 43.97 | 33.32 |
| Deprived neighborhood (%) | 9.07 | 6.20 |
| Single-parent (%) | 26.09 | 24.82 |
| At least 1 parent passed maturity exam (%) | 70.56 | 82.25 |
| Recieved a high track offer (%) | 41.62 | 55.72 |
| Acceptable to at least 1 high-track program (%) | 58.52 | 100.00 |
| 10th grade enrollment | | |
| High track (grammar school) (%) | 42.99 | 61.64 |
| Intermediate track (vocational secondary) (%) | 41.37 | 37.53 |
| Low track (vocational school) (%) | 15.64 | 0.83 |
| Listed schools | | |
| Listed any high-track program (%) | 61.38 | 100.00 |
| Listed high-track program first (%) | 45.06 | 82.21 |
| High track share of listed programs (mean) | 43.27 | 60.61 |
| $N$ | 54,631 | 2,518 |

*Notes:* The table presents summary statistics for the overall sample, and the sample of students in the at-risk sample who are empirically close to the assignment cut-off.

Table 3: Statistical Tests for Balance

| **Dep. Var.: High Track Offer** | (1) | (2) |
|---|---|---|
| Baseline 8th-grade math test score (std.) | 0.717*** | 0.019 |
| | (0.007) | (0.060) |
| Baseline 8th-grade reading test score (std.) | 0.828*** | -0.041 |
| | (0.007) | (0.061) |
| Female | 0.157*** | -0.064 |
| | (0.004) | (0.047) |
| Age (in years) | -0.087*** | 0.042 |
| | (0.004) | (0.038) |
| SES (CSH-index) | 0.757*** | 0.076 |
| | (0.007) | (0.067) |
| Any social benefits | -0.127*** | -0.032 |
| | (0.004) | (0.048) |
| Deprived neighborhood | -0.048*** | -0.022 |
| | (0.002) | (0.026) |
| Single-parent | -0.065*** | -0.025 |
| | (0.004) | (0.045) |
| Parent w/ maturity exam or higher | 0.262*** | 0.004 |
| | (0.004) | (0.037) |
| *N* | 51,135 | 2,518 |
| Propensity score FE | | ✓ |
| RDD controls | | ✓ |

Table 3 presents regressions of student characteristics pre-track assignment in the 8th grade on an indicator for whether the student received a high-track offer. Column (1) refers to the full sample. Column (2) refers to the selected sample of "at-risk" students with non-degenerate assignment risk, and controls both for high-track assignment risk and running variables. The bandwidth used to compute this restricted sample is 0.25, with a uniform kernel. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Summary of Control Variables Used in Main Analysis

| Category | Var. Name | Notes |
|---|---|---|
| Propensity score FE | propensity score | aggregated by rounding the nearest .05 to avoid cell sizes of 1 |
| RDD controls | distance above$_n$ | (scaled) distance below the cutoff of individual programs |
| | distance below$_n$ | (scaled) distance above the cutoff of individual programs |
| | CID$_n$ | indicator for each program applied to, used to compute fixed effects |
| Student controls | age | continuous |
| | gender | indicator var |
| | SES (CSH-Index) | standardized family background index |
| | deprived neighborhood | scale 1 (worst) - 5 (best) |
| | single-mother | indicator var |
| | child-related benefits | indicator var: discounted or free lunch, textbooks, or receipt of govt. child support |
| | grade retention | indicator var |
| | parental education | dominance approach for available parent(s), 1 (< primary) to 5 (masters +) |
| | financial | "compared to other families how well does your family live?", 1 (worst) to 5 (best) |
| Baseline test scores | mathematics | standardized 8th-grade tests |
| | english | standardized 8th-grade tests |

Table 5: First-Stage: Program Offers and High Track Accession

| | (1) | (2) | (3) |
|---|---|---|---|
| High Track Offer | 0.667*** | 0.660*** | 0.660*** |
| | (0.029) | (0.029) | (0.029) |
| $F$-statistic | 523.149 | 509.648 | 509.366 |
| $N$ | 2,518 | 2,518 | 2,518 |
| Propensity score FE | ✓ | ✓ | ✓ |
| RDD controls | ✓ | ✓ | ✓ |
| Student controls | | ✓ | ✓ |
| Test score controls | | | ✓ |

Table 5 reports the results of the first stage, wherein enrollment in the highest track is instrumented with receipt of an offer of a place in a high-track program. In the fully saturated model, we control for the propensity score, running variable controls, individual student characteristics, and 8th-grade baseline test scores. The sample is limited to applicants with non-missing baseline test scores. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: High Track Access and 10th-Grade Student Outcomes

| | 2SLS | | | OLS |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| **Panel A: Dep. Var.: 10th-Grade Average Test Scores** | | | | |
| High Track | 0.109** | 0.103** | 0.108*** | 0.181*** |
| | (0.050) | (0.050) | (0.038) | (0.005) |
| **Panel B: Dep. Var.: 10th-Grade Math Test Scores** | | | | |
| High Track | 0.164*** | 0.144** | 0.143*** | 0.165*** |
| | (0.060) | (0.058) | (0.048) | (0.006) |
| **Panel C: Dep. Var.: 10th-Grade Reading Test Scores** | | | | |
| High Track | 0.054 | 0.063 | 0.073 | 0.198*** |
| | (0.059) | (0.060) | (0.048) | (0.006) |
| **Panel D: Dep. Var.: 10th-Grade University Aspirations** | | | | |
| High Track | 0.075 | 0.078 | 0.083 | 0.256*** |
| | (0.050) | (0.051) | (0.051) | (0.005) |
| $N$ | 2,518 | 2,518 | 2,518 | 51,135 |
| Propensity score FE | ✓ | ✓ | ✓ | |
| RDD controls | ✓ | ✓ | ✓ | |
| Student controls | | ✓ | ✓ | ✓ |
| Baseline test scores | | | ✓ | ✓ |

*Notes:* Table 6 presents regressions of 10th-grade student test scores and university aspirations on high-track program attendance. Estimates in columns 1-3 are from 2SLS models, where attendance is instrumented for using receipt of a high-track offer. In addition to saturated propensity score and running variable controls, we iteratively control for student characteristics and baseline test scores in the 8th grade. Column 4 presents estimates from OLS regressions of high-track attendance on the same outcomes, using the full 2015 student sample. Both samples are limited to applicants with non-missing baseline test scores. Robust standard errors are reported in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01

Table 7: Gender Heterogeneity in High Track Access Effects

| | (1) | (2) | (3) |
|---|---|---|---|
| **Panel A: Dep. Var.: 10th-Grade Average Test Scores** | | | |
| High Track × Male | 0.186*** | 0.165*** | 0.170*** |
| | (0.051) | (0.060) | (0.047) |
| High Track × Female | 0.033 | 0.053 | 0.058 |
| | (0.053) | (0.057) | (0.043) |
| **Panel B: Dep. Var.: 10th-Grade Math Test Scores** | | | |
| High Track × Male | 0.349*** | 0.192*** | 0.180*** |
| | (0.060) | (0.070) | (0.059) |
| High Track × Female | -0.018 | 0.104 | 0.113** |
| | (0.062) | (0.065) | (0.053) |
| **Panel C: Dep. Var.: 10th-Grade Reading Test Scores** | | | |
| High Track × Male | 0.023 | 0.138* | 0.160*** |
| | (0.061) | (0.072) | (0.060) |
| High Track × Female | 0.085 | 0.002 | 0.003 |
| | (0.063) | (0.069) | (0.056) |
| **Panel D: Dep. Var.: 10th-Grade University Aspirations** | | | |
| High Track × Male | 0.025 | -0.004 | 0.001 |
| | (0.053) | (0.065) | (0.064) |
| High Track × Female | 0.122** | 0.133** | 0.138** |
| | (0.052) | (0.054) | (0.054) |
| *N* | 2,518 | 2,518 | 2,518 |
| Propensity score FE | ✓ | ✓ | ✓ |
| RDD controls | ✓ | ✓ | ✓ |
| Student controls | | ✓ | ✓ |
| Baseline test scores | | | ✓ |

*Notes:* Table 7 presents regressions of 10th-grade student test scores and university aspirations on high-track program attendance, where enrollment is instrumented for using receipt of a high-track offer interacted with students' gender. In addition to saturated propensity score and running variable controls, we iteratively control for student characteristics and baseline test scores in the 8th grade. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Mechanisms

| | 10th grade class averages (leave-own-out) | | | 10th grade percentile rank in class (normalized) | |
|---|---|---|---|---|---|
| | 8th grade test scores | Girls | SES (index) | 8th grade test scores | SES (index) |
| | (1) | (2) | (3) | (4) | (5) |
| High Track Offer | 0.168*** | 0.052*** | 0.202*** | -0.071*** | -0.106*** |
| | (0.024) | (0.015) | (0.023) | (0.014) | (0.013) |
| $R^2$ | 0.819 | 0.692 | 0.809 | 0.852 | 0.892 |
| $N$ | 2,518 | 2,518 | 2,518 | 2,518 | 2,518 |
| Propensity score FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| RDD controls | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student controls | ✓ | ✓ | ✓ | ✓ | ✓ |
| Test score controls | ✓ | ✓ | ✓ | ✓ | ✓ |

*Notes:* Table 8 presents regressions of 10th-grade student characteristics on an indicator for whether the student received a high-track offer. All estimates were computed with saturated high-track propensity scores and running variable controls, as well as controls for student characteristics and baseline test scores in the 8th grade. First, using the peer leave-own-out mean calculated at the 10th-grade class cohort level, column (1) reports differences in 8th-grade baseline test scores, column (2) reports the female classroom share, and column (3) reports class-average socioeconomic status (SES). Second, using the student's class rank in the 10th grade (normalized at the class level), column (4) reports differences in 8th-grade baseline test scores, and column (5) reports differences in socioeconomic status (SES). The sample is limited to applicants with non-missing 8th and 10th-grade test scores. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Effects of High Track Access with High Quality Peers (Peer Achievement)

| | (1) | (2) | (3) |
|---|---|---|---|
| **Panel A: Class Average 8th-Grade Test Scores (Mechanism)** | | | |
| High-Quality Peer Program | 0.412*** | 0.344*** | 0.344*** |
| | (0.025) | (0.029) | (0.029) |
| **Panel B: 10th-Grade Average Test Scores (Effect)** | | | |
| High-Quality Peer Program | 0.113 | 0.035 | 0.029 |
| | (0.071) | (0.083) | (0.062) |
| **Panel C: 8th-Grade Average Test Scores (Placebo)** | | | |
| High-Quality Peer Program | 0.047 | 0.014 | |
| | (0.079) | (0.090) | |
| *N* | 1,052 | 1,052 | 1,052 |
| Propensity score FE | ✓ | ✓ | ✓ |
| RDD controls | ✓ | ✓ | ✓ |
| Student controls | | ✓ | ✓ |
| Test score controls | | | ✓ |

*Notes:* Table 9 presents estimates from 2SLS models for enrollment in a high-track program with high-quality peers, where enrollment is instrumented with receipt of an offer for a high-track program with high-quality peers, including saturated propensity score and running variable controls. High-track programs with high-quality peers are those with above-median class-average 8th-grade test scores. Columns (2) and (3) iteratively include controls for student characteristics and baseline 8th-grade test scores, respectively. Panel A demonstrates the mechanism, or the leave-own-out 10th-grade class-level average of baseline 8th-grade test scores. Panel B reports the effect of enrollment in a high-track program with high-quality peers on a student's own 10th-grade average test scores. Panel C contains the results of a placebo test, reporting the effect of enrollment in a high-track program with high-quality peers on a student's own 8th-grade average test scores. The sample is limited to applicants with non-missing baseline test scores, a high track propensity score of one, and a high-quality peer high track propensity between one and zero. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Heterogeneous Effects of High Track Access with High Quality Peers (Peer Achievement)

| | (1) | (2) | (3) |
|---|---|---|---|
| **Panel A: Dep. Var.: 10th-Grade Average Test Scores** | | | |
| High-Quality Peer Program × Male | 0.186*** | 0.053 | 0.056 |
| | (0.071) | (0.088) | (0.067) |
| High-Quality Peer Program × Female | 0.022 | 0.016 | -0.002 |
| | (0.077) | (0.100) | (0.076) |
| *N* | 1,052 | 1,052 | 1,052 |
| **Panel B: Dep. Var.: 10th-Grade Math Test Scores** | | | |
| High-Quality Peer Program × Male | 0.306*** | 0.181 | 0.089 |
| | (0.089) | (0.110) | (0.094) |
| High-Quality Peer Program × Female | -0.187** | -0.164 | -0.075 |
| | (0.088) | (0.109) | (0.088) |
| *N* | 1,273 | 1,273 | 1,273 |
| **Panel C: Dep. Var.: 10th-Grade Reading Test Scores** | | | |
| High-Quality Peer Program × Male | 0.106 | 0.198* | 0.162* |
| | (0.084) | (0.108) | (0.085) |
| High-Quality Peer Program × Female | 0.186** | 0.044 | 0.013 |
| | (0.092) | (0.117) | (0.093) |
| *N* | 1,135 | 1,135 | 1,135 |
| **Panel D: Dep. Var.: 10th-Grade University Aspirations** | | | |
| High-Quality Peer Program × Male | 0.003 | -0.024 | -0.032 |
| | (0.054) | (0.069) | (0.069) |
| High-Quality Peer Program × Female | 0.003 | 0.005 | 0.006 |
| | (0.057) | (0.075) | (0.075) |
| *N* | 1,052 | 1,052 | 1,052 |
| Propensity score FE | ✓ | ✓ | ✓ |
| RDD controls | ✓ | ✓ | ✓ |
| Student controls | | ✓ | ✓ |
| Test score controls | | | ✓ |

*Notes:* Table 10 presents heterogeneity analyses for the results reported in Table 9, where "at risk" samples exposed to high-quality (above median) peers are constructed with respect to peer performance in the relevant subject. E.g., Panel A measures peer ability in terms of peers' average test scores, Panels B and C are based on peers' average mathematics and reading test scores, respectively. Panel D additionally uses the average test score sample. All samples are limited to applicants with non-missing baseline test scores, a high track propensity score of one, and a high-quality peer high track propensity between one and zero. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Effects of High Track Access with High Quality Peers (Peer Behavior)

| | (1) | (2) | (3) |
|---|---|---|---|
| **Panel A: Class Average 8th-Grade Test Scores (Mechanism)** | | | |
| High-Quality Peer Program | 0.098*** | 0.076*** | 0.076*** |
| | (0.027) | (0.021) | (0.021) |
| **Panel B: 10th-Grade Average Test Scores (Effect)** | | | |
| High-Quality Peer Program | 0.144*** | 0.108** | 0.113*** |
| | (0.054) | (0.053) | (0.040) |
| **Panel C: 8th-Grade Average Test Scores (Placebo)** | | | |
| High-Quality Peer Program | 0.009 | -0.008 | — |
| | (0.054) | (0.054) | — |
| *N* | 2,033 | 2,033 | 2,033 |
| Propensity score FE | ✓ | ✓ | ✓ |
| RDD controls | ✓ | ✓ | ✓ |
| Student controls | | ✓ | ✓ |
| Test score controls | | | ✓ |

*Notes:* Table 11 presents estimates from 2SLS models for enrollment in a high-track program with high-quality peers, where enrollment is instrumented with receipt of an offer for a high-track program with high-quality peers, including saturated propensity score and running variable controls. High-track programs with high-quality peers are those with above-median class-average grades for behavior measured in the 8th grade. Columns (2) and (3) iteratively include controls for student characteristics and baseline 8th-grade test scores, respectively. Panel A demonstrates the mechanism, or the leave-own-out 10th-grade class-level average of baseline 8th-grade test scores. Panel B reports the effect of enrollment in a high-track program with high-quality peers on a student's own 10th-grade average test scores. Panel C contains the results of a placebo test, reporting the effect of enrollment in a high-track program with high-quality peers on a student's own 8th-grade average test scores. The sample is limited to applicants with non-missing baseline test scores, a high track propensity score of one, and a high-quality peer high track propensity between one and zero. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Heterogeneous Effects of High Track Access with High Quality Peers (Peer Behavior)

| | (1) | (2) | (3) |
|---|---|---|---|
| **Panel A: Dep. Var.: 10th-Grade Average Test Scores** | | | |
| High-Quality Peer Program × Male | 0.161*** | 0.063 | 0.096** |
| | (0.057) | (0.062) | (0.048) |
| High-Quality Peer Program × Female | 0.126** | 0.140** | 0.125*** |
| | (0.056) | (0.059) | (0.043) |
| **Panel B: Dep. Var.: 10th-Grade Math Test Scores** | | | |
| High-Quality Peer Program × Male | 0.306*** | 0.061 | 0.095* |
| | (0.065) | (0.072) | (0.057) |
| High-Quality Peer Program × Female | 0.072 | 0.185*** | 0.154*** |
| | (0.063) | (0.066) | (0.051) |
| **Panel C: Dep. Var.: 10th-Grade Reading Test Scores** | | | |
| High-Quality Peer Program × Male | 0.020 | 0.070 | 0.103 |
| | (0.068) | (0.076) | (0.063) |
| High-Quality Peer Program × Female | 0.183*** | 0.097 | 0.098* |
| | (0.069) | (0.071) | (0.058) |
| **Panel D: Dep. Var.: 10th-Grade University Aspirations** | | | |
| High-Quality Peer Program × Male | 0.043 | 0.034 | 0.036 |
| | (0.027) | (0.030) | (0.029) |
| High-Quality Peer Program × Female | 0.055* | 0.045 | 0.045 |
| | (0.029) | (0.030) | (0.030) |
| *N* | 2,033 | 2,033 | 2,033 |
| Propensity score FE | ✓ | ✓ | ✓ |
| RDD controls | ✓ | ✓ | ✓ |
| Student controls | | ✓ | ✓ |
| Test score controls | | | ✓ |

*Notes:* Table 12 presents heterogeneity analyses for the results reported in Table 11, where enrollment is instrumented for using receipt of an offer for a high-track program with high-quality peers interacted with students' gender. Columns (2) and (3) iteratively include controls for student characteristics and baseline 8th-grade test scores, respectively. The sample is limited to applicants with non-missing baseline test scores, a high track propensity score of one, and a high-quality peer high track propensity between one and zero. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Effects of High Track Access with High Quality Peers (Peer Diligence)

| | (1) | (2) | (3) |
|---|---|---|---|
| **Panel A: Class Average 8th-Grade Test Scores (Mechanism)** | | | |
| High-Quality Peer Program | 0.332*** | 0.207*** | 0.209*** |
| | (0.034) | (0.032) | (0.032) |
| **Panel B: 10th-Grade Average Test Scores (Effect)** | | | |
| High-Quality Peer Program | 0.132 | 0.105 | 0.096 |
| | (0.083) | (0.089) | (0.066) |
| **Panel C: 8th-Grade Average Test Scores (Placebo)** | | | |
| High-Quality Peer Program | 0.034 | 0.015 | — |
| | (0.083) | (0.091) | — |
| $N$ | 1,127 | 1,127 | 1,127 |
| Propensity score FE | ✓ | ✓ | ✓ |
| RDD controls | ✓ | ✓ | ✓ |
| Student controls | | ✓ | ✓ |
| Test score controls | | | ✓ |

*Notes:* Table 13 presents estimates from 2SLS models for enrollment in a high-track program with high-quality peers, where enrollment is instrumented with receipt of an offer for a high-track program with high-quality peers, including saturated propensity score and running variable controls. High-track programs with high-quality peers are those with above-median class-average grades for diligence measured in the 8th grade. Columns (2) and (3) iteratively include controls for student characteristics and baseline 8th-grade test scores, respectively. Panel A demonstrates the mechanism, or the leave-own-out 10th-grade class-level average of baseline 8th-grade test scores. Panel B reports the effect of enrollment in a high-track program with high-quality peers on a student's own 10th-grade average test scores. Panel C contains the results of a placebo test, reporting the effect of enrollment in a high-track program with high-quality peers on a student's own 8th-grade average test scores. The sample is limited to applicants with non-missing baseline test scores, a high track propensity score of one, and a high-quality peer high track propensity between one and zero. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Heterogeneous Effects of High Track Access with High Quality Peers (Peer Diligence)

| | (1) | (2) | (3) |
|---|---|---|---|
| **Panel A: Dep. Var.: 10th-Grade Average Test Scores** | | | |
| High-Quality Peer Program × Male | 0.196** | 0.087 | 0.118 |
| | (0.086) | (0.100) | (0.074) |
| High-Quality Peer Program × Female | 0.087 | 0.116 | 0.082 |
| | (0.084) | (0.096) | (0.072) |
| **Panel B: Dep. Var.: 10th-Grade Math Test Scores** | | | |
| High-Quality Peer Program × Male | 0.349*** | 0.048 | 0.126 |
| | (0.102) | (0.118) | (0.100) |
| High-Quality Peer Program × Female | -0.011 | 0.069 | 0.063 |
| | (0.095) | (0.109) | (0.092) |
| **Panel C: Dep. Var.: 10th-Grade Reading Test Scores** | | | |
| High-Quality Peer Program × Male | 0.042 | 0.126 | 0.110 |
| | (0.102) | (0.121) | (0.091) |
| High-Quality Peer Program × Female | 0.185* | 0.164 | 0.101 |
| | (0.101) | (0.117) | (0.089) |
| **Panel D: Dep. Var.: 10th-Grade University Aspirations** | | | |
| High-Quality Peer Program × Male | -0.080 | -0.171** | -0.168** |
| | (0.057) | (0.070) | (0.069) |
| High-Quality Peer Program × Female | 0.005 | -0.039 | -0.042 |
| | (0.055) | (0.072) | (0.072) |
| *N* | 1,127 | 1,127 | 1,127 |
| Propensity score FE | ✓ | ✓ | ✓ |
| RDD controls | ✓ | ✓ | ✓ |
| Student controls | | ✓ | ✓ |
| Test score controls | | | ✓ |

*Notes:* Table 14 presents heterogeneity analyses for the results reported in Table 13, where enrollment is instrumented for using receipt of an offer for a high-track program with high-quality peers interacted with students' gender. Columns (2) and (3) iteratively include controls for student characteristics and baseline 8th-grade test scores, respectively. The sample is limited to applicants with non-missing baseline test scores, a high track propensity score of one, and a high-quality peer high track propensity between one and zero. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 15: Statistical Tests for Balance Computed via Local Bandwidth

| | Full Sample | Selected Sample |
|---|---|---|
| | (1) | (2) |
| Baseline 8th-grade math test score (std.) | 0.717*** | -0.029 |
| | (0.007) | (0.063) |
| Baseline 8th-grade reading test score (std.) | 0.828*** | -0.071 |
| | (0.007) | (0.063) |
| Female | 0.157*** | -0.103** |
| | (0.004) | (0.050) |
| Age (in years) | -0.087*** | 0.025 |
| | (0.004) | (0.041) |
| SES (CSH-index) | 0.757*** | 0.093 |
| | (0.007) | (0.068) |
| Any social benefits | -0.127*** | -0.009 |
| | (0.004) | (0.050) |
| Deprived neighborhood | -0.048*** | -0.019 |
| | (0.002) | (0.027) |
| Single-parent | -0.065*** | -0.022 |
| | (0.004) | (0.049) |
| Parent w/ maturity exam or higher | 0.262*** | -0.007 |
| | (0.004) | (0.040) |
| *N* | 51,135 | 2,228 |
| Propensity score FE | | ✓ |
| RDD controls | | ✓ |

*Notes:* Table 15 presents regressions of student characteristics pre-track assignment on an indicator for whether the student was offered a place in a high-track program. Column (1) refers to the full sample. Column (2) refers to the selected sample of "at-risk" students with non-degenerate assignment risk, and controls both for high-track assignment risk and running variables. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 16: Local Bandwidth Selected Sample: High Track Access and 10th-Grade Outcomes

| | (1) | (2) | (3) |
|---|---|---|---|
| **Panel A: Dep. Var.: 10th-Grade Average Test Scores** | | | |
| High Track | 0.138*** | 0.119** | 0.154*** |
| | (0.053) | (0.053) | (0.040) |
| High Track × Boy | 0.209*** | 0.172*** | 0.210*** |
| | (0.054) | (0.063) | (0.050) |
| High Track × Girl | 0.069 | 0.079 | 0.111** |
| | (0.057) | (0.061) | (0.046) |
| **Panel B: Dep. Var.: 10th-Grade Mathematics Test Scores** | | | |
| High Track | 0.222*** | 0.180*** | 0.215*** |
| | (0.063) | (0.062) | (0.051) |
| High Track × Boy | 0.397*** | 0.220*** | 0.242*** |
| | (0.064) | (0.074) | (0.062) |
| High Track × Girl | 0.051 | 0.150** | 0.195*** |
| | (0.066) | (0.070) | (0.058) |
| **Panel C: Dep. Var.: 10th-Grade Reading Test Scores** | | | |
| High Track | 0.054 | 0.058 | 0.093* |
| | (0.063) | (0.063) | (0.051) |
| High Track × Boy | 0.020 | 0.124 | 0.178*** |
| | (0.064) | (0.076) | (0.063) |
| High Track × Girl | 0.086 | 0.008 | 0.028 |
| | (0.068) | (0.073) | (0.059) |
| $N$ | 2,228 | 2,228 | 2,228 |
| Propensity score FE | ✓ | ✓ | ✓ |
| RDD controls | ✓ | ✓ | ✓ |
| Student controls | | ✓ | ✓ |
| Baseline test scores | | | ✓ |

*Notes:* In addition to saturated propensity score and running variable controls, Columns (2) and (3) iteratively include controls for student characteristics and baseline 8th-grade test scores, respectively. The sample is limited to applicants with non-missing baseline test scores. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

# Appendix for Online Publication

## A The Deferred Acceptance Algorithm

The student-proposing deferred acceptance algorithm proceeds as follows (Abdulkadiroğlu and Sönmez, 2003):

> *Step 1:* Each student proposes to their first choice. Each school tentatively assigns its seats to its proposers one at a time following their priority order. Any remaining proposers are rejected.
>
> In general, at
>
> *Step k:* Each student who was rejected in the previous step proposes to their next choice. Each school considers the students it has seated together with its new proposers and tentatively assigns its seats to these students one at a time following their priority order. Any remaining proposers are rejected. The algorithm terminates when no student proposal is rejected and each student is assigned their final tentative assignment.

Per Gale and Shapley (1962), the resulting matches are both stable and student-optimal. Given all students weakly prefer the school they are matched to, there is no justified envy.

// IMPRINT

**ZEW – Leibniz-Zentrum für Europäische Wirtschaftsforschung GmbH Mannheim**

ZEW – Leibniz Centre for European Economic Research

L 7,1 · 68161 Mannheim · Germany
Phone +49 621 1235-01
info@zew.de · zew.de